Enterprise Computing Stage 6 (Year 12) – teacher support resource

**Data visualisation**

# Teacher support resource

**Teacher note:** this resource has been designed to facilitate the ready conversion into a student booklet by removing the answers within the response windows. Teacher notes can be deleted before distributing to students.

Student name:

Class:

Teacher:

Contents

[Teacher support resource 1](#_Toc182842402)

[Unit overview 5](#_Toc182842403)

[Glossary 6](#_Toc182842404)

[NESA glossary keywords 8](#_Toc182842405)

[The design and production process 9](#_Toc182842406)

[Using data to tell a story 10](#_Toc182842407)

[Explain the purposes of data visualisation 13](#_Toc182842408)

[Simplifying understanding 15](#_Toc182842409)

[Telling a story 18](#_Toc182842410)

[Highlighting significant results 22](#_Toc182842411)

[Describe how features of software contribute to a better understanding of datasets through data visualisation, including spreadsheets, creative design applications and combining applications to track trends and forecast 25](#_Toc182842412)

[Identify patterns in data by interpreting and comparing datasets for an enterprise, social or ethical issue to highlight trends and for predictive data analytics 28](#_Toc182842413)

[Investigate the impact of the evolution of hardware and software on the field of data analytics 32](#_Toc182842414)

[Processing power 32](#_Toc182842415)

[Storage/memory 33](#_Toc182842416)

[Communication media 34](#_Toc182842417)

[Describe online analytical processing (OLAP) 39](#_Toc182842418)

[Assess data integrity in the development of a data visualisation 43](#_Toc182842419)

[Ownership 43](#_Toc182842420)

[Source 44](#_Toc182842421)

[Validation 44](#_Toc182842422)

[Risk 44](#_Toc182842423)

[Explain the impact of enterprise data warehousing on data visualisation 48](#_Toc182842424)

[Analysis and use of historical data trends and patterns 48](#_Toc182842425)

[Correlation with current data 49](#_Toc182842426)

[Data refinement/optimisation 49](#_Toc182842427)

[Explain how big data affects the design and development of data visualisation 52](#_Toc182842428)

[Scope of visible information 53](#_Toc182842429)

[Types and depth of insight provided by the data 53](#_Toc182842430)

[Evaluate bias in data collection, storage and analysis when developing visualisations 56](#_Toc182842431)

[Accuracy 56](#_Toc182842432)

[Audience 57](#_Toc182842433)

[Data source 57](#_Toc182842434)

[Unconscious bias 57](#_Toc182842435)

[Interpreting data visualisations 59](#_Toc182842436)

[Evaluate the effectiveness of software tools used to develop data visualisations 62](#_Toc182842437)

[Spreadsheets used to develop dashboards 64](#_Toc182842438)

[Presentation software used to present data analysis 65](#_Toc182842439)

[Business analytics services, including ‘as a service’ products 65](#_Toc182842440)

[Custom software solutions 66](#_Toc182842441)

[Interrogate data from a data visualisation 67](#_Toc182842442)

[Interpreting what you see 68](#_Toc182842443)

[Aggregation 68](#_Toc182842444)

[Filtering 69](#_Toc182842445)

[The effect of outliers 69](#_Toc182842446)

[Reasoning 70](#_Toc182842447)

[Designing for user experience 74](#_Toc182842448)

[Use graphic design tools to assist in the graphic development of a data visualisation 74](#_Toc182842449)

[Key principles of UX design 74](#_Toc182842450)

[Explain how user experience (UX) influences the development of effective data visualisations 78](#_Toc182842451)

[Relevance to the audience 78](#_Toc182842452)

[Audience interpretation 79](#_Toc182842453)

[Customisation 79](#_Toc182842454)

[Live analysis 80](#_Toc182842455)

[Develop and implement criteria for evaluating the effectiveness of user experiences 82](#_Toc182842456)

[Investigate the impact of emerging hardware and software technologies on user interface (UI) and UX design and development 89](#_Toc182842457)

[Introduction to UI and UX 89](#_Toc182842458)

[Creating data visualisations 94](#_Toc182842459)

[Research, source, organise and store data appropriate for a data visualisation 94](#_Toc182842460)

[Design and develop a data visualisation for a specific scenario to represent trends, patterns and relationships, and illustrate predictive analysis incorporating big data 98](#_Toc182842461)

[Investigate and implement methods to maintain data security 100](#_Toc182842462)

[Cybersecurity 100](#_Toc182842463)

[Data backup 103](#_Toc182842464)

[References 108](#_Toc182842465)

# Unit overview

In this unit, students will develop a fundamental understanding of the principles and practices of data visualisation. The lessons and sequences are designed to equip students with the knowledge and skills needed to collect, analyse, and present data using various visualisation techniques. Throughout the unit, students will explore how data visualisations can effectively communicate insights and support decision-making processes.

Weeks 1 to 5 will focus on using data to tell a compelling story. Students will learn to transform raw data into meaningful visualisations that highlight key messages. They will explore the importance of user interfaces (UI) and user experience (UX) in designing visualisations that are accessible, clear, and engaging. This will extend their understanding of how data can be translated into information that influences audience perception and comprehension.

Weeks 6 to 10 will involve more advanced data analysis and interrogation techniques, where students will create visualisations that inform decisions. During this time, they will delve into methods for evaluating data bias and ensuring the accuracy of insights. Students will also reflect on the impact of their visualisations on decision-making and explore how data-driven conclusions can lead to improved outcomes.

The unit will culminate in an assessment task where students will use spreadsheets, databases, data analysis, modelling and simulation software to create data visualisations. They will be required to present and communicate a ‘data story’ that showcases their ability to analyse and visualise data effectively. Additionally, students will be encouraged to integrate their skills in data visualisation into their Enterprise Computing project, identifying real-world applications and opportunities for deeper exploration.

# Glossary

Many of the following words will gather more meaning to you as you work through this booklet.

Each time you see a new word in bold throughout this workbook you can add its definition in the table below in case you need to refer back later.

|  |  |
| --- | --- |
| Word | Definition |
| Aggregation | The process of combining things or amounts into a single group or total. |
| Bias | A particular tendency, trend, inclination, feeling, or opinion, especially one that is preconceived or unreasoned. |
| Big data | Extremely large datasets that may be analysed computationally to reveal patterns, trends and associations. |
| Cybersecurity | The protection of information technology elements, including hardware and software, data or network services. |
| Data backup | Data backup is the practice of copying data from a primary to a secondary location, to protect it in case of a disaster, accident or malicious action. |
| Data integrity | Data integrity is the overall accuracy, completeness and consistency of data. Data integrity also refers to the safety of data. |
| Data visualisation | Data visualisation is the graphical representation of information and data. By using visual elements like charts, graphs and maps, data visualisation tools provide an accessible way to see and understand trends, outliers and patterns in data. |
| Data warehousing | The practice of collecting data from a wide range of sources within a company into a single database that may be used to guide management decisions. |
| Hardware | The collection of physical components that constitute a computer system, all of which are tangible objects. Hardware includes both internal and external components of the system, as well as any peripheral devices that can be connected. |
| Online analytical processing (OLAP) | OLAP (online analytical processing) is a computing method that enables users to easily and selectively extract and query data in order to analyse it. |
| Patterns | In a technological context, a pattern might be recurring sequences of data over time that can be used to predict trends, particular configurations of features in images that identify objects, frequent combinations of words and phrases, or particular clusters of behaviour. |
| Predictive analysis | Predictive analytics is the use of data to predict future trends and events. It uses historical data to forecast potential scenarios. |
| Relationships | The way in which 2 or more people or things are connected, or the state of being connected. |
| Software | The programs and other operating information used by a computer. |
| Trends | A general direction of change: a way of behaving or proceeding, that is developing and becoming more common. For example, the downward/upward trend of the stock market. Digital technology is the latest/current trend in television. |
| User experience (UX) | The usability, ease of use and enjoyment provided in the interaction between the customer and the product. |

**Teacher note:** for students with an English as an additional language or dialect (EAL/D) background – the glossary can be provided complete so that they have additional time to understand the key terms with bilingual dictionaries. The glossary can be provided to students in their preferred communication mode.

# NESA glossary keywords

NESA keywords can be used in the syllabus and in the Higher School Certificate examination. Familiarisation with these keywords can assist in understanding how to write and respond to questions.

|  |  |
| --- | --- |
| Key term | Definition |
| Apply | Use, utilise, employ in a particular situation |
| Compare | Show how things are similar or different |
| Describe | Provide characteristics and features |
| Evaluate | Make a judgement based on criteria; determine the value of |
| Examine | Inquire into |
| Explain | Relate cause and effect; make the relationships between things evident; provide why and/or how |
| Investigate | Plan, inquire into and draw conclusions about |
| Outline | Sketch in general terms; indicate the main features of |

[NESA: Glossary of key words](https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/hsc/hsc-student-guide/glossary-keywords)

**Teacher note:** configure, design, explore and implement are used in this topic and are not listed.

# The design and production process

Throughout your study of Enterprise Computing, you will learn about design processes and how to apply them. You will explore different types of design processes and learn how to apply them in your design project.

The design and production process:

* involves a sequence of organised steps which provide a solution to design needs and opportunities
* may take a few seconds or minutes, such as when you select what clothes to wear, or may take years as in the case with the design of a motor vehicle
* may involve one person or may involve many people
* may be simple or complex, depending on the task
* involves questioning (or evaluating) throughout the iterative process.

Figure 1 – flowchart of design and production process

Design and production process diagram
A flowchart labelled 'Ongoing evaluation' with a two-headed arrow indicating both directions. 
The first part of the flowchart is called '1. Identifying and defining'. It says 'identify and define the needs, opportunities and wants of a computing challenge, practise the technical skills, develop evaluation criteria.' There is an arrow pointing to the next section, which is labelled '2. researching and planning'. It says 'research, generate and practise ideas, be creative and propose new approaches to problems, explore new design opportunities.' An arrow points to the next section, labelled '3. producing and implementing', it says 'build and implement ideas, apply a variety of skills and techniques to create products that meet set criteria, modify and iterate solutions'. The arrow points to the next section, labelled '4. testing and evaluating'. It says 'test and evaluate solutions/products, evaluate quality and effectiveness against the criteria, make judgements throughout the solution and use these to refine the product.'
After testing and evaluating is a big arrow called 'Review if required to improve' and it goes all the way back up to the first part of the flowchart, indicating a cycle.

# Using data to tell a story

As a class watch [Enterprise Computing Year 12 Unit 2: Data visualisation (23:58)](https://www.youtube.com/watch?v=g4biYeAfkmU) for an overview of the focus area.

Data visualisation is the graphical representation of data to produce easily interpreted information.

By using visual elements like charts, graphs and maps, data visualisation tools provide an accessible way to see and understand trends, outliers and patterns in data.

In the world of big data, data visualisation tools and technologies are essential to analyse massive amounts of information and make data-driven decisions

Infographics are everywhere, they allow highly visual representations of, at times, very complex data sets.

Figure 2 – Australian Government visualisation

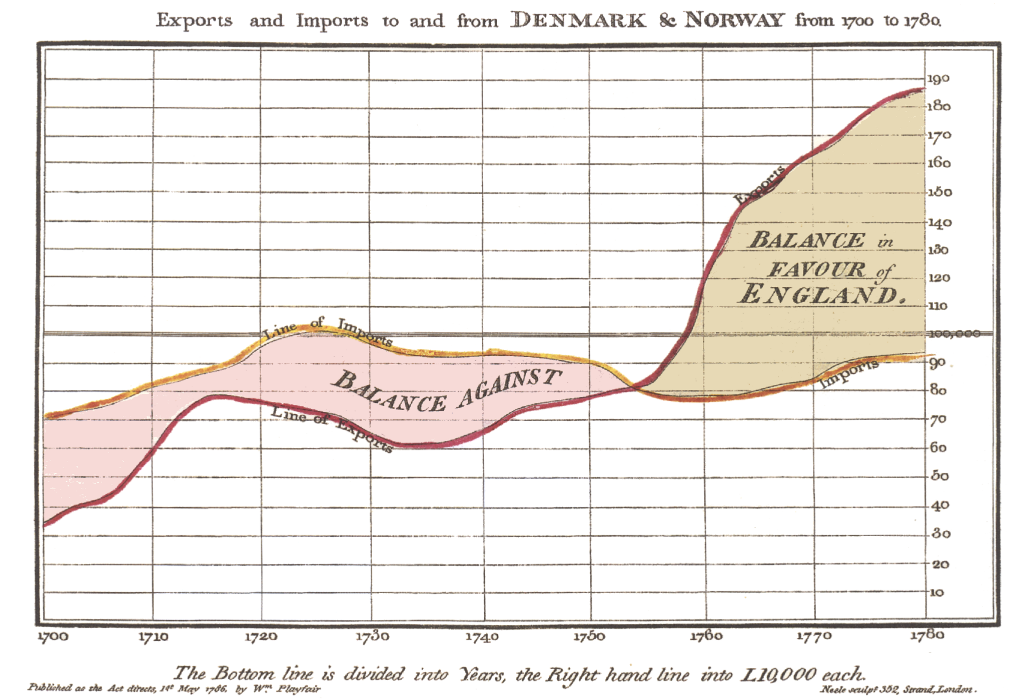
Example of a graphical user interface or timeline with 9 different data visualisations of COVID-19. 


© Commonwealth of Australia (2020)

Visualising data has been around for a very long time.

These 2 examples from 1786 and 1858 provide interesting case studies of how raw data that would have been recorded in a ledger, can be shown visually.

Figure 3 – William Playfair: Exports and Imports 1786



[‘William Playfair's Time Series of Exports and Imports of Denmark and Norway’](https://commons.wikimedia.org/wiki/File:Playfair_TimeSeries-2.png) is in the public domain.

Figure 4 – Florence Nightingale: Mortality 1858

This graphic indicates the annual rate of mortality per 1,000 in each month that occurred from preventable diseases (in blue), those that were the results of wounds (in red), and those due to other causes (in black).

The legend reads:

The areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in November 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red, & the black lines enclosing them.


[‘Example of polar area diagram by Florence Nightingale (1820–1910)’](https://en.wikipedia.org/wiki/File:Nightingale-mortality.jpg) is in the public domain.

****As a class, watch [Mapping data visualisations meteoric rise from Victorian London to today (6:00)](https://aeon.co/videos/mapping-data-visualisations-meteoric-rise-from-victorian-london-to-today).

**Activity 1:** visualisation simplifies complex datasets for better understanding and their use in enterprise systems

As a class, look at [Visualization as Argument, William Playfair’s Time-Series Charts](https://dataxdesign.io/chapters/playfair) and discuss the value of data visualisation.

Describe how visualisation simplifies complex datasets for better understanding.

|  |
| --- |
| **Sample answer:**  Data visualisation is the process of representing data and information graphically through visual elements such as charts, graphs, maps and infographics. It allows complex datasets to be communicated in a clear and accessible way, making it easier for users to identify trends, patterns and insights. By transforming raw data into visual formats, data visualisation simplifies understanding, enhances storytelling, and highlights key information, enabling informed decision-making and more effective communication. |

Identify key reasons for using data visualisation in enterprise systems.

|  |
| --- |
| **Sample answer:**  Data visualisation in enterprise systems makes it easier to understand large amounts of data by turning it into simple graphs or charts. This helps people make better decisions and communicate more clearly with others in the business. It also allows them to track important information in real-time, which saves time and improves how the company works. By spotting patterns and trends, businesses can plan for the future and make smarter choices, and it helps companies stay on top of regulations by making reporting easier. Overall, it lets everyone make faster and more informed decisions. |

## Explain the purposes of data visualisation

As a class watch [Data Visualization in 2024 | The Ultimate Guide (11:01)](https://www.youtube.com/watch?v=loYuxWSsLNc) and discuss the purposes of data visualisation.

Data visualisation helps to simplify complex data by breaking it down into a more accessible and understandable format. This makes it easier for people to comprehend the information and identify patterns and trends that may not be immediately apparent from the raw data.

Data visualisation is a powerful tool for communicating data and information to others. By visualising data, you can convey information quickly and effectively, making it easier for people to understand and retain the information.

Data visualisation is often used to support decision-making processes by providing clear and concise information that can be used to inform decisions based on data-driven insights.

Data visualisation can help you identify patterns and trends in your data that may not be immediately apparent. By using different types of graphs and charts, you can highlight relationships between different data points and gain insights into your data.

Data visualisation is also a valuable tool for exploring and understanding data. By visualising data, you can quickly spot patterns and relationships that may not be immediately apparent, allowing you to gain deeper insights into your data and make more informed decisions.

**Activity 2:** data visualisation uses and purpose

Describe data visualisation you use on a regular basis.

|  |
| --- |
| **Sample answers:**  Timetable – a colour-coded weekly schedule or calendar that visually represents subjects, class periods, and extracurricular activities.  Weather apps – visuals like temperature charts, hourly forecasts, and rain probability graphs are seen daily when checking weather apps to plan for school days.  Fitness or health trackers – use smartwatches or phone apps that display data like steps taken, active minutes or sleep patterns through bar charts, pie charts and line graphs.  Game analytics – in online games or mobile apps, performance dashboards may display statistics about playing time, progress, achievements or rankings, often visualised as bar charts or progress meters. |

Explain the purpose of data visualisation.

|  |
| --- |
| **Sample answer:**  The purpose of data visualisation is to transform complex data into easily understandable visual formats, simplifying the interpretation of information for a broad audience. By using charts, graphs and other visual tools, data visualisation helps to ‘tell a story’ by guiding viewers through the data, making trends and patterns more apparent. It also highlights significant results by drawing attention to key insights, enabling quicker decision-making and deeper understanding. This process turns raw data into clear, actionable information that can be used to communicate findings effectively and drive informed decisions. |

### Simplifying understanding

**Use of visual cues**

Visual cues, such as colours, shapes and sizes can make it easier to understand data by allowing the viewer to quickly identify trends and patterns. For example, a bar graph with different colours for each category can make it easier to compare the relative size of each category.

Figure 5 – visual cues



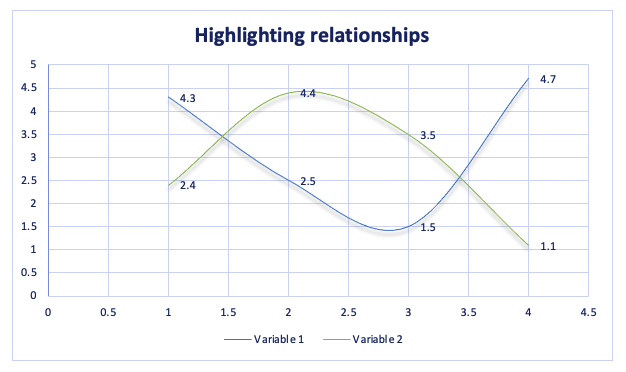
**Condensation of information**

Data visualisation condenses large amounts of information into a compact and easily digestible format. This makes it easier for the viewer to understand and retain the information, as they don't have to process and remember a large amount of raw data.

**Highlighting relationships**

Data visualisation can highlight relationships between different data points, making it easier for the viewer to see the connections between them. For example, a scatter plot can show the relationship between 2 variables, making it easier to see if there is a correlation between them.

Figure 6 – highlighting relationships



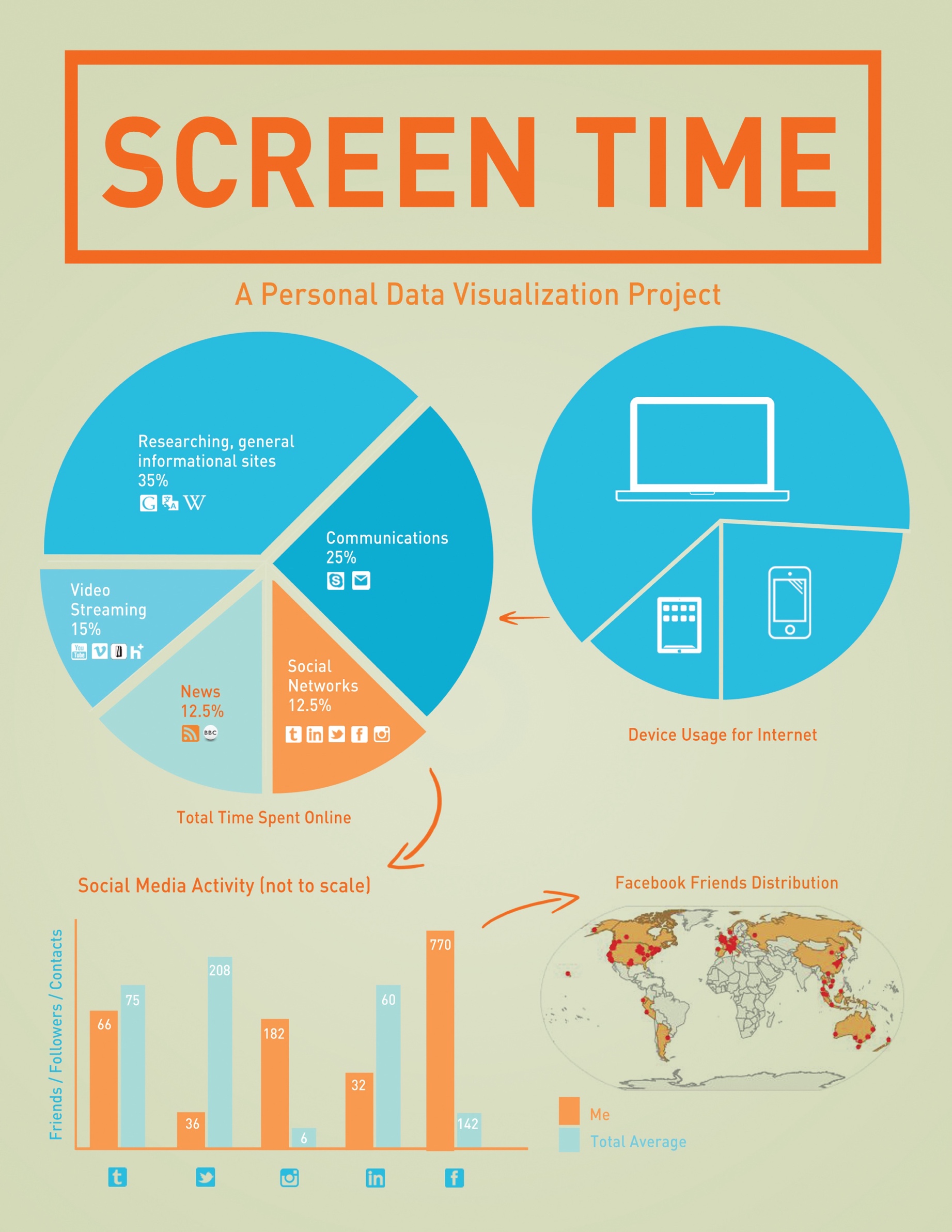
**Providing context**

Data visualisation provides context for the data by showing it in relation to other data points. This can help the viewer to understand the meaning of the data and see how it fits into the bigger picture.

**Making data more accessible**

Data visualisation can make data more accessible to a wider audience, including people who may not have a background in statistics or data analysis. By visualising the data, people can understand it without having to know how to interpret raw data.

Figure 7 – screen time visualisation



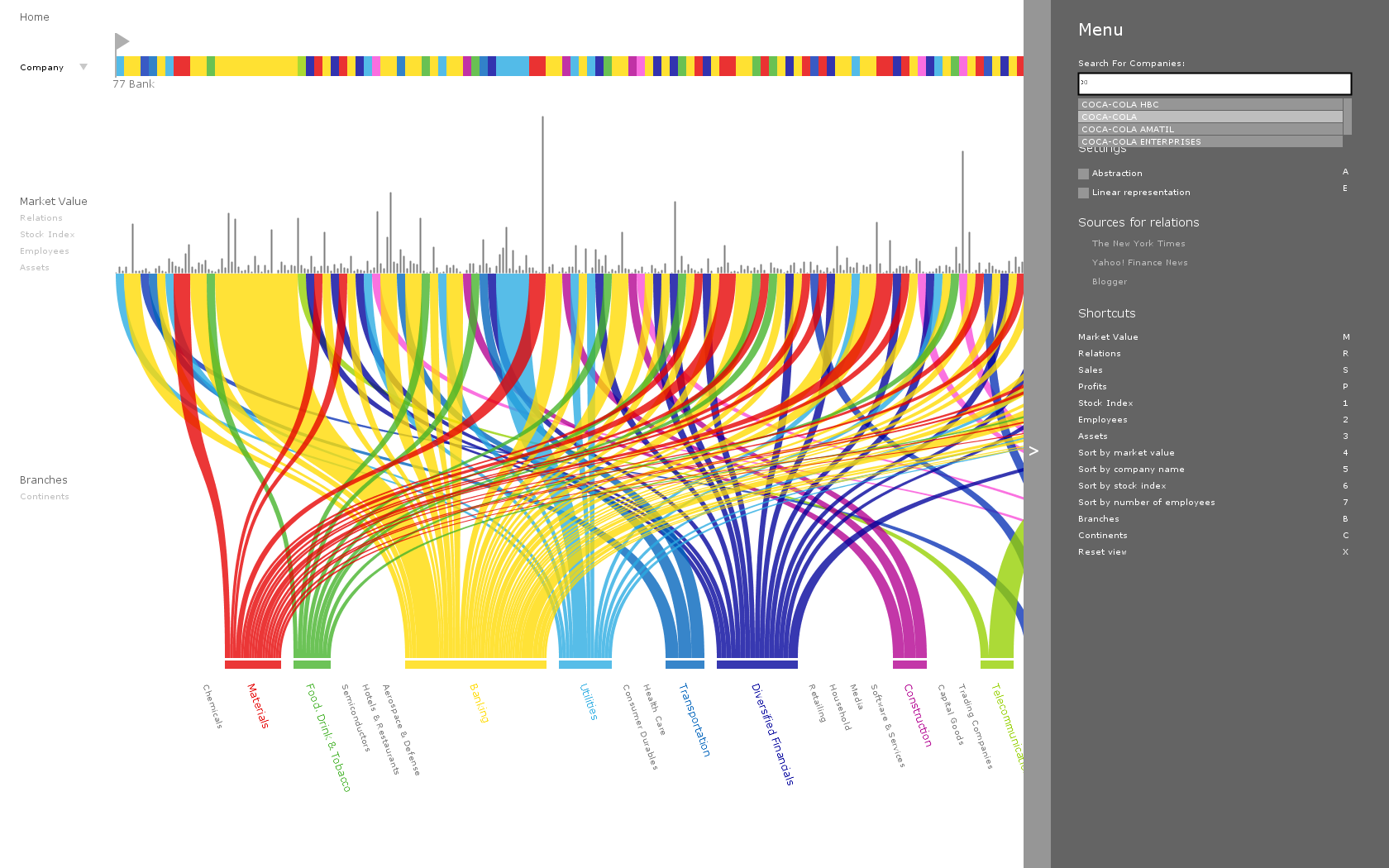
‘[Data Visualization](https://www.flickr.com/photos/17266692@N05/9002633633)’ by [Tran Dinh](https://www.flickr.com/photos/17266692@N05) is licensed under [CC BY-NC-SA 2.0](https://creativecommons.org/licenses/by-nc-sa/2.0/?ref=openverse).

### Telling a story

Data visualisation can tell a story by creating a visual narrative that conveys information and insights about the data.

Effective data visualisation should be able to communicate the key messages and insights from the data in an intuitive and engaging way.

Figure 8 – visual data mining



‘[KP “Visual Data Mining” 4](https://www.flickr.com/photos/31907740@N00/4860841137)’ by [Stefan Wagner](https://www.flickr.com/photos/31907740@N00) is licensed under [CC BY-NC-SA 2.0](https://creativecommons.org/licenses/by-nc-sa/2.0/?ref=openverse).

**Contextualising the data**

Data visualisation can provide context for the data by showing it in relation to other data points. This can help to tell the story of how the data has changed over time, or how it compares to other datasets.

**Highlighting trends and patterns**

Data visualisation can highlight trends and patterns in the data that may not be immediately apparent from raw data. This can help to tell the story of how the data is evolving and what insights can be gleaned from it.

**Showing relationships**

Data visualisation can show relationships between different data points, making it easier to see how they are connected. This can help to tell the story of how different factors are impacting the data and what the relationships between them are.

**Making the data more engaging**

Data visualisation can make the data more engaging by using visuals that are appealing and easy to understand. This can help to tell the story of the data in a way that is interesting and memorable for the viewer.

**Communicating insights**

Data visualisation can communicate insights from the data in a way that is easy to understand and retain. By highlighting the key insights, data visualisation can tell the story of what the data means and what actions can be taken based on that information.

Figure 9 – communicating insight into relationships on social media

A data visualisation of social media with  colourful dots and lines.



‘[My Twitter account as data visualization](https://www.flickr.com/photos/13518023@N03/3985827494)’ by [Michael Gallagher](https://www.flickr.com/photos/13518023@N03) is licensed under [CC BY-SA 2.0](https://creativecommons.org/licenses/by-sa/2.0/?ref=openverse).

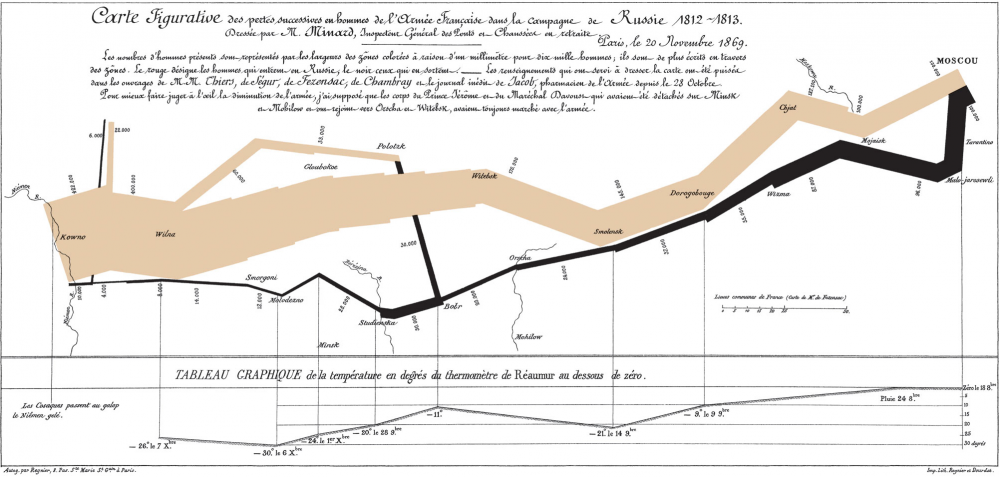
**Activity 3:** the best data visualisation

****As a class read and investigate the [10 Of The Best Data Visualization Examples From History & Today](https://www.tableau.com/learn/articles/best-beautiful-data-visualization-examples).

[Think, Pair, Share](https://app.education.nsw.gov.au/digital-learning-selector/LearningActivity/Card/645) in a small group about what in your opinion, is the best data visualisation you have seen.

Look at the data of Napoleon’s 1812 march on Russia.

Figure 10 – Napoleon’s March by Charles Minard



‘[Charles Minard's 1869 chart showing the number of men in Napoleon’s 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path. Lithograph, 62 × 30 cm](https://en.wikipedia.org/wiki/File:Minard.png)’ is in the public domain.

As a class watch [Minard's Famous "Napoleon's March" Chart – What It Shows, What It Doesn't (11:23)](https://www.youtube.com/watch?v=hlb1uM_SOcE).

In 1869 Charles Joseph Minard, a French civil engineer known for his information graphics, produced what is widely regarded as the best visualisation graph of all time.

What does this data visualisation show?

|  |
| --- |
| **Sample answer:**  The visualisation shows Napoleon’s army, departing at full-strength from the Polish border to Russia and then their subsequent retreat. |

What information can be gleaned from these numbers?

|  |
| --- |
| **Sample answer:**  The numbers alone don’t tell much of a story. |

Compare the thickness of the 2 lines and discuss the scale of the losses.

|  |
| --- |
| **Sample answer:**  The thickness of the lines represents the size of the army, with the beige line representing the march on Moscow and the black line representing the retreat back to Poland. |

What other forms of data have been represented in the visualisation?

|  |
| --- |
| **Sample answer:**  Napoleon set off with 422,000 troops and returned with just 10,000.  Geographical features such as locations and rivers crossed have been plotted, as well as the varying temperatures at different points on the march and retreat. |

Closely inspect this visualisation.

Explain where and why tragedy occurred at various points on the journey.

|  |
| --- |
| **Sample answer:**  During the retreat on 28 September 1813, the data labels on the black line read 50,000, before dropping to 28,000, showing that 22,000 men died crossing the Berezina River near Minsk! |

Explain how this data visualisation highlights significant trends and insights.

|  |
| --- |
| **Sample answer:**  Minard’s visualisation is highly effective because it combines geography, time and data about troop losses, clearly showing the disastrous impact of both the Russian army and the severe winter. It visually tells the story of Napoleon's failed campaign better than raw numbers could, making it easy to grasp the scope of the disaster at a glance. |

### Highlighting significant results

Data visualisation can highlight significant results by using various graphical and pictorial elements to emphasise the most important insights and findings from the data.

Here are some ways that data visualisation can highlight significant results:

* **Emphasising key trends:** data visualisation can highlight key trends in the data by using graphs and charts that emphasise changes over time. For example, a line graph can show how a variable has changed over time, making it easy to see any significant trends.
* **Comparing data points:** data visualisation can compare data points by using graphs and charts that show the relative size or importance of each data point. For example, a bar graph can show the comparison between different categories, making it easy to see which categories are the most significant.
* **Highlighting outliers:** data visualisation can highlight outliers by using graphs and charts that emphasise data points that are significantly different from the rest of the data. For example, a scatter plot can show the relationship between 2 variables and emphasise data points that are significantly outside of the expected range.
* **Showing patterns and relationships:** data visualisation can show patterns and relationships between different data points by using graphs and charts that highlight these relationships. For example, a heatmap can show the relationship between 2 variables, making it easy to see any significant patterns or correlations.
* **Using colour to emphasise significance:** data visualisation can use colour to emphasise significance by using different colours to represent different data points. For example, red can be used to represent data points that are significantly higher than expected, while blue can be used to represent data points that are significantly lower.

**Activity 4:** favourite topping on Superbowl pizza state by state

Students explore trends in pizza toppings using the article [Super Bowl pizza sales reach into the millions](https://www.usatoday.com/in-depth/graphics/2023/02/09/national-pizza-day-super-bowl-favorite-toppings-graphics/11201105002/) as a source of information. They analyse how the data is presented in the article.

Are there any pizza toppings that are outliers (either much more or much less popular than others)? Describe these outliers and what might explain their difference from the rest of the data.

|  |
| --- |
| **Sample answer:**  Yes, pesto is only voted by one state as a favourite topping and anchovies are a clear outlier in terms of popularity. According to the data, anchovies are much less popular than other pizza toppings, with only about 2% of people choosing it. This low popularity might be explained by the fact that anchovies have a strong, salty taste that many people find overpowering or unpleasant. Additionally, they are a less common topping in many regions, which may further reduce their popularity. |

How does identifying outliers help us understand the overall trends in pizza topping preferences? Are there any outliers that skew the data?

|  |
| --- |
| **Sample answer:**  Identifying outliers, like anchovies, helps us understand which toppings are exceptions to the general trend of popular choices. In this case, the fact that anchovies are much less popular than other toppings tell us that people generally prefer more common, familiar flavours like pepperoni, sausage, or mushrooms. Outliers like anchovies don’t necessarily skew the data in this case, but they do show that certain toppings appeal to a smaller, more niche group of consumers, which is important when looking at the overall preferences. |

How does the popularity of meat toppings compare to vegetable toppings in the graph? Which category appears more significant, and why?

|  |
| --- |
| **Sample answer:**  In the graph, meat toppings such as pepperoni and sausage are significantly more popular than vegetable toppings like mushrooms and onions. For example, pepperoni, the most popular meat topping, has about 36% of people choosing it, while mushrooms, the most popular vegetable topping, has a lower percentage. This suggests that meat toppings are generally more preferred by consumers, which might be because they are viewed as more flavourful or indulgent. Meat toppings appear more significant because they dominate the top rankings, showing a clear preference for heartier, more savory options. |

Why is a bar graph a useful tool for comparing data points like the most hated pizza topping? What advantage does it offer over other types of visualisations?

|  |
| --- |
| **Sample answer:**  A bar graph is a useful tool for comparing data points like the most hated pizza topping because it allows you to visually see the differences in popularity between each topping. Bar graphs make it easy to compare the length of each bar, representing the percentage of people who dislike certain toppings. This makes it immediately clear which topping is the most disliked. The advantage of a bar graph over other visualisations, like pie charts, is that it allows for direct, side-by-side comparisons across multiple categories. This makes it easier to spot the most significant data points at a glance, especially when there are many items to compare. |

Perform a search to find data on Australian pizza consumer habits and list publicly available datasets below.

|  |
| --- |
| **Sample answer:**  [Most popular types of pizza ordered among consumers in Australia as at February 2023.](https://www.statista.com/statistics/1410088/australia-popular-pizza-orders/) |

**Extension activity**

****Create a data visualisation of Australian pizza preferences using data that tells a story and share with your group your findings.

## Describe how features of software contribute to a better understanding of datasets through data visualisation, including spreadsheets, creative design applications and combining applications to track trends and forecast

Different software features can contribute to a better understanding of datasets through data visualisation by providing tools for organising, analysing and visualising data in a way that highlights key insights and trends. By combining spreadsheets, creative design applications and other tools, it is possible to create sophisticated data visualisations that help users to gain deeper insights into their data and make informed decisions.

**Spreadsheets**

Spreadsheets such as Microsoft Excel or Google Sheets are a commonly used tool for organising and analysing data. They provide a range of built-in charting tools that allow users to quickly create visual representations of the data. The ability to create pivot tables and custom formulas makes it possible to analyse complex data and create visualisations that highlight key insights.

**Creative design applications**

Creative design applications such as Adobe Illustrator or Canva provide advanced design tools that can be used to create more sophisticated and visually appealing data visualisations. These applications allow users to add graphics, text and visual elements to their visualisations, making it possible to create visual representations that are both informative and engaging.

**Combining applications to track trends and forecast**

By combining different software applications, it is possible to create more sophisticated data visualisations that track trends and forecast future trends. For example, spreadsheets can be used to organise and analyse data, while creative design applications can be used to create visual representations of the data. The combination of these 2 applications can provide a powerful tool for tracking trends and forecasting future trends, making it possible to gain deeper insights into the data.

Investigate the [Surfline](https://www.surfline.com/) website and look at how the charts and data visualisations would use spreadsheets and design applications as part of their creation. Surfline is used by surfers to predict the surf conditions coming up over the next week. The website is an example of combining applications to track trends and forecast conditions such as wind, swell and tide.

**Activity 5:** key features and combing applications

****Describe the key features of spreadsheets and design applications used in data visualisation.

|  |
| --- |
| **Sample answers:**  Spreadsheets and design applications are pivotal in data visualisation, each offering distinct features that enhance the way data is interpreted and communicated.  **Charts and graphs**  Spreadsheets provide built-in tools to create various types of charts, such as bar, line, pie and scatter plots. These are essential for visualising trends and relationships within datasets.  **Conditional formatting**  This feature allows users to highlight cells based on values, helping to identify patterns, trends or anomalies in the data quickly.  **Pivot tables**  Pivot tables summarise large datasets, making it easier to explore data relationships and trends by aggregating information across different categories.  **Trendlines and forecasting**  Spreadsheet tools like Microsoft Excel include options to add trendlines to charts, providing visual representations of data trends over time. Forecasting functions allow predictions based on historical data.  **Formulas and functions**  Spreadsheets offer an array of built-in functions that allow users to manipulate data, which is then represented visually for deeper insights  **Customisable visual elements**  Design applications offer full control over the appearance of charts, graphs, and infographics. Users can modify colours, fonts, and layouts to improve clarity and aesthetic appeal.  **Interactive dashboards**  Tools like Tableau or Power BI enable users to create interactive dashboards where viewers can drill down into specific data points, filter results, and view real-time updates.  **Data integration**  Many design applications allow the combination of data from multiple sources. This feature helps to create more complex visualisations by merging datasets, improving tracking of trends across systems or time periods  **Animation and dynamic visuals**  Some advanced design software supports the creation of animated visualisations, which can help in demonstrating how data evolves over time or in highlighting key insights dynamically.  **Geospatial mapping**  Applications like Tableau enable users to create maps that visualise geographic data, offering insights into location-based trends or patterns. |

****Explain how software can be combined to create more effective data visualisations.

|  |
| --- |
| **Sample answer:**  **Data preparation and analysis (spreadsheets + data management tools)**  Spreadsheets like Microsoft Excel or Google Sheets are widely used for initial data collection, cleaning, and preparation. They offer functions like sorting, filtering, and formula application to organise large datasets. Once the data is refined, it can be exported to more advanced visualisation tools (for example Tableau or Power BI) for more complex representations.  For example, after analysing trends using formulas in Excel, users can import the cleaned data into Tableau to create interactive, dynamic dashboards that allow deeper exploration of the trends. This integration improves efficiency by combining the familiar data manipulation capabilities of spreadsheets with the advanced visualisation and interactivity features of business intelligence platforms.  **Visual design and customisation (spreadsheets + design software)**  While spreadsheets offer basic charting tools, they often lack flexibility in design and aesthetics. To overcome this, users can create initial visualisations in a spreadsheet and then export them to creative design applications like Canva. These design tools allow for extensive customisation, including adding branding elements, adjusting typography, and enhancing visual appeal. This combination is especially useful for creating infographics or presentation-ready visuals that are both data-driven and visually compelling.  For example, an Excel chart can be exported as a vector image to Illustrator, where users can modify colours, add annotations, or incorporate images to tell a more engaging story. |

## Identify patterns in data by interpreting and comparing datasets for an enterprise, social or ethical issue to highlight trends and for predictive data analytics

Identifying patterns in data is an important step in interpreting and comparing datasets for an enterprise, social or ethical issue.

By collecting data, cleaning and organising it, exploring it, comparing datasets and using predictive analytics, it is possible to identify patterns in data, highlight trends and use this information to make informed decisions.

The following are steps that can be taken to identify patterns in data.

1. **Collect data**

The first step in identifying patterns in data is to collect relevant data. This data can come from a variety of sources, including surveys, transactions or other sources of information.

1. **Clean and organise data**

Once data has been collected, it needs to be cleaned and organised so that it can be easily analysed. This may involve removing duplicate data, handling missing values or transforming the data into a format that is more suitable for analysis.

1. **Explore the data**

The next step is to explore the data to identify any patterns or trends. This can involve using various techniques such as data visualisation, statistical analysis or machine learning algorithms to uncover insights and relationships within the data.

1. **Compare datasets**

To identify patterns in data, it is often necessary to compare datasets. This can involve comparing data from different time periods or comparing data from different groups or categories. By comparing datasets, it is possible to identify similarities, differences and trends that can provide valuable insights into the data.

1. **Use predictive analytics**

Predictive data analytics can be used to identify patterns in data and make predictions about future trends. This can involve using machine learning algorithms to analyse data and identify relationships and patterns that can be used to make predictions. Predictive analytics can help enterprises to identify potential opportunities and risks and make informed decisions.

**Activity 6:** examining trends in car colours using data visualisation

As a class read the article [New Car Colours More Monochrome Than Ever Before, Data Shows](https://www.abc.net.au/news/2024-10-10/new-car-colours-more-monochrome-than-ever-before-data-shows/104451484) as a source of information. Analyse data presented in the article and practice interpreting trends through data visualisation techniques and answer the questions below.



The article highlights that cars have become more monochromatic over time. Which specific colours have increased in popularity according to the data in the article?

|  |
| --- |
| **Sample answer:**  The article and data identify that the trend is for shades like grey, silver, black and white have become increasingly dominant in car colours, overtaking more vibrant colours. |

What colours have decreased in popularity based on the data presented?

|  |
| --- |
| **Sample answer:**  According to the article, colours like red, blue and yellow, which were once more common, have seen a decline in popularity. |

If you were to visualise the data from the article using a pie chart, what would the chart represent and why might a pie chart be effective for this type of data?

|  |
| --- |
| **Sample answer:**  A pie chart could represent the percentage distribution of car colours, showing which colours dominate the market. This type of visualisation would effectively highlight the overwhelming share of monochromatic colours compared to others. |

Using the data available on the website, create a bar chart that shows the year-by-year changes in the popularity of grey, black and white cars. How do the trends for these colours compare across the years?

|  |
| --- |
| **Sample answer:**  The chart should show a steady increase in grey, black and white car purchases over the years, indicating a shift toward more neutral tones. |

The article suggests that car colour trends may reflect broader societal or cultural preferences. Based on the data, what cultural or societal trends could be influencing the shift toward monochrome colours?

|  |
| --- |
| **Sample answer:**  Monochrome colours such as black, grey and white are often associated with minimalism, professionalism and technological advancement. The growing preference for these colours might reflect a societal shift toward simplicity, functionality, or a desire for cars to project a sleek, modern image. |

Based on the data trends shown in the article, predict what might happen to car colour preferences in the next 5 to 10 years. Justify your prediction using evidence from the article.

|  |
| --- |
| **Sample answer:**  Given the increasing dominance of monochromatic colours like grey, black and white, it is likely that these colours will continue to be popular. However, there could also be a rise in new variations of these shades (for example matte or metallic finishes) as people seek ways to stand out while staying within the neutral colour palette. |

If you were tasked with creating a dashboard for car manufacturers to help them visualise car colour trends over time, what types of visualisations would you include and why?

|  |
| --- |
| **Sample answer:**  I would include a line graph to show how the popularity of each car colour has changed over time, a pie chart to show the current market share of each colour and possibly an interactive map to show regional differences in car colour preferences. |

After completing the questions, discuss the following in groups:

* How do data visualisations help you better understand the trends in car colours?
* What other industries could benefit from analysing colour trends and how could this information be used?

**Extension activity:**

Use software like Google Sheets, Microsoft Excel or any data visualisation tool to create your own visualisations based on the trends described in the article. Create at least one bar chart and one pie chart, then share your visualisations.

## Investigate the impact of the evolution of hardware and software on the field of data analytics

The evolution of processing power, storage and memory and communication media has had a significant impact on the field of data analytics.

The increase in processing power has enabled the development of new and more sophisticated algorithms and tools that allow enterprises to process, analyse and make sense of large and complex datasets.

The increase in storage and memory capacity has made it possible to store and process large amounts of data, enabling enterprises to perform more sophisticated and comprehensive data analysis.

This evolution of hardware has impacted software and has opened new possibilities to gain deeper insights into their data and make more informed decisions.

This has opened new possibilities for enterprises to gain deeper insights into their data and make more informed decisions.

### Processing power

**Big data processing**: the rise of big data has made it necessary to process and analyse larger and more complex datasets. The increase in processing power has enabled enterprises to process these large data sets in a timely and efficient manner. This has made it possible to extract valuable insights from big data that were previously not possible.

**Advanced analytics**: the evolution of processing power has made it possible to run more advanced analytics on data. This has led to the development of new algorithms and models that can help enterprises to identify patterns, trends and relationships in their data.

**Machine learning**: the increase in processing power has also enabled the development of machine learning algorithms. These algorithms can be used to automatically identify patterns and relationships in data, making it possible to make predictions and forecasts based on the data.

**Real-time analytics**: the increase in processing power has also made it possible to perform real-time analytics on data. This allows enterprises to quickly respond to changing conditions and make data-driven decisions in real-time.

**Cloud computing**: the increase in processing power has also led to the rise of cloud computing. This has made it possible to process and analyse large datasets in the cloud, reducing the need for expensive hardware and making it possible for enterprises of all sizes to access and make use of powerful data analytics tools.

### Storage/memory

**Big data storage**

The rise of big data has made it necessary to store large amounts of data. The increase in storage and memory capacity has enabled enterprises to store and process these large datasets, making it possible to extract valuable insights from big data that were previously not possible.

**Advanced analytics**

The increase in storage and memory capacity has made it possible to perform more advanced analytics on data. This has led to the development of new algorithms and models that can help enterprises to identify patterns, trends and relationships in their data and make more informed decisions.

**Machine learning**

The increase in storage and memory capacity has also enabled the development of machine learning algorithms. These algorithms require large amounts of data to train and make accurate predictions and the increase in storage and memory capacity has made it possible to process and store this data.

**Real-time analytics**

The increase in storage and memory capacity has also made it possible to perform real-time analytics on data. This allows enterprises to quickly respond to changing conditions and make data-driven decisions in real-time.

**Data warehousing**

The increase in storage and memory capacity has also made it possible to build data warehouses, where enterprises can store large amounts of data for analysis. This has made it possible to perform more comprehensive and sophisticated data analysis and has opened new possibilities for enterprises to gain deeper insights into their data.

### Communication media

**Data collection**

The development of new communication media, such as the Internet and mobile devices, has made it possible to collect large amounts of data from a wide range of sources. Data analytics can perform more comprehensive and sophisticated data analysis and has opened up new possibilities for enterprises to gain deeper insights into their data.

**Real-time data**

The increase in the speed and reliability of communication media has made it possible to collect and share data in real-time. This has made it possible to perform real-time analytics, allowing enterprises to quickly respond to changing conditions and make data-driven decisions.

**Cloud computing**

The evolution of communication media has also enabled the rise of cloud computing. Data analytics can store and process large amounts of data in the cloud, reducing the need for expensive hardware and making it possible for enterprises of all sizes to access and make use of powerful data analytics tools.

**Collaboration**

The development of new communication media has also made it possible for individuals and enterprises to collaborate more effectively. Data analytics can share data and insights and has opened up new possibilities for enterprises to work together to gain deeper insights into their data.

**Data visualisation**

The development of new communication media has also made it possible to share data and insights more effectively through data visualisation. This has made it easier for enterprises to communicate their findings to a wide range of stakeholders and has opened up new possibilities for enterprises to make data-driven decisions.

**Activity 7:** analysing the impact of evolving hardware and software in sports

As a class watch [Socceroos World Cup Performance and Data Analysis (1:38–39:50)](https://schoolsnsw.sharepoint.com/sites/TASNSWStatewideStaffroom/_layouts/15/stream.aspx?id=%2Fsites%2FTASNSWStatewideStaffroom%2FShared%20Documents%2F02%2E%20TAS%20professional%20learning%2FAnalysing%20Data%20with%20the%20Socceroos%2FSocceroos%20World%20Cup%20Performance%20and%20Data%20Analysis%20%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2E01b06464%2D4597%2D46ee%2D8f5b%2D9fe894226650).

**Teacher note:** this video is part of the [Professional learning library](https://schoolsnsw.sharepoint.com/sites/TASNSWStatewideStaffroom/SitePages/TAS-Professional-Learning-Library.aspx) of the TAS Statewide Staffroom. The selected duration shows a presentation from Doug Kour, Data Analyst who had just worked with the Socceroos during the World Cup. This video is downloadable and relevant for students, however the introduction and question and answer session at the end are aimed specifically at teacher professional learning.

**Scenario:**

You are a data analyst working for an Australian soccer organisation that tracks the performance of players across different leagues and levels. Over the past decade, the organisation has upgraded its hardware and software infrastructure to enhance its data analytics capabilities. These upgrades include more powerful processing systems, increased storage for player data, and improved communication tools for real-time analysis and collaboration.

****In small groups, discuss the following questions and assess how these technological advancements have improved the organisation’s ability to analyse player performance data and predict future trends in soccer.

**Step 1:** understanding the key areas of technological evolution.

**Big data processing power**

With advanced processing power, the organisation can now analyse millions of data points on player performance, such as speed, distance covered and goal accuracy during each match.

* How does increased processing power help you quickly analyse and make sense of large volumes of player data from multiple games in a season?

**Machine learning**

Machine learning algorithms allow the organisation to predict future player performance by analysing patterns in past data (for example how a player’s performance changes based on their workload, injury history or age).

* How has machine learning allowed the organisation to make better predictions about player performance and fitness levels?

**Real-time analytics**

Increased processing power enables real-time analysis of player data during live matches. This allows coaches to make immediate tactical adjustments based on player fatigue or team performance.

* How do real-time analytics improve decision-making during a match?

**Big data storage**

The organisation now stores vast amounts of historical data on every player’s performance, allowing for more comprehensive analysis.

* How does increased storage capacity allow you to track player performance over multiple seasons and compare current data with historical trends?

**Machine learning**

Machine learning models require vast amounts of training data to make accurate predictions. The increased storage capacity allows the organisation to keep detailed data on players and matches, enabling more accurate model training.

* Why is large storage capacity essential for running machine learning models in soccer performance analysis?

**Real-time data**

The development of real-time communication tools allows coaches and analysts to receive live data during matches. For example, sensors worn by players can transmit data instantly to the analytics team.

* How does real-time data collection improve the way you track player fitness and performance during live games?

**Data visualisation and collaboration**

Improved communication media allows analysts to create data visualisations and share them instantly with coaching staff. For example, heat maps of player movements during a game can be shared on tablets or screens for immediate review.

* How has the ability to quickly share data visualisations improved collaboration between the data team and coaching staff?

**Step 2:** analysing the impact on soccer performance tracking.

1. How has increased processing power improved the organisation’s ability to process and analyse large amounts of player performance data?

|  |
| --- |
| **Sample answer:**  With increased processing power, it is possible to process data from sensors on all players during a match, tracking their speed, distance covered, and fatigue levels in real-time. This helps identify when a player is underperforming or at risk of injury. |

1. How has machine learning changed the way you predict player performance and potential injuries?

|  |
| --- |
| **Sample answer:**  Machine learning models predict when a player is likely to experience a decline in performance or sustain an injury based on their workload, training patterns, and previous injuries. This helps manage player workloads and prevent injuries before they occur. |

1. Why is increased storage capacity important for analysing long-term player development?

|  |
| --- |
| **Sample answer:**  With increased processing power, we can now process data from sensors on all players during a match, tracking their speed, distance covered, and fatigue levels in real-time. This helps identify when a player is underperforming or at risk of injury. |

1. How does real-time data collection help coaches make better decisions during a match?

|  |
| --- |
| **Sample answer:**  Real-time data collection provides coaches with instant feedback on player performance, allowing them to make quick substitutions or tactical changes when a player is fatigued or underperforming. |

1. How have data visualisation tools helped improve communication between analysts and coaches?

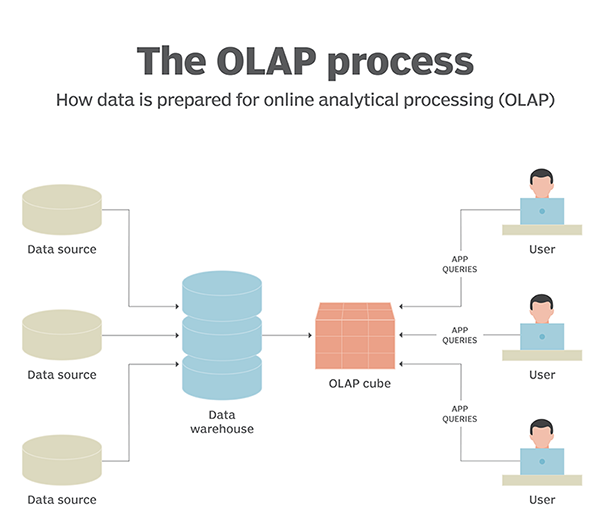
|  |
| --- |
| **Sample answer:**  Data visualisation tools allow us to create heat maps and charts that show player movements and game statistics, which we can share with coaches during or after the match. This makes it easier to explain complex data and make quick adjustments. |

## Describe online analytical processing (OLAP)

Online analytical processing (OLAP) is a data analysis technique used to enable business intelligence and decision-making. It is a type of software that enables users to analyse and manipulate large, multi-dimensional data sets. OLAP provides a powerful set of tools for data analysis and allows users to perform advanced analytics on data stored in a data warehouse or multidimensional database.

The main advantage of OLAP is its ability to perform complex data analysis quickly and easily. OLAP provides a user-friendly interface for performing data analysis and it allows users to perform complex calculations, such as roll-ups, drill-downs and slicing and dicing, on large, multi-dimensional datasets.

Figure 11 – the OLAP process



[‘OLAP process’](https://commons.wikimedia.org/w/index.php?curid=97417812%22%3eFile:Crm-olap%20desktop.png) by Margaret Rouse is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/?ref=openverse).

OLAP is typically used to analyse data in areas such as sales, finance and marketing and it is designed to provide business decision-makers with quick and easy access to the data they need to make informed decisions.

OLAP is based on the concept of multi-dimensional data analysis, where data is organised into a cube-like structure, with each dimension representing a different aspect of the data, such as time, geography, product, or customer. The cube-like structure of OLAP data makes it possible to perform complex data analysis quickly and easily and it provides a powerful set of tools for data analysis and visualisation.

As a class watch [OLAP cubes: Unleashing Multidimensional Data Insights (7:47)](https://www.youtube.com/watch?v=s8UDqSRXgbg).

**Activity 8:** building an OLAP cube

****In small groups, imagine that you are building an OLAP cube to analyse sales data for a retail company. Each group will choose 3 dimensions (for example time, geography, product) and build a cube-like structure to organise the data.

**Steps:**

1. **Choose dimensions**: pick 3 dimensions (for example time, geography and product) and describe how each one will be structured. For example:

* Time: Year → Quarter → Month → Week
* Geography: Country → State → City
* Product: Category → Brand → Product

1. **Define measures**: choose at least 2 measures to analyse within the cube (for example total sales, average order size).
2. **Create a scenario**: write a scenario in which your company needs to use OLAP to make a business decision. For example, your company wants to know which product category had the highest sales in Australia during Quarter 1 of 2024.
3. **Perform analysis**: use OLAP techniques (roll-up, drill-down, slicing, dicing) to answer your scenario’s questions. Present your findings to the class and explain how the OLAP cube helped you make an informed business decision.

|  |
| --- |
| **Sample answer:**  **A table showing electronics and clothing sales across Australia.**  **Note**: this is just for key cities; other states and cities could be added in a complete OLAP cube.  Perform OLAP Analysis.  **Roll-up**: we can roll-up the data by summarising total sales across all cities in Australia. This gives us a high-level view of how each product category performs across the entire country.  **Summary**:   * Electronics total sales: $2.03 million AUD * Clothing total sales: $1.05 million AUD.   **Drill-down**: to investigate which city contributes the most to the total sales in Australia, we can drill-down into the individual city data for each product category.  **Insight**: Sydney had the highest sales for both categories, with $600,000 AUD in electronics and $300,000 AUD in clothing.  **Slicing**: by slicing the data, we can focus only on electronics sales in Australia, filtering out clothing data entirely. This allows us to analyse the electronics category independently.  **Insight**: electronics performed significantly better than clothing, with total sales of $2.03 million AUD compared to $1.05 million AUD for clothing.  **Dicing**: we can dice the data to compare electronics and clothing sales in specific cities, such as Sydney and Melbourne, during Quarter 1 2024. This helps us compare performance across different regions for both product categories.  **Insight**: in both Sydney and Melbourne, electronics sales outperformed clothing sales. Electronics sales were $600,000 AUD in Sydney and $550,000 AUD in Melbourne, whereas clothing sales were $300,000 AUD and $280,000 AUD in the same cities.  **Key finding**: the Electronics category had the highest total sales in Australia during Quarter 1 of 2024, generating $2.03 million AUD in revenue. The Clothing category performed significantly lower with $1.05 million AUD in sales.  **Business insight**: the company should consider investing more in marketing electronics, especially in cities like Sydney and Melbourne, where the demand is highest. The average order size for electronics is also higher than for clothing, suggesting that customers are willing to spend more on electronics.  This analysis demonstrates how using OLAP techniques – roll-ups, drill-downs, slicing, and dicing – can provide business insights from large, multi-dimensional datasets. In this scenario, the focus was on the Australian market, allowing for region-specific business decision-making. |

**Activity 9:** OLAP reflection questions

How does OLAP allow businesses to make more informed decisions?

|  |
| --- |
| **Sample answer:**  OLAP allows businesses to quickly analyse large sets of data across multiple dimensions, helping them identify trends, patterns, and outliers. By providing tools like roll-ups and drill-downs, OLAP enables decision-makers to explore data at both high-level and detailed views, which leads to more accurate and informed decisions. |

Why is multi-dimensional data analysis important in business?

|  |
| --- |
| **Sample answer:**  Multi-dimensional data analysis is important because it allows businesses to explore complex relationships between different factors like time, location, and product type. This type of analysis provides a deeper understanding of what drives performance, which can help companies optimise operations, improve marketing strategies, and increase sales. |

## Assess data integrity in the development of a data visualisation

Before creating any data visualisation, it is crucial to assess data integrity to ensure that the visualisation is both accurate and reliable. Data integrity refers to the consistency, accuracy and trustworthiness of the data being used, and this is influenced by several key factors.

First, **ownership** of the data addresses who controls and maintains the data, helping establish accountability for its quality. Next, the **source** of the data must be verified to ensure that it is credible, up-to-date, and relevant for the context. **Validation** processes, such as checking for consistency, completeness, and correctness, help ensure that the data is free of errors or inconsistencies. Finally, the potential **risks** involved with using unreliable or misrepresented data need to be considered, as these can lead to misleading visualisations and poor decision-making.

|  |  |
| --- | --- |
| Feature | Description |
| Ownership | Refers to the person or team responsible for the data and the data visualisation. |
| Source | Indicates where the data comes from, including the format and method of collection. |
| Validation | Refers to the process of checking the accuracy and completeness of the data, including outliers and errors. |
| Risk | Refers to the likelihood of data corruption, errors or privacy concerns. |

### Ownership

* Determines who is responsible for the accuracy, quality and completeness of the data.
* This can include the person or team who collected the data, the organisation that owns the data, or the data visualisation team.
* Ownership can also determine who has the authority to make changes or corrections to the data or visualisation.

### Source

* Indicates the origin of the data, including the type of data, the data format and the method of collection.
* This can include the dataset used, data processing and cleaning methods and data storage locations.
* Understanding the source is crucial for interpreting and analysing the data.

### Validation

* Refers to the process of verifying the accuracy and completeness of the data.
* This can include checking for data anomalies, missing data, outliers and data errors.
* Validation helps ensure that the data is reliable and can be trusted for analysis and decision-making.

### Risk

* Refers to the likelihood of data corruption, errors or privacy concerns.
* This can include the risk of data breaches, data loss or unauthorised access to the data.
* Understanding the potential risks associated with the data can help inform decisions about how to handle and protect the data.

Understanding and addressing these components helps create data visualisations that are trustworthy and insightful.

**Activity 10:** assess a dataset

1. **Choose a dataset**

****Select a dataset to analyse. You could use the dataset you intend to use for Assessment task 2 or the scenario of school attendance or a dataset from an online source (for example a public government dataset).

1. **Analyse the data integrity**

****For each of the 4 categories (ownership, source, validation and risk), answer the questions below to assess the integrity of the data.

**Step 1: Ownership**

* Who owns the data?
* Who is responsible for collecting and maintaining the data? Is it an individual, a team or an organisation?
* Who is responsible for ensuring the data’s accuracy and quality?
* Does the data owner verify the data’s accuracy before sharing or visualising it?
* Who can make changes or corrections to the data?
* If you find errors in the data, who has the authority to fix them?

|  |
| --- |
| **Sample answer:**  The data is owned by the school’s administrative team, which is responsible for collecting and maintaining attendance records. The team ensures that attendance is recorded accurately each day. The school’s data management team has the authority to correct any errors in the data. |

**Step 2: Source**

* Where does the data come from?
* What is the origin of the data? (for example, sensor readings, surveys, manually entered data)
* What format is the data in?
* Is it a CSV file, database or other format? How was the data processed and cleaned before being stored?
* How was the data collected?
* Was the data collected automatically or manually? Are there any known limitations in the collection process?

|  |
| --- |
| **Sample answer:**  The data comes from manual daily attendance sheets that are entered into a school database. The data is stored in a CSV format, and staff members input the data at the end of each day. A limitation of this process is that manual entry can lead to potential errors, such as typos. |

**Step 3: Validation**

* Has the data been validated for accuracy?
* Were any checks performed to ensure the data is accurate and complete?
* Are there any anomalies or outliers in the data?
* Are there any values that don’t make sense or fall outside the expected range?
* Are there any missing data points?
* Is any data missing or incomplete?

|  |
| --- |
| **Sample answer:**  The data was partially validated by checking for missing entries at the end of each week. However, a closer inspection revealed outliers where attendance records were extremely low on certain days, possibly due to manual entry errors. There are also some missing data points where attendance wasn't recorded for a few students. |

**Step 4: Risk**

* What are the potential risks of using this data?
* Could there be privacy concerns, data breaches, or unauthorised access to the data?
* Is there a risk of data corruption or loss?
* How is the data protected from accidental loss or corruption?
* What steps can be taken to mitigate these risks?
* What measures should be in place to protect the data?

|  |
| --- |
| **Sample answer:**  There is a risk of data loss if the system crashes or if manual records are misplaced before being entered into the system. Additionally, there is a privacy risk as the attendance data includes personal information about students, which could be exposed if not properly secured. To mitigate these risks, the school could use data encryption and ensure that only authorised personnel can access or edit the data. |

## Explain the impact of enterprise data warehousing on data visualisation

As a class watch [What Is a Data Warehouse? (3:31)](https://www.youtube.com/watch?v=AHR_7jFCMeY).

Enterprise data warehousing (EDW) plays a pivotal role in enhancing data visualisation by providing a centralised, organised and optimised repository of data. By enabling the analysis of historical trends, correlation with current data and refinement of information, an EDW empowers businesses to make data-driven decisions with confidence. Data visualisation, supported by a robust EDW, transforms complex data into actionable insights, making it an essential tool for modern enterprises.

**Introduction to enterprise data warehousing (EDW)**

An enterprise data warehouse (EDW) is a central repository where an organisation stores and manages its data. This data comes from various sources, such as sales records, customer information, financial transactions and more. EDWs are designed to handle large volumes of data, organise it effective and make it accessible for analysis and decision-making. One of the critical roles of an EDW is to support data visualisation, which is the process of representing data in graphical formats like charts, graphs and dashboards.

### Analysis and use of historical data trends and patterns

**Historical data storage**

An EDW collects and stores data over extended periods, sometimes spanning several years. This extensive collection allows analysts to study historical trends and patterns. For example, a company might analyse sales data over the past 5 years to identify seasonal trends or the impact of marketing campaigns.

**Enhanced analysis**

With an EDW, organisations can quickly access historical data, compare it with current data and visualise changes over time. Data visualisation tools, connected to an EDW, can generate graphs that show trends such as increasing or decreasing sales, customer behaviour changes, or shifts in market demand. This historical analysis is crucial for making informed business decisions, as it helps identify what strategies worked in the past and what didn’t.

**Predictive insights**

By analysing historical data, businesses can also predict future trends. For instance, if historical data shows a consistent increase in product sales during certain months, the company can prepare for similar demand in the future. Data visualisation makes these predictions easier to understand by presenting them in clear, graphical formats.

### Correlation with current data

**Real-time data integration**

An EDW not only stores historical data but also integrates current, real-time data from various sources. This integration enables businesses to see the complete picture by correlating past and present data. For example, a retailer can compare last year's holiday sales with the current year's data to see if their new marketing strategies are effective.

**Visualisation of correlations**

Data visualisation tools connected to an EDW can create charts and graphs that show correlations between historical and current data. For instance, a line graph might display last year’s monthly sales alongside this year’s data, making it easy to see how current trends align with or diverge from past patterns. This correlation helps businesses understand the factors driving current performance.

**Impact on decision-making**

Understanding these correlations is vital for making strategic decisions. If a business sees that current trends are not aligning with historical patterns, they may need to adjust their strategies. Data visualisation simplifies this process by making complex correlations easier to interpret and act upon.

### Data refinement/optimisation

**Data cleaning and transformation**

Before data can be useful, it often needs to be cleaned and transformed. An EDW supports this process by providing tools to refine and optimise data. This might involve removing duplicates, filling in missing values, or converting data into a consistent format. Clean, well-organised data is crucial for accurate analysis and effective visualisation.

**Optimised data for visualisation**

Once the data is refined, it can be used to create more accurate and meaningful visualisations. Clean data ensures that the charts and graphs generated are reliable and easy to interpret. For example, a scatter plot showing the relationship between customer age and purchasing habits is only useful if the data is accurate and complete.

**Improving data insights**

Data optimisation also enhances the ability to uncover deeper insights. Refined data, when visualised, can reveal subtle patterns and trends that might be missed in raw data. For example, a heatmap of customer activity might show that certain regions are consistently underperforming, prompting further investigation and targeted strategies.

**Activity 11:** enterprise data warehouse questions

****Read [Enterprise Data Warehouse: EDW Components, Key Concepts, and Architecture Types](https://www.altexsoft.com/blog/enterprise-data-warehouse-concepts/).

Define data warehousing and its role in enterprise data management.

|  |
| --- |
| **Sample answer:**  Data warehousing refers to the process of collecting, storing, and managing large volumes of structured data from various sources within an organisation.  A data warehouse serves as a central repository where data is systematically organised and made available for query and analysis.  Its primary role in enterprise data management is to consolidate data from disparate systems (such as databases, transactional systems, and external sources) into a unified environment, ensuring that data is clean, consistent and ready for use.  This allows businesses to gain a comprehensive view of their operations, enabling more accurate reporting, historical analysis and data-driven decision-making.  Data warehouses often support online analytical processing (OLAP), which enables the complex querying of data across multiple dimensions, further enhancing business intelligence and analytics capabilities within an enterprise. |

****Explain how data warehouses support the creation of accurate and comprehensive visualisations.

|  |
| --- |
| **Sample answer:**  Data warehouses play a critical role in supporting the creation of accurate and comprehensive visualisations by providing a reliable foundation of clean, integrated data. Because data warehouses aggregate and standardise data from various sources, visualisations derived from this data are more likely to be accurate and free of inconsistencies or errors.  The structured nature of a data warehouse allows users to slice and dice data in different ways, enabling in-depth analysis and exploration across multiple dimensions, such as time, geography or product categories.  Furthermore, data warehouses are designed to handle large datasets, making it easier to generate comprehensive visualisations that reflect trends and patterns over extended time periods. By ensuring the integrity and consistency of data, warehouses enable businesses to produce visualisations that offer meaningful insights and drive informed decision-making. |

## Explain how big data affects the design and development of data visualisation

Big data has a significant impact on the design and development of data visualisation due to its volume, velocity and variety. Here are a few ways in which big data affects data visualisation:

**Handling large volumes of data**

* Big data sets can contain millions or even billions of data points, which can be challenging to manage and process.
* Data visualisation designers must choose appropriate techniques and tools to handle large data volumes effectively, such as using data reduction methods to eliminate redundant or insignificant data.
* This means that the visualisation must be designed to handle and present large amounts of data effectively and efficiently.

**Dealing with high velocity data**

* Big data is often generated in real-time or at a high velocity, requiring real-time data visualisation and analysis.
* Designers must choose visualisation techniques that allow for real-time processing and display of data, such as dashboards and live updates.
* The visualisation must be designed to keep up with the rapid pace of incoming data to provide up-to-date information.

**Handling data variety**

* Big data often comes in a variety of formats and types, such as structured, semi-structured and unstructured data.
* Designers must choose visualisation techniques that are flexible enough to handle multiple data types, including data from different sources.
* This means that the visualisation must be designed to present different data formats in a meaningful and coherent way.

### Scope of visible information

In terms of the scope of visible information, big data visualisation requires a focus on the most relevant data points.

Designers must choose what information to display and what to leave out, based on the intended audience and the insights they want to provide. In other words, big data visualisation requires prioritisation and selective attention to present the most relevant and valuable information.

This can involve using filtering or aggregation techniques to reduce the volume of data while still conveying essential insights. Ultimately, the scope of visible information must strike a balance between displaying enough information to provide insights while not overwhelming the audience with too much data.

### Types and depth of insight provided by the data

Big data has a significant impact on the design and development of data visualisation as it provides an opportunity to gain deeper insights and make better-informed decisions. Here are some ways in which big data affects data visualisation in terms of the types and depth of insight provided:

* **Increased complexity**:big data often includes a wide variety of data types, which can be difficult to analyse and visualise effectively. Visualisation designers must choose techniques that can handle complex data relationships, such as network visualisations and multi-dimensional visualisations.
* **More granular insights:** big data provides a much more detailed view of trends and patterns in the data, which can lead to more granular insights. Designers must choose visualisation techniques that can present data at a fine-grained level, such as heatmaps and scatter plots.
* **Improved predictive analytics:** big data provides an opportunity to use advanced analytics techniques such as machine learning to gain predictive insights from the data. Designers must choose visualisation techniques that can present predictive insights in an intuitive and actionable way, such as decision trees and dynamic charts.
* **Increased speed of insights:** big data often requires real-time or near-real-time analysis and insights, which can help organisations to make timely decisions. Designers must choose visualisation techniques that can present insights quickly and efficiently, such as dashboards and live data feeds.

**Activity 12:** investigate the Transport for NSW Open Data website

****The [Transport for NSW Open Data](https://opendata.transport.nsw.gov.au/dataset/) website contains large datasets of information. Investigate some of the 200 available datasets and list potential datasets that could inspire the theme of your data story in Assessment task 2.

|  |
| --- |
| **Sample answers:**  **Opal trip data**  This dataset contains information about train, bus, ferry, and metro usage, collected through Opal tap-on and tap-off data. It could be used to analyse public transport trends, peak travel times, or commuter behaviour across different regions in NSW.  **Road crash data**  Data on road crashes in NSW, which could be useful for analysing road safety trends, identifying high-risk areas, and exploring the impact of infrastructure improvements on accident rates.  **Cycling data**  Data on cycling activity, including cycleway counts and cycling propensity in various areas. This dataset could help students investigate the rise in cycling as a mode of transport, assess infrastructure effectiveness, or highlight environmental impacts.  **Public transport timetables and performance**  This dataset includes real-time data on public transport schedules, delays, and performance metrics, offering insights into public transport reliability, efficiency, and accessibility in different regions. |

****Imagine you work for the government in Transport NSW. Explain how big data affects the design and development of data visualisation.

|  |
| --- |
| **Sample answer:**  **Handling volume of data**  Big data sources, such as Opal card tap-on/tap-off data, generate millions of data points daily. Visualising this massive volume requires designing scalable visualisation tools that can efficiently process and display vast amounts of data. For instance, when visualising trends in commuter behaviour across NSW, we might need to aggregate data to make patterns more understandable while ensuring no critical information is lost.  **Real-time data and velocity**  Transport NSW often deals with real-time data from public transport systems, road traffic sensors, and GPS devices, which must be processed quickly to inform decision-making. The velocity of this data demands that visualisations can dynamically update, allowing for real-time monitoring of traffic, transport delays, or incidents.  **Data variety and integration**  The variety of data sources, ranging from transport usage patterns, road crash reports, weather data, to public feedback, means that we must create visualisations capable of integrating multiple datasets. Designing visualisations that merge different types of data—such as combining Opal card usage with weather data to identify how rain impacts public transport use—helps us to uncover hidden relationships and deliver more comprehensive insights to policymakers and the public.  **Ensuring accessibility and usability**  Big data visualisation must ensure that users of various expertise levels can easily interact with the visualisation. For instance, Transport NSW must cater to both technical users, such as urban planners, and the general public. Interactive visualisations that allow users to drill down into specific data points (for example, zooming in on a specific station’s usage data or filtering by time) are critical for making complex data understandable and actionable. |

## Evaluate bias in data collection, storage and analysis when developing visualisations

Evaluating bias in data collection, storage, and analysis is crucial when developing visualisations because bias can significantly distort the accuracy and fairness of the insights presented. Bias in data collection might occur if certain groups or variables are overrepresented or underrepresented, leading to skewed results that don’t reflect reality.

In data storage, bias can arise if the data is organised in a way that emphasises specific outcomes or excludes important factors. During analysis, biased algorithms or methods may prioritise certain patterns while ignoring others, resulting in misleading conclusions. If these biases are not addressed, the visualisations created may reinforce inaccurate or unfair assumptions, leading to poor decision-making and perpetuating inequalities.

Therefore, identifying and minimising bias ensures that visualisations are objective, balanced and truly reflective of the underlying data.

**Activity 13:** bias in accuracy, audience, data source and unconscious bias

Complete the tables below to show the impact on bias in data collection, storage and analysis when developing visualisations.

### Accuracy

|  |  |  |
| --- | --- | --- |
| Evaluation | Examples | Impact |
| The accuracy of data collected needs to be checked for errors and inconsistencies to avoid bias in the visualisation. | Inaccurate data could result in skewed visualisations and misinterpretations of data.  For example, a survey that only includes responses from a specific demographic could create bias in the data. | **Sample answer:**  Biased data collection, storage and analysis can lead to visualisations that perpetuate stereotypes, further marginalise certain groups, or lead to misguided decisions. |

### Audience

|  |  |  |
| --- | --- | --- |
| Evaluation | Examples | Impact |
| The audience should be taken into consideration when designing the visualisation to avoid any bias that could lead to misinterpretation of the data. | Designing a visualisation with the audience in mind, such as using simpler language and clear design, can help avoid bias in interpretation. | **Sample answer:**  Biased visualisations can lead to incorrect conclusions, poor decision-making and damage to reputations. |

### Data source

|  |  |  |
| --- | --- | --- |
| Evaluation | Examples | Impact |
| Data sources should be evaluated for potential bias, such as selective sampling or omitted variables and attempts should be made to mitigate these issues. | Data sources, such as surveys or studies, may have limitations that lead to biased results. For example, a survey may only sample certain regions or populations, leading to incomplete or inaccurate data. | **Sample answer:**  Biased data sources can lead to incomplete or inaccurate conclusions, further perpetuating bias. |

### Unconscious bias

|  |  |  |
| --- | --- | --- |
| Evaluation | Examples | Impact |
| Unconscious biases in the selection, analysis and interpretation of data can lead to inaccurate or misleading visualisations and should be avoided. | Unconscious biases, such as favouring certain data or ignoring data that doesn't support a particular narrative, can lead to inaccurate visualisations and conclusions. | **Sample answer:**  Unconscious biases can lead to misleading visualisations that perpetuate stereotypes, misrepresent data and lead to poor decision-making. |

**Activity 14:** propose methods to minimise bias in developing data visualisations

What methods would minimise bias in developing data visualisations?

|  |
| --- |
| **Sample answers:**  **Use clear and consistent scales**  When creating visualisations, ensure that all axes and scales are consistent and accurate. Misleading scales, such as uneven intervals or inconsistent baselines, can create false impressions of trends.  **Choose appropriate graph types**  Selecting the right graph type for the data is crucial. For example, using pie charts to show percentages or using bar graphs to compare different categories helps in accurately conveying information.  **Label everything clearly**  Clearly labelling axes, data points, and legends helps viewers understand what each part of the visualisation represents. This reduces the chance of misinterpretation due to missing or vague labels.  **Avoid overloading with unnecessary information**  Simplify visualisations to only include the necessary elements. Adding too many colours, lines or markers can create confusion and mislead viewers into focusing on irrelevant data.  **Provide context with titles and descriptions**  A good title and a brief description can help provide context for the visualisation. Describing what the data represents, where it comes from, and why it matters helps in avoiding biased interpretations. |

# Interpreting data visualisations

As a class view the Data visualisation PowerPoint.

Data visualisations are graphical representations of data that help communicate complex information clearly and efficiently. Common types of visualisations include bar charts, line graphs, pie charts and scatter plots. They enable readers to quickly grasp patterns, trends and outliers in datasets. Effective interpretation of these visualisations involves understanding the context, identifying the variables being compared and analysing the key takeaways the visual is meant to convey.

For example, a line graph might show a company's revenue growth over time, with the x-axis representing time and the y-axis showing revenue. A sudden spike or drop may indicate important events, such as product launches or market shifts.

**Tips for interpreting data visualisations**

* **Identify the chart type**: each chart serves a different purpose (for example bar charts for comparisons, pie charts for proportions, line charts for trends).
* **Check the axes**: ensure you understand what is being compared and measured.
* **Look for patterns or trends**: peaks, valleys and plateaus in charts can indicate significant events.
* **Consider the source**: always consider the credibility of the data and the source presenting the visualisation.

As a class watch [Using Design Techniques for Clear and Appealing Data Visualization (4:16)](https://www.youtube.com/watch?v=0Smgm2UTUSo).

**Activity 15:** reading a graph

Figure 12 – graph

A graph with different coloured bars that shows smart phone users across countries and their preferred brands. 



‘[Smartphone Brands Retain Home Turf Advantage](http://www.statista.com/chart/25032/brand-of-main-smartphone-survey/)’ by statista is licensed under [CC BY-ND](https://creativecommons.org/licenses/by-nd/4.0/).

Analyse the graph above and write a reflection on what you think the graph is communicating to you. To guide you with your response, start with some observations:

* What is the topic of the graph?
* What quantities are being compared?
* What are some observations that you can make based on the graphs?
* What surprises you about the graph? What do you wonder?
* What do you foresee happening in the next 10 years?

|  |
| --- |
| **Sample answer:**  The graph displays data on the most popular smartphone brands used by consumers. It compares the percentages of users for each major brand, such as Apple, Samsung, and others, across a certain period.  Some observations that can be made from the graph include the clear dominance of Apple and Samsung as leading brands, each holding a significant portion of the market.  One surprising element might be the position of other brands, such as Huawei or Xiaomi, which may have a strong presence in certain regions but struggle to gain market share in others. This leads to a question: Will the rising popularity of budget-friendly smartphones shift these numbers in the coming years?  Looking ahead, I foresee that in the next 10 years, Apple and Samsung will continue to dominate, but newer brands may rise as technology advances, particularly in emerging markets where affordability plays a key role. |

**Teacher note:** [Turner’s Graph of the Week](https://www.turnersgraphoftheweek.com/) is a great resource for finding thought-provoking data visualisations that are current and regularly updated.

## Evaluate the effectiveness of software tools used to develop data visualisations

As a class watch [Data Analytics – The 9 Essential Tools! 2024 (4:07)](https://www.youtube.com/watch?v=jgXp1EE4Wms) and discuss the tools available to students to use in the Assessment task.

Several software tools allow users to create data visualisations, each offering different features, flexibility and usability. Common tools include Microsoft Power BI, Tableau, Google Data Studio and Microsoft Excel. These tools vary in terms of complexity, with some focusing on advanced analytics and others offering simple, user-friendly interfaces for basic visualisation needs.

Tableau, for instance, is highly effective for complex, interactive dashboards, while Excel remains a versatile tool for simpler visualisations.

Key factors to consider when evaluating data visualisation tools:

* **Ease of use:** How intuitive is the interface for users with varying levels of expertise?
* **Data integration:** How well does the tool connect with different data sources (databases, spreadsheets, cloud services)?
* **Customisation:** Does the tool allow users to customise visualisations based on specific requirements?
* **Interactivity:** Can users interact with the data, such as filtering or zooming into certain views?
* **Cost:** Does the pricing model fit the needs of the organisation or individual user?

**Activity 16:** evaluate data visualisation software tools

Complete the table choosing 3 software tools to evaluate and describe the benefits of using these tools.

|  |  |
| --- | --- |
| Software tool | Evaluation |
| Microsoft Excel | **Sample answer:**  Microsoft Excel is a versatile and widely used tool for data visualisation, especially for those with basic to intermediate skills.  Its built-in chart features allow users to create a range of visualisations like bar charts, line graphs, pie charts and scatter plots.  Excel’s accessibility is also a significant benefit, as most users are familiar with its interface, making it ideal for quick and straightforward visual representations.  Excel allows for extensive customisation of charts, including data labels, colour schemes, and formatting. It is particularly useful for smaller datasets and quick visualisations. However, it may be limited for advanced analytics or interactive features.  Cost – Excel is included in most Office subscriptions, making it an economical option for data visualisation. |
| Microsoft Power BI | **Sample answer:**  Power BI provides more sophisticated and interactive data visualisation capabilities compared to Excel.  It is designed to connect with multiple data sources and automatically update visualisations, making it suitable for dynamic reporting.  One of the core benefits is the interactivity in reports. Users can click through different data points and drill down into specific segments, enhancing the analytical experience.  Power BI supports scalable reports that can handle large datasets efficiently.  It also allows collaboration through shared dashboards and reports, making it ideal for team projects and business settings. However, it requires more advanced knowledge to use effectively. |
| Flourish in Canva | **Sample answer:**  Flourish creates interactive visualisations that go beyond basic charts, allowing users to create compelling narratives with their data. It offers a range of advanced templates, such as animated bar charts, scatter plot maps and story-based visualisations.  Flourish's interactivity allows users to dive deeper into the data by exploring visual elements. It is a great tool for storytelling and presenting data in a dynamic way, especially for projects or presentations. However, users need some understanding of interactivity design principles to maximise its potential. Flourish offers free access for public visualisations and paid options for private projects and additional features. |
| Tableau | **Sample answer:**  Tableau is renowned for its advanced data visualisation capabilities, with a broad range of charts and graphical tools that support complex analysis. Its drag-and-drop interface enables users to create detailed visualisations quickly, even with limited coding skills.  Tableau strikes a balance between ease of use and the power to create highly customised, interactive visualisations. It is particularly effective in visualising large, complex datasets.  Tableau can connect to numerous data sources seamlessly and offers high interactivity, enabling users to explore data deeply through dashboards. It is favoured in industries requiring robust analytics, though it may come at a higher cost compared to other tools. |

### Spreadsheets used to develop dashboards

Spreadsheets, such as Microsoft Excel or Google Sheets, are powerful tools for developing data dashboards. Dashboards provide an at-a-glance overview of key metrics and performance indicators and spreadsheets offer an accessible platform to build these dashboards without the need for advanced programming knowledge. They allow for the combination of various charts, graphs and data tables into one coherent view.

For example, a sales performance dashboard might include a bar chart showing monthly sales, a pie chart representing the product distribution and a table summarising key performance indicators (KPIs).

**Advantages of using spreadsheets for dashboards**

* **Accessibility**: most users are familiar with spreadsheets, reducing the learning curve.
* **Customisation**: spreadsheets offer extensive options for customising the appearance and structure of dashboards.
* **Data handling**: they handle large datasets through built-in formulas, pivot tables and charts.
* **Cost-effective**: spreadsheets are often included in standard office software suites, making them cost-effective solutions for smaller organisations.

### Presentation software used to present data analysis

Presentation software, such as Microsoft PowerPoint or Google Slides is commonly used to present data analysis to stakeholders. These tools allow users to combine text, images and data visualisations into cohesive presentations that tell a story. Effective use of presentation software enhances the clarity and impact of the data analysis by ensuring that the audience can easily follow along with the key findings.

For example, a financial analysis presentation might include slides featuring graphs showing quarterly earnings, along with a breakdown of budget allocations through pie charts.

**Tips for using presentation software to present data analysis**

* **Keep it simple**: avoid overloading slides with too much data or text. Focus on key points.
* **Use visual aids**: graphs, charts and infographics help make data more digestible.
* **Highlight insights**: clearly state what the data shows and what decisions should be made based on the analysis.
* **Engage the audience**: use animations or interactive elements to maintain engagement.

### Business analytics services, including ‘as a service’ products

As a class watch [Software as a Service (SaaS) Explained in 5 mins (5:22)](https://www.youtube.com/watch?v=20QUNgFIrK0).

Business analytics services, including Software as a Service (SaaS) products, offer organisations powerful tools for data analysis without requiring the infrastructure and maintenance of traditional on-premises software. These services, such as Google Analytics, IBM Watson Analytics and Salesforce Analytics Cloud, are hosted in the cloud and accessed via the web, offering scalability, flexibility and ease of use.

SaaS analytics platforms typically include tools for data integration, visualisation and predictive analytics. They can automate tasks like generating reports, running queries and even providing machine learning-driven insights.

**Advantages of ‘as a service’ products**

* **Scalability:** these products can scale with the size of the business, from small startups to large enterprises.
* **Lower costs**: with SaaS, there are no large upfront investments in infrastructure; businesses pay a subscription fee instead.
* **Automatic updates:** the service provider handles software updates and maintenance, reducing the workload for IT departments.
* **Accessibility:** since the tools are cloud-based, they can be accessed from anywhere with an internet connection.

### Custom software solutions

Custom software solutions are tailored to meet the specific needs of a business or organisation. Unlike off-the-shelf software, which is designed for a wide range of users, custom solutions are developed with particular use cases and business processes in mind. This approach can offer significant advantages in terms of efficiency, workflow integration and user experience.

For instance, a custom data analysis tool developed for a logistics company might include unique features such as integration with proprietary databases, real-time tracking and predictive models designed specifically for their shipping routes and inventory.

**Benefits of custom software solutions**

* **Tailored functionality**: custom software is designed to solve specific business problems, ensuring a perfect fit with organisational needs.
* **Scalability**: custom software can be built to grow alongside the business, accommodating new features and increased demand.
* **Competitive advantage**: having a unique software solution can provide a competitive edge by improving efficiency, performance and decision-making.
* **Integration**: custom solutions can be designed to seamlessly integrate with existing systems and processes.

## Interrogate data from a data visualisation

Interrogating data from a data visualisation involves critically analysing the graphical representation to draw meaningful insights.

The process goes beyond simply observing a chart or graph – it requires asking questions about the data source, understanding the context and exploring the variables presented. A thorough interrogation ensures that conclusions drawn from data visualisations are accurate, relevant and actionable.

Key questions to consider when interrogating a data visualisation include:

* What is the purpose of the visualisation?
* What trends, patterns or anomalies are visible?
* Are there any outliers or unusual data points?
* What assumptions are being made by the visualisation?

Effective interrogation involves breaking down the visual into its components (axes, labels, data points) and understanding the story behind the numbers.

**Activity 17:** interrogate data

As a class watch [How to spot a misleading graph (4:09)](https://www.youtube.com/watch?v=E91bGT9BjYk).

Facilitate a class discussion on how when graphs are used well, they can help us intuitively grasp complex data. While visual software has enabled more usage of graphs throughout all media, it has also made them easier to use in a careless or dishonest way – and as a result, there are plenty of ways graphs can mislead and outright manipulate.

### Interpreting what you see

When interpreting a data visualisation, it is essential to start with the basics:

1. **Understand the chart type**

Different chart types convey different messages. For example, line charts are ideal for showing trends over time, while bar charts are best for comparing discrete categories.

1. **Examine the axes**

The x-axis and y-axis (or more in complex visualisations) provide context for the data. Ensure that you understand what is being measured and the units used.

1. **Look for patterns**

Identify any clear trends, such as upward or downward movements, clusters, or gaps in the data. These patterns often tell the main story behind the visualisation.

1. **Contextualise the data**

Visualisations are often snapshots of larger datasets. Ask if the visualisation provides enough context. Does it tell the whole story, or are more variables or details needed?

By systematically reviewing the visualisation, you can ensure that your interpretation aligns with the intended message of the data.

### Aggregation

**Aggregation** in data visualisation refers to the process of combining multiple data points into a single summary value or statistic, such as a sum, average or count.

Aggregated data simplifies complex datasets, making it easier to identify patterns and trends. For example, a sales report may aggregate daily sales figures into monthly totals to provide a broader view of performance.

While aggregation helps make data more digestible, it can also obscure important details. For instance, averaging data can mask variability or outliers. Thus, while aggregated data provides a useful high-level overview, it is crucial to drill down into disaggregated data when necessary to avoid misinterpretation.

**Examples of aggregation**

* Average sales revenue over a year
* Total population of a region
* Median household income.

### Filtering

**Filtering** allows users to focus on specific subsets of data within a visualisation. By applying filters, viewers can narrow down the scope of the data being displayed, making it easier to identify relevant trends or relationships. For example, a sales dashboard might allow filtering by region, product category, or time period.

Filters are particularly useful when the dataset is large or complex. They enable more targeted analysis and help users draw insights that are specific to their needs. However, it is essential to understand the impact of filtering, as it can sometimes lead to biased conclusions if key data points are excluded or overlooked.

**Common use cases for filtering**

* Analysing customer feedback by demographic group
* Filtering sales data by product category or region
* Examining website traffic by device type (for example mobile versus desktop).

### The effect of outliers

**Outliers** are data points that deviate significantly from the rest of the dataset. In a visualisation, outliers can appear as isolated points far removed from the main data clusters. Outliers may indicate:

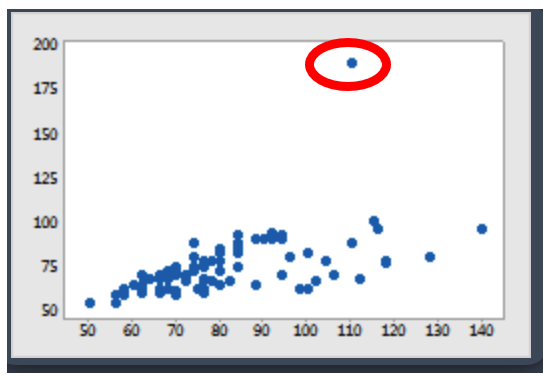
* **Errors**: data entry mistakes or faulty measurements
* **Anomalies**: unique but valid data points that represent unusual or extreme cases
* **Important insights**: outliers can reveal unexpected findings, such as sudden market shifts or extreme user behaviour.

When interpreting a visualisation, it is important to consider the effect of outliers on the overall data set. Outliers can skew averages, distort trends, or mislead viewers if not properly understood. Analysts should decide whether to exclude or further investigate outliers, based on their potential significance to the analysis.

**Examples of outliers**

* A single day of unusually high sales compared to an otherwise stable trend
* A drastic dip in website traffic caused by a technical issue.

Figure 13 – example of an outlier



### Reasoning

Reasoning in the context of data visualisation is the process of using logic and critical thinking to draw conclusions from the data presented. This involves interpreting the data accurately, considering the broader context and making inferences based on the evidence.

Good reasoning in data interpretation requires:

* **Identifying trends**: what story does the data tell? Are there clear upward or downward movements, seasonal patterns, or correlations between variables?
* **Assessing causality:** are the trends caused by a particular factor, or could there be other underlying causes?
* **Avoiding bias:** reasoning should be objective and based on evidence. Avoid jumping to conclusions based on preconceived notions.
* **Questioning assumptions:** does the data visualisation present the full picture, or is more information needed?

Effective reasoning helps in making data-driven decisions that are logical, sound and well-informed. For example, if sales data shows a drop in one quarter, reasoning should explore possible causes, such as seasonality, external market conditions, or internal business decisions, and not simply attribute the drop to a single factor without further analysis.

**Activity 18:** Netflix's usage of data visualisation to manage network traffic

Netflix, a global streaming service, utilises extensive data visualisations to monitor and manage its network traffic across different regions.

Given the diverse geographical locations of its users, Netflix's data visualisations help the company detect and analyse outliers in data traffic, which can indicate potential issues such as server overload, content delivery network (CDN) performance issues or unauthorised access attempts.

**Background**

Netflix experiences varying levels of demand in different regions based on factors like time of day, newly released content, or local internet speeds. The company's engineers use real-time visualisations to monitor these fluctuations. Outliers in this data are particularly significant as they can signal problems that may affect user experience, such as buffering or slow load times.

1. What is the importance of identifying outliers in data visualisations used by a company like Netflix?

|  |
| --- |
| **Sample answer:**  Identifying outliers is crucial for Netflix because they can indicate unusual activity that might impact service delivery. For example, a sudden spike in traffic could be due to a new show release or it might be a sign of a distributed denial of service (DDoS) attack. Recognising these outliers helps Netflix respond quickly to potential issues to maintain a smooth streaming experience for users. |

1. How can outliers affect the operational decisions at Netflix?

|  |
| --- |
| **Sample answer:**  Outliers can lead to significant operational decisions such as scaling server capacity to handle sudden increases in demand or adjusting the CDN strategies to optimise content delivery. For instance, if an outlier in traffic data shows a consistent increase in viewership in a certain region, Netflix might decide to upgrade infrastructure or increase bandwidth allocation in that area to improve streaming quality. |

1. Describe a scenario where ignoring outliers might be detrimental for Netflix.

|  |
| --- |
| **Sample answer:**  Ignoring outliers in traffic data might result in server overloads and poor streaming quality, especially during peak times or popular show releases. For example, if Netflix ignores a traffic spike that occurs every Friday night in Europe, they might not allocate enough resources, leading to buffering issues and a poor user experience, which could ultimately cause a loss of subscribers. |

1. What types of data visualisations might Netflix use to identify and analyse these outliers?

|  |
| --- |
| **Sample answer:**  Netflix likely uses heat maps to visualise global traffic data, time series graphs to track traffic changes over time and scatter plots to identify unusual points of data that deviate from normal patterns. These visualisations enable quick identification and response to anomalies in real-time. |

1. How does the interpretation of outliers contribute to improving the Netflix user experience?

|  |
| --- |
| **Sample answer:**  By interpreting outliers, Netflix can optimise its network to handle discrepancies in data traffic, ensuring that high demand does not compromise streaming quality. This proactive approach in managing anomalies allows Netflix to maintain consistent performance across all regions, thus enhancing user satisfaction and loyalty. |

1. What might be the consequences if Netflix's data visualisations are poorly designed and fail to highlight outliers effectively?

|  |
| --- |
| **Sample answer:**  If Netflix’s visualisations are poorly designed, critical outliers may go unnoticed. This oversight can lead to unresolved issues, such as inadequate server capacity or inefficient content delivery, negatively impacting user experience. Over time, this could reduce viewer satisfaction, increase frustration and potentially lead to a higher churn rate as subscribers leave Netflix for a more reliable service. |

# Designing for user experience

## Use graphic design tools to assist in the graphic development of a data visualisation

User experience (UX) design focuses on how a product or service feels to the end user. It involves understanding the needs, values, abilities, and limitations of users, as well as the goals and objectives of the product. When designing for UX, it is crucial to create products that are not only functional but also pleasurable and intuitive to use.

### Key principles of UX design

* Always design with the user in mind. Understand their needs, behaviours, and preferences through user research methods like surveys, interviews, and usability testing.
* Ensure your design can be used by a diverse range of users, including those with disabilities or different technological skills.
* Maintain consistency in the product to create familiarity for users.
* Provide clear feedback for user actions to build a sense of control and reassurance.
* Aim for intuitive interfaces that require minimal effort to navigate and understand.

**Graphic design tools for data visualisation**

In the realm of user experience design, data visualisation plays a crucial role in helping users understand and interpret data efficiently. The purpose is to present data in a way that makes insights easy to comprehend immediately. Effective data visualisation requires thoughtful design and the right tools.

**Selecting the right graphic design tool**

Several graphic design tools can assist in developing effective data visualisations, each offering unique features. Here are some common tools and their key applications in the context of data visualisation:

* **Canva**: a user-friendly, drag-and-drop tool suitable for beginners. Canva includes templates and pre-built graphics to help streamline the creation of data visualisations like bar charts, pie charts, and timelines.
* **Tableau**: primarily focused on data analysis and visualisation, Tableau allows users to create interactive and dynamic charts directly from data sources. It is an excellent tool for visualising large datasets.
* **Figma**: a collaborative web-based design tool that allows for the creation of dynamic prototypes, wireframes, and layouts. Figma’s integration with design systems helps maintain consistency across visualisations.
* **Microsoft Excel**: often underrated, Excel provides basic but functional charting capabilities, allowing users to quickly generate various graphs and charts with built-in tools.

**Activity 19:** build practical skills

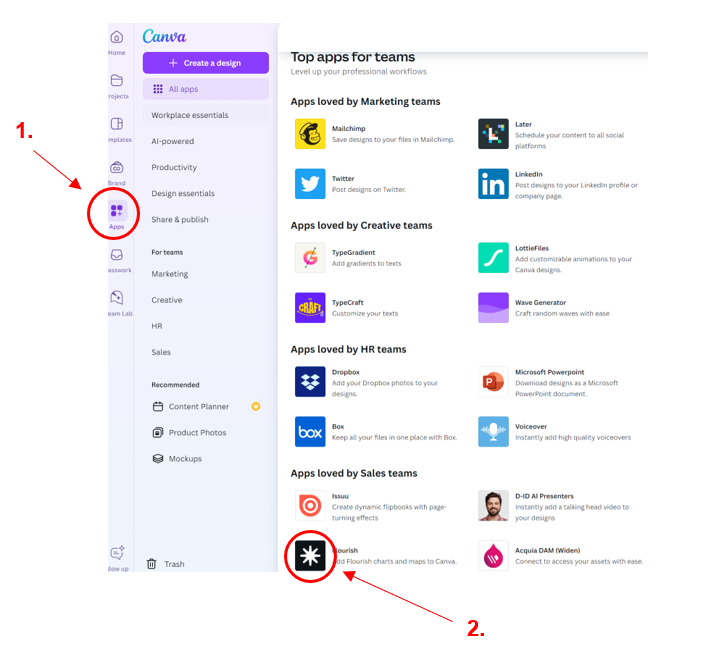
Students choose one of the following software (Canva, Microsoft Excel or Power BI) or research other software and complete tutorials to develop skills to tell a story and apply the principles of UX. This activity leads to the development of skills to create a data story using their chosen data and software for Assessment task 2.

**Canva/Flourish**

As a class watch [Storytelling with Data: Infographics and data visualization | Canva for Journalists (10:04)](https://www.youtube.com/watch?v=SuFcF2uS1zY). Students can complete the activities in the Canva [Workbook](https://canva.me/data-workbook-desktop) – Charts and Tables -Journalist Essentials.

To use advanced data visualisation tools in Canva you need to add Flourish to your apps within Canva. To finalise adding Flourish you will need to create a free account with your educational email.

Figure 14 – Canva – browse to **Apps** section and select **Flourish**



*Screenshots of Canva reproduced for non-commercial, educational purposes under the*[*Canva for Education*](https://aus01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.canva.com%2Fpolicies%2Fedu-additional-terms%2F&data=05%7C02%7Celizabeth.rose5%40det.nsw.edu.au%7Ce985d6f4d6564f03a12208dd0a01071a%7C05a0e69a418a47c19c259387261bf991%7C0%7C0%7C638677721733578221%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=Tpqnl8FfHW2Q47lWwydzaFV7dBcoobwItJ8a1ZyTndg%3D&reserved=0)*terms.*

Figure 15 – add Flourish in apps

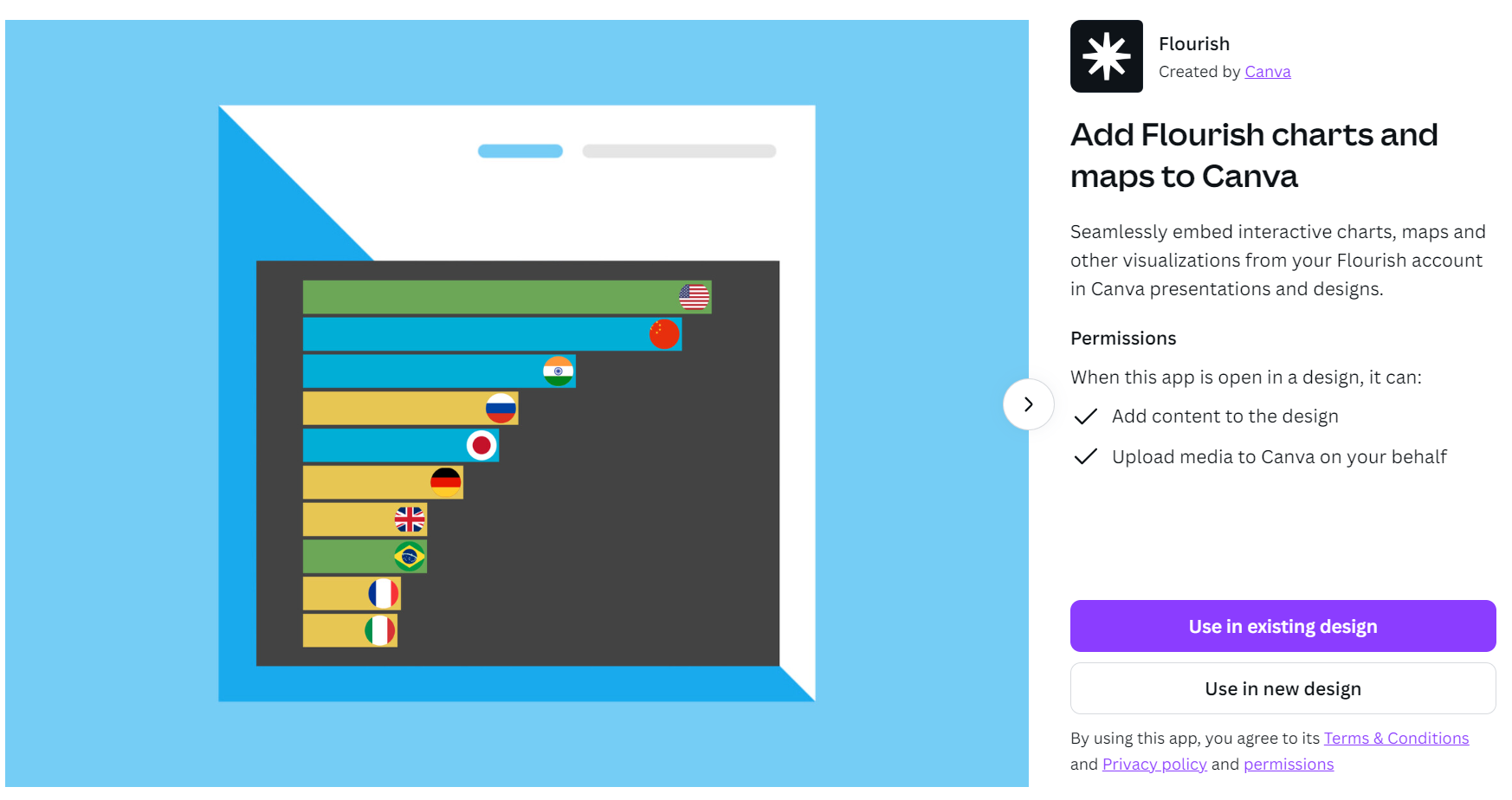
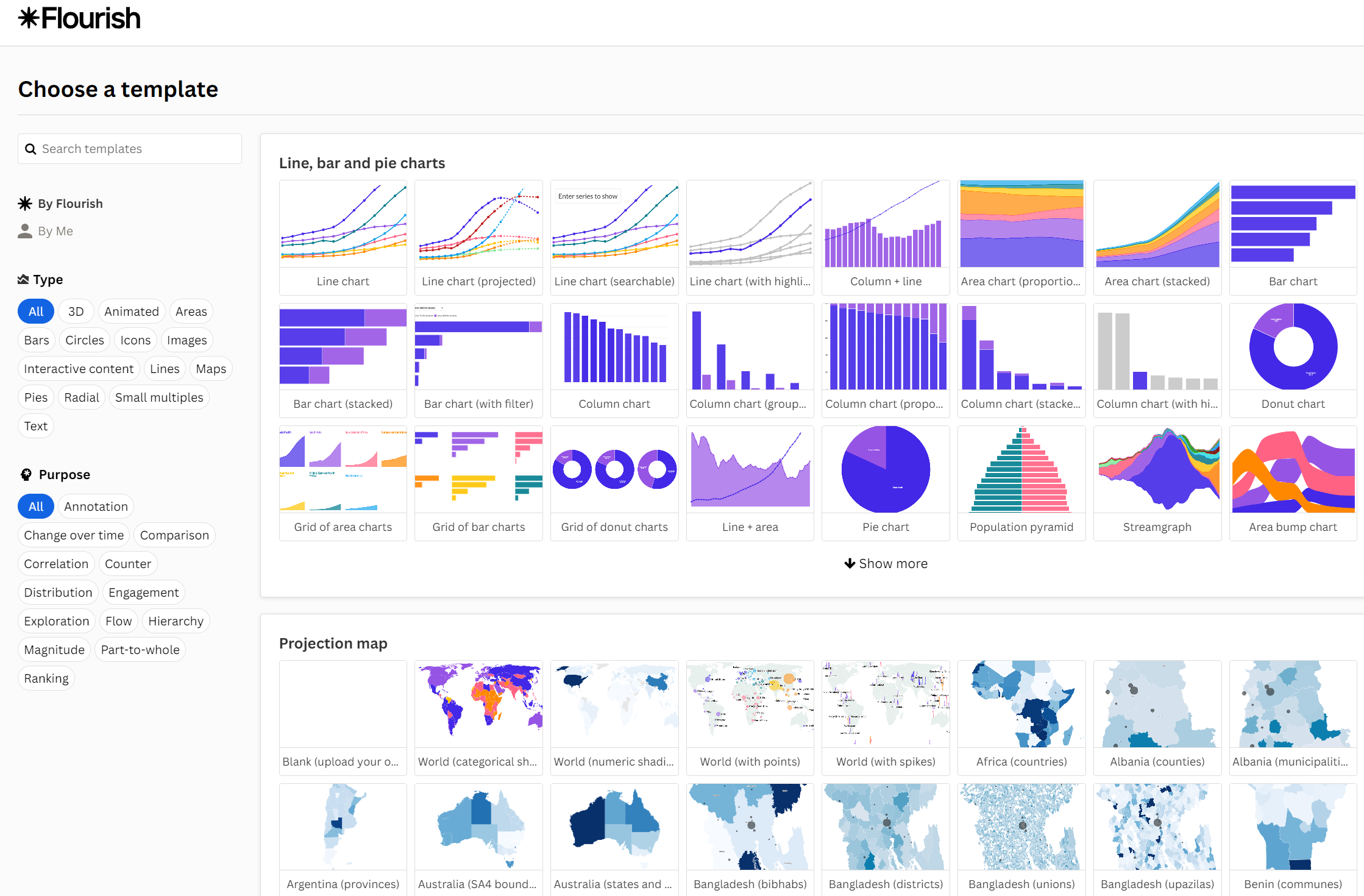
*****Screenshots of Canva reproduced for non-commercial, educational purposes under the*[*Canva for Education*](https://aus01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.canva.com%2Fpolicies%2Fedu-additional-terms%2F&data=05%7C02%7Celizabeth.rose5%40det.nsw.edu.au%7Ce985d6f4d6564f03a12208dd0a01071a%7C05a0e69a418a47c19c259387261bf991%7C0%7C0%7C638677721733578221%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=Tpqnl8FfHW2Q47lWwydzaFV7dBcoobwItJ8a1ZyTndg%3D&reserved=0)*terms.*

Figure 16 – choose a template

****

*Screenshots of Canva reproduced for non-commercial, educational purposes under the*[*Canva for Education*](https://aus01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.canva.com%2Fpolicies%2Fedu-additional-terms%2F&data=05%7C02%7Celizabeth.rose5%40det.nsw.edu.au%7Ce985d6f4d6564f03a12208dd0a01071a%7C05a0e69a418a47c19c259387261bf991%7C0%7C0%7C638677721733578221%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=Tpqnl8FfHW2Q47lWwydzaFV7dBcoobwItJ8a1ZyTndg%3D&reserved=0)*terms.*

**Microsoft Excel**

Students may have already mastered the use of Microsoft Excel in earlier learning. The [YouTube playlist of Excel Tutorials by EasyClick Academy](https://www.youtube.com/watch?v=hElkEzAjd5o&list=PLAE-5Y383viiGxNskXM443Ir1mvaSMJqU) has short, accessible tutorials for creating a range of charts and can be referenced as needed.

**Power BI**

As a class watch [From Excel to Power BI in 12 Minutes (Beginner Essentials) (12:42)](https://www.youtube.com/watch?v=gOs7EC-FebE)

How can you find, collect, and clean data from different sources? Power BI is a tool for making sense of your data. Completing the Microsoft learning course [Get data with Power BI Desktop (1 hr 15 min)](https://learn.microsoft.com/en-gb/training/modules/get-data-power-bi/) you will learn tricks to make data-gathering easier.

## Explain how user experience (UX) influences the development of effective data visualisations

User experience (UX) is a critical component in the development of effective data visualisations. It refers to how easy, intuitive and enjoyable the experience is for users when interacting with a visualisation. A well-designed data visualisation should not only present data clearly but also ensure that the user can interpret and engage with it effortlessly. When UX principles are applied, users can more easily extract insights, which makes the visualisation more effective in conveying its intended message.

Several factors influence the development of data visualisations from a UX perspective, including relevance to the audience, ease of interpretation, customisation options and the ability to provide live analysis.

### Relevance to the audience

The success of a data visualisation depends on how relevant it is to the audience's needs and interests. A well-designed visualisation considers who the audience is and what they are looking to learn or understand from the data. For instance, a visualisation meant for financial analysts might include detailed numerical data, while one aimed at a general audience would focus on high-level trends and key takeaways.

**Key UX considerations for audience relevance**

* **User personas**: developers of data visualisations should consider the specific personas of their audience, such as their level of expertise, their role and the context in which they will be using the visualisation.
* **Tailored insights**: a visualisation should provide insights that are directly relevant to the audience’s decision-making process. For example, executives may need quick summaries of key metrics, while data scientists might require in-depth breakdowns.

Ensuring the relevance of data to the audience enhances engagement and makes the visualisation much more effective in fulfilling its purpose.

### Audience interpretation

A data visualisation is only as good as the audience's ability to understand and interpret it. Clear, intuitive designs that guide users through the data make interpretation easier and faster. Complex or cluttered visualisations can lead to confusion, misinterpretation or an inability to extract key insights. By ensuring that the visualisation is intuitive and clear, the audience will be able to interpret the information accurately, enhancing their user experience.

**Key UX factors for enhancing audience interpretation**

* **Clarity**: charts, graphs and other visuals should be easy to understand at first glance. Avoid overloading visualisations with too much information and instead focus on simplicity and readability.
* **Effective labelling**: axes, data points and legends must be clearly labelled. Ambiguous labels or unclear legends can lead to misinterpretation of the data.
* **Data presentation**: the use of colour, size and positioning of elements can greatly influence how data is perceived. For example, colours should be used consistently to represent certain variables and important data points should be emphasised through size or placement.

### Customisation

Customisation features in data visualisations allow users to adapt the visual to their specific needs, making it more useful and personalised. Effective UX design considers that different users may want to interact with the data in different ways, depending on their objectives and preferences. Customisation features can include filtering data, changing chart types, or selecting different time ranges.

**Key UX customisation features**

* **Filtering options**: allowing users to filter data based on criteria such as region, time, or category helps them focus on the most relevant information.
* **Interactive elements**: interactive data visualisations that let users drill down into specific data points or hover over sections to reveal more details improve engagement and provide deeper insights.
* **Adjustable visuals**: users may want to switch between different types of charts (for example from a bar chart to a line chart) to better understand the data. Offering this flexibility ensures that users can view the data in the format that best suits their needs.

Customisation enhances the UX by making the data visualisation adaptable to various user needs, leading to more effective data exploration and analysis.

### Live analysis

Live analysis refers to the ability of a data visualisation to update in real time based on the latest data. This feature is particularly useful for users who need to make decisions based on current information, such as stock market analysts, business executives monitoring performance metrics, or emergency responders tracking real-time events.

**Key UX considerations for live analysis**

* **Real-time data updates:** the visualisation should automatically refresh with the most up-to-date data without the need for user intervention. This ensures that users are always working with the most relevant and accurate information.
* **Performance:** live data visualisations must perform efficiently without causing delays or lag, which can lead to frustration and a negative user experience.
* **Alerts and notifications:** for visualisations used in critical contexts, such as financial markets or disaster response, UX considerations include providing alerts or notifications when certain thresholds are reached, or significant changes occur in the data.

**Activity 20:** how Spotify uses UX to enhance data visualisations in Wrapped

Spotify Wrapped is a popular year-end feature that gives users a visual summary of their music listening habits for the year. Each year, Spotify presents this data through a series of fun, interactive and personalised data visualisations, summarising everything from users' most-played songs to the total time spent listening to music.

Spotify’s Wrapped feature is an excellent example of how user experience (UX) can influence the development of effective data visualisations. The service uses engaging graphics, clear labels and intuitive design to ensure users can easily interpret their data while feeling connected to their listening habits. The data is presented in a fun and creative way, making it relevant to users while also allowing for customisation, such as the ability to share highlights on social media.

1. Why is it important for data visualisations to be relevant to the audience in a feature like Spotify Wrapped?

|  |
| --- |
| **Sample answer:**  It is important for data visualisations to be relevant to the audience because users are more likely to engage with information that is meaningful to them. In Spotify Wrapped, the data visualisation is based on each user's unique listening habits, which makes it highly personal and engaging. By seeing their favourite songs, artists and genres, users are more connected to the data, making the experience fun and memorable. |

1. How does Spotify Wrapped make the data easy to interpret for users?

|  |
| --- |
| **Sample answer:**  Spotify Wrapped makes data easy to interpret by using clear labels, simple charts and bold colours to highlight key information. For example, instead of overwhelming users with raw numbers, Spotify uses fun infographics, like pie charts and bar graphs, to show data like top genres or minutes listened. The design is clean and the interface is intuitive, so users can understand their stats without any confusion. |

1. What role does customisation play in improving user experience in Spotify Wrapped?

|  |
| --- |
| **Sample answer:**  Customisation plays a significant role in improving the user experience by allowing users to interact with their data in a way that feels personal. In Spotify Wrapped, users can see personalised results tailored to their listening habits, and they can also share these results with friends on social media. This personal touch, along with the ability to control how they view and share their data, makes the experience feel more unique and relevant to each individual. |

1. How does live analysis or real-time updates enhance the experience of Spotify Wrapped throughout the year?

|  |
| --- |
| **Sample answer:**  While Spotify Wrapped itself is an annual feature, the ability to see real-time listening data through daily updates in the Spotify app adds to the overall user experience. Users can track their favourite songs and artists throughout the year, which keeps them engaged and excited for the final Wrapped report at the end of the year. Real-time updates also allow users to stay aware of trends in their listening habits, creating anticipation for what their final Wrapped will look like. |

1. What impact do outliers or unusual data points have on the visualisation and how does Spotify Wrapped handle this?

|  |
| --- |
| **Sample answer:**  Outliers or unusual data points, such as a user listening to one song on repeat hundreds of times, can skew the data. Spotify Wrapped handles this by showcasing these outliers in a fun, light-hearted way. For example, if a user listens to one particular song much more than others, Spotify might highlight it as their ‘most obsessive’ track. This makes the outliers feel special rather than problematic, enhancing the overall user experience. |

1. Why do you think Spotify Wrapped is a successful example of using UX principles to create effective data visualisations?

|  |
| --- |
| **Sample answer:**  Spotify Wrapped is successful because it uses UX principles like relevance, ease of interpretation, customisation and interactivity to create an engaging, personalised experience for users. The visualisations are designed to be fun and easy to understand, which makes users excited to interact with them. The data is presented in a way that feels personal and relevant. |

## Develop and implement criteria for evaluating the effectiveness of user experiences

Data visualisation is a powerful tool that transforms raw data into visual formats like charts, graphs and dashboards, allowing users to interpret and understand complex information quickly. However, just like with any user experience (UX), the effectiveness of a data visualisation depends on how well it meets users' needs. To create meaningful and accessible visualisations, it is important to develop and implement criteria for evaluating their effectiveness.

Evaluating the effectiveness of data visualisations is essential for ensuring that users can understand and engage with the data in meaningful ways. By developing criteria focused on clarity, usability, accuracy, accessibility, efficiency, aesthetics and context, you can systematically assess and improve any data visualisation. Through user testing, heuristic evaluation and data analytics, you can gather the insights needed to refine visualisations and ensure they meet the needs of their audience.

**What is heuristic evaluation?**

A **heuristic** is a practical problem-solving approach or rule of thumb that helps people make decisions or solve problems efficiently when a perfect or optimal solution may not be immediately available. Heuristics are often used when facing complex problems where time and resources are limited and they allow people to arrive at a satisfactory solution quickly without needing to evaluate every possible option.

In the context of user experience (UX) and data visualisation, **heuristics** refer to widely accepted best practices or principles that guide the design and evaluation of user interfaces or visualisations. These principles are based on experience and general rules that often lead to good results, even if they don’t guarantee a perfect solution.

One well-known example is Jakob Nielsen's **10 Usability Heuristics for User Interface Design**, which include guidelines such as:

* **Visibility of system status**: users should always be informed of what is happening within a system.
* **Match between system and the real world**: the design should use concepts, language and workflows familiar to users.
* **User control and freedom**: users should be able to undo and redo actions easily.

In data visualisation, heuristic principles help ensure that visualisations are easy to understand and interpret, such as:

* avoiding **chartjunk** (unnecessary elements in the visualisation that don’t add value)
* using **consistent scales** and axes for accurate comparisons
* ensuring **appropriate colour usage** for clarity and accessibility.

Heuristics help designers and evaluators focus on key aspects of usability and design without needing to perform exhaustive, time-consuming testing for every possible scenario.

The following 5 steps will help you develop and implement criteria for evaluating the effectiveness of user experiences.

**Step 1: understanding user needs in data visualisation**

Before evaluating a data visualisation, you must first understand the users and their goals. Consider the following:

* **Who is the target audience?** Are they data experts, like statisticians or casual users who need to quickly grasp key trends?
* **What are the key insights users are seeking?** Do they want an overview or detailed, in-depth data?
* **How will the visualisation be used?** Will users be interacting with it, like in a dashboard, or simply viewing it in a static report?

Knowing these factors helps tailor your evaluation criteria to the specific needs of your audience. For example, a data dashboard for business managers will prioritise ease of use and quick insights, while a scientific report may require more detailed and accurate visualisations.

**Step 2: developing evaluation criteria for data visualisation**

To evaluate the effectiveness of data visualisations, you need clear criteria. Below are key aspects to consider:

**Clarity**

* Is the data easy to understand? The visualisation should convey the message clearly without overwhelming the user.
* Is the labelling accurate and appropriate? Axes, legends and data points must be labelled clearly and consistently.
* Does the visual match the data’s purpose? For example, pie charts work well for showing proportions, while line charts are better for trends over time.

**Usability**

* Is the visualisation easy to navigate and interpret? Users should be able to extract key insights without confusion. If it is interactive, are users able to explore the data smoothly?
* Is it responsive? For digital visualisations, ensure that they adjust properly on different devices and screen sizes.

**Accuracy**

* Does the visualisation accurately represent the data? Distorted scales, misaligned axes or omitted data can lead to misinterpretation. Make sure the visualisation faithfully represents the data’s reality.
* Is the data up-to-date and reliable? Outdated or incorrect data can compromise the usefulness of the visualisation.

**Accessibility**

* Is the visualisation accessible to all users, including those with disabilities? Features like colourblind-friendly palettes, text alternatives for visual elements and the ability to navigate with a keyboard are critical for accessibility.
* Is the contrast between elements sufficient? Ensure that all text, lines and colours are easily distinguishable, especially for users with visual impairments.

**Efficiency**

* How quickly can users gain insights? Users should be able to extract key takeaways without spending too much time decoding the visualisation.
* Is the data visualisation free of unnecessary clutter? Visual elements like unnecessary gridlines, excessive labels, or irrelevant data points should be removed to make the visualisation more streamlined.

**Aesthetics**

* Is the visualisation visually appealing and engaging? The design should be aesthetically pleasing without sacrificing functionality. Use appropriate colours, fonts and shapes that draw attention to the most important data.
* Are elements well-organised? The layout should guide the user’s eyes smoothly through the data without confusion.

**Context and relevance**

* Does the visualisation provide enough context for users to understand it? Users should be able to see how the data fits into a larger picture. Add labels, titles and descriptions where needed to clarify the message.
* Is the visualisation tailored to the audience’s knowledge level? The complexity of the visualisation should match the users’ understanding of the data.

**Step 3: implementing the evaluation criteria**

Once you have developed your criteria, it is time to put them into action. Here is how you can evaluate a data visualisation:

1. **User testing**

Bring in real users to test the visualisation. This could be done through:

* Observation: watch how users interact with the visualisation. Can they extract the necessary insights quickly and accurately?
* Surveys and feedback: ask users to rate how easy the visualisation was to understand and whether they found it useful and engaging.
* Task-based testing: set specific questions or tasks related to the data and see how easily users can answer or complete them using the visualisation.

1. **Heuristic evaluation**

Apply best practices and heuristics for data visualisation:

* avoid ‘chartjunk’ (unnecessary decoration that doesn’t improve the understanding of the data).
* use colours effectively (for example avoiding too many or too few colours, ensuring colourblind accessibility).
* ensure labels and scales are logical and consistent.

1. **Data analytics**

For interactive visualisations, use data analytics tools (like Google Analytics for websites or built-in dashboard tracking) to monitor:

* how users are interacting with the visualisation
* which parts are frequently viewed or clicked
* where users might struggle or abandon the visualisation.

**Step 4: measuring the effectiveness of data visualisations**

Now that you’ve gathered feedback and data, it is time to evaluate the results. Use the following methods to measure effectiveness:

1. **Scoring based on criteria**

Rate the visualisation on each of the criteria you’ve developed (for example clarity, usability, accuracy and so on). You could create a scoring system (for example 1 to 10 for each criterion), then average the scores for an overall rating.

1. **Quantitative measures**

Use objective data to evaluate effectiveness.

* Time on task: how long does it take users to interpret the visualisation?
* Error rate: do users misunderstand or misinterpret the data?
* Engagement: how frequently do users interact with the visualisation (for example clicks or hovers)?

1. **Qualitative feedback**

Analyse feedback from user testing:

* Are users satisfied with the visualisation?
* What suggestions do they have for improvement?

1. **Comparative analysis**

Compare the current visualisation with past versions or competitors:

* Have you made improvements?
* Does your visualisation stand out in terms of usability or aesthetics?

**Step 5: improving data visualisations based on evaluation**

Once you’ve evaluated your data visualisation, the final step is to make improvements based on your findings. For instance:

* simplify the visualisation if users found it too complex
* improve labelling and annotations to enhance clarity
* adjust colour schemes to make the visualisation more accessible.

Remember, just like with any user experience, data visualisations should be continuously refined to ensure they remain relevant, accurate and effective.

****In groups, discuss how you will apply this learning to Assessment task 2 when developing your own data visualisation and criteria to evaluate effectiveness.

**Teacher note:** further reading about how Jakob Nielsen created the [10 Usability Heuristics](https://www.uxtigers.com/post/usability-heuristics-history)may assist students understanding.

## Investigate the impact of emerging hardware and software technologies on user interface (UI) and UX design and development

Emerging hardware and software technologies such as virtual reality (VR), artificial intelligence (AI), cloud computing and touchscreens are transforming how user interfaces are designed and how users experience products. These technologies enhance UI/UX by making interactions more natural, responsive and personalised. However, designers must also navigate challenges like ensuring usability across devices, preventing cognitive overload and balancing personalisation with user control.

By investigating these technologies and their impacts, you can gain a deeper understanding of the evolving field of UI/UX design and how it will continue to change in the future.

### Introduction to UI and UX

User interface (UI) refers to the visual elements of a product, such as buttons, menus and icons, that users interact with directly. User experience (UX) is a broader concept that encompasses the overall feeling and experience users have while using a product, including usability, functionality and satisfaction.

As hardware and software technologies continue to evolve, they have a profound impact on how UI and UX are designed and developed.

#### Emerging hardware technologies and their impact on UI/UX

**Touchscreens and gestural interfaces**

* Impact on UI: with touchscreens, UI design has moved beyond the traditional mouse and keyboard setup. Buttons and menus need to be larger and spaced out to be finger friendly. Swipe, pinch and drag gestures become part of the interface, demanding intuitive design.
* Impact on UX: the transition to touchscreens enhances the user experience by making devices more interactive and natural to use. However, poor implementation of gesture controls can lead to frustration if gestures are not responsive or intuitive.

**Voice-controlled interfaces**

* Impact on UI: voice-controlled systems like Siri or Alexa reduce the need for traditional UIs like buttons and menus. Designers must focus on crafting effective voice prompts, feedback systems and integrating natural language processing.
* Impact on UX: voice interfaces improve accessibility and convenience, enabling hands-free operation. They can make interactions faster and more efficient, though they can also cause frustration if the system fails to understand commands correctly.

**Virtual reality (VR) and Augmented reality (AR)**

* Impact on UI: VR and AR require entirely new user interfaces that are immersive and three-dimensional. Traditional buttons and menus may be replaced by spatial controls, where users interact with objects in a virtual or augmented environment.
* Impact on UX: VR and AR provide highly immersive experiences but can also cause issues like motion sickness or cognitive overload if not designed carefully. The user experience depends heavily on the fluidity of interactions and the level of immersion.

**Wearable devices**

* Impact on UI: devices like smartwatches and fitness trackers offer limited screen real estate, requiring highly simplified UIs. Designers need to focus on minimalist designs with clear, essential information presented immediately.
* Impact on UX: wearable devices emphasise convenience and instant access to information, but poor design can lead to user frustration due to the small interface size. UX also needs to consider how these devices integrate with other systems, such as smartphones or apps.

#### Emerging software technologies and their impact on UI/UX

**Artificial intelligence (AI) and machine learning (ML)**

* Impact on UI: AI is increasingly being used to create personalised user interfaces that adapt to user behaviour. For example, recommendation systems on streaming platforms or e-commerce sites adjust UI elements to align with user preferences.
* Impact on UX: AI-driven personalisation improves the overall experience by anticipating user needs and delivering more relevant content or features. However, too much personalisation can feel invasive or lead to ‘filter bubbles’, limiting a user's exploration of new content.

**Cloud computing**

* Impact on UI: cloud-based applications allow for UIs that are accessible from any device with an internet connection, driving the design of responsive UIs that work seamlessly across platforms and screen sizes.
* Impact on UX: cloud computing enhances UX by enabling real-time collaboration, automatic backups and access to data anywhere, providing users with a more flexible and convenient experience.

**Progressive web apps (PWAs)**

* Impact on UI: PWAs blur the line between web and mobile apps, allowing UIs to function similarly across devices. Designers must create UI elements that are responsive, lightweight and can work offline.
* Impact on UX: PWAs improve UX by offering fast, reliable and engaging experiences on any device without requiring downloads. They combine the best of web and mobile app design, increasing accessibility and engagement.

**Natural language processing (NLP)**

* Impact on UI: NLP allows for more conversational interfaces, such as chatbots and voice assistants. UIs can become simpler, focusing more on dialogue-driven interactions rather than graphical buttons and menus.
* Impact on UX: NLP enhances UX by allowing users to interact with systems in a more natural and intuitive way. However, poor understanding of context or complex language by NLP systems can diminish the user experience.

**Activity 21:** investigating the impact of emerging technologies on UI/UX

Research and analyse how emerging hardware and software technologies impact the design and development of user interfaces (UI) and user experiences (UX).

1. **Research**

Choose one emerging hardware technology (for example VR, touchscreens, wearables) and one emerging software technology (for example AI, cloud computing, NLP).

Investigate how each technology impacts the design of user interfaces and the overall user experience.

1. **Analysis**

For each technology, answer the following questions:

* How has the technology influenced UI design? Provide specific examples.
* How has it impacted UX, either positively or negatively? Provide specific examples.
* What are some challenges or limitations in designing for this technology?

1. **Group discussion**

****Share your findings with your group or class and compare insights. Identify trends across different technologies and discuss how designers might overcome the challenges associated with these technologies.

|  |
| --- |
| **Sample answer:**  **Example 1: Virtual reality (VR)**   * **UI Impact**: VR has transformed traditional 2D UIs into 3D immersive environments where users interact with spatial elements like virtual objects, or menus that respond to hand movements or controller inputs. For example, in VR gaming, menus can appear as holographic displays that users can interact with by reaching out and selecting objects in the virtual space. * **UX Impact**: VR provides a highly immersive experience that allows users to feel present in a simulated environment. This enhances UX in areas like education and entertainment, offering unique experiences. However, users may experience motion sickness or discomfort due to poorly designed VR interfaces, which can break immersion. * **Challenges**: designing intuitive interactions in 3D space can be challenging because users may not always understand how to interact with virtual objects. Additionally, ensuring compatibility across various VR headsets is difficult, as performance and controls can vary between devices.   **Example 2: Artificial intelligence (AI)**   * **UI impact:** AI-driven personalisation has led to dynamic UIs that adapt based on user preferences. For instance, Netflix’s UI changes the layout and recommendations based on the user’s watch history and preferences, making it easier to find relevant content. * **UX impact:** AI enhances the user experience by providing personalised recommendations, improving the relevance of content and increasing user satisfaction. However, if the personalisation is too aggressive, it can feel invasive or limit the user's exploration of diverse content. * **Challenges:** designers must ensure that AI doesn’t over-personalise or make assumptions about user behaviour that could lead to a frustrating or limited experience. Balancing user control with AI-driven suggestions is key. |

# Creating data visualisations

## Research, source, organise and store data appropriate for a data visualisation

Creating effective data visualisations requires not only technical skills but also the ability to find and manage accurate, relevant data. Learning how to research, source, organise and store data that is suitable for creating meaningful visualisations is important. This process involves several key steps: identifying the right data, making sure it is trustworthy, organising it effectively and storing it safely for future use.

1. **Researching data**

The first step in creating a data visualisation is finding the right data. This involves:

**Defining the purpose of your visualisation**

Before you start looking for data, think about the story you want to tell or the question you want to answer. This helps you know what kind of data you need. For example, if you want to create a visualisation showing trends in climate change, you’ll need data on temperature, emissions or weather patterns over time.

**Types of data**

There are 2 main types of data:

* Qualitative data: this is descriptive data that can’t easily be measured, such as customer feedback or interview responses.
* Quantitative data: this is numerical data that can be measured, such as sales numbers, population statistics or exam scores.

**Sources of data**

Data can come from a variety of sources, depending on your needs:

* Primary data: data you collect yourself through surveys, experiments, or observations.
* Secondary data: data that has already been collected by someone else, such as government reports, research papers, or data from online databases.

1. **Sourcing data**

Once you know what kind of data you need, it is time to find it.

**Reliable sources**

Ensure that you get data from reliable sources to maintain accuracy. Some trusted sources include:

* government websites (for example Australian Bureau of Statistics)
* research institutions and universities
* industry reports
* open data portals.

**Data formats**

Data can come in different formats, and it is important to understand how to work with them:

* CSV (Comma-Separated Values): this is one of the most common formats for storing structured data. It is easy to import into spreadsheet tools like Microsoft Excel or Google Sheets.
* JSON (JavaScript Object Notation): a format often used for web-based data. It is popular in programming and allows data to be stored in a hierarchical structure.
* APIs (Application Programming Interfaces): some websites provide data via APIs, which allow you to pull in live data directly from a web service. This is useful for dynamic visualisations.

1. **Organising data**

Before you can use data for visualisation, it is important to organise it properly.

**Clean your data**

Data often comes with errors or inconsistencies, such as missing values, duplicates or incorrect formatting. Cleaning your data involves:

* Removing duplicates: ensure that you don’t have multiple copies of the same data point.
* Handling missing data: you might need to fill in missing values or decide to exclude incomplete records.
* Correcting errors: look for typos, incorrect data types (for example dates written as text), or outliers that don’t make sense (for example an age of 150 years in a demographic dataset).

**Organise by categories**

Once your data is clean, it is important to structure it logically. If you are working with a spreadsheet, this means organising data into:

* Rows: each row represents an individual data point (for example one student, one day’s weather).
* Columns: each column represents a different variable (for example date, temperature, sales amount).

**Use consistent labels**

Make sure that column names and categories are consistent throughout your dataset to avoid confusion when visualising. For example, if you have data from multiple regions, make sure you use the same label (for example ‘NSW’ versus ‘New South Wales’).

1. **Storing data**

Storing data properly is essential for ensuring that it is accessible and protected for future use.

**File naming conventions**

Always use clear and consistent names for your data files. For example, a good file name might be ‘Climate\_Data\_Australia\_2023.csv’, which tells you what the data is about, where it comes from, and when it was collected. Avoid vague names like ‘data1.csv’ that don’t provide any context.

**Storage locations**

Store your data in a safe and accessible location:

* Cloud storage: use platforms like Google Drive, Dropbox, or OneDrive for easy access and sharing.
* Local storage: save files on your computer, but always create backups to avoid losing data if your device fails.
* Database storage: if you are working with a large dataset, consider using a database to manage and store data more efficiently.

**Data security**

Make sure your data is secure, especially if it contains sensitive information. Use password protection or encryption to secure your files and ensure only authorised users can access the data.

1. **Final considerations**

**Keep data up to date**

Data can become outdated quickly, especially in fast-moving fields like technology or economics. Always check that your data is current and accurate before using it for a visualisation.

**Document your process**

Keep a record of where your data comes from, how it was collected, and any cleaning or processing steps you performed. This helps ensure transparency and makes it easier to explain your work to others.

**Activity 22:** how to approach data

Imagine you are creating a data visualisation to show the rise of renewable energy usage in Australia. Describe the processes of how you would approach research, source, organise and store data appropriate for a data visualisation.

|  |
| --- |
| **Sample answer:**   * Research: decide that you need data on renewable energy production over the last 10 years. * Source: find a reliable dataset from the Australian government’s energy statistics website. * Organise: clean the data by removing any rows with missing information, ensure all years are in the same format (YYYY), and sort the data by year. * Store: save the cleaned dataset to Google Drive as ‘Australia\_Renewable\_Energy\_2013-2023.csv’ with a backup on your computer, ensuring the file is accessible when you begin building your visualisation. |

## Design and develop a data visualisation for a specific scenario to represent trends, patterns and relationships, and illustrate predictive analysis incorporating big data

**Teacher note:** students commence Assessment task 2 and are guided with one-to-one support in class time. Students use the steps to success in accompaniment to the following steps.

**How to design and develop a data visualisation for a specific scenario (Assessment task 2)**

**Step 1:** understand the purpose of data visualisation.

Data visualisation is more than just creating charts or graphs; it is about communicating complex information in a way that is clear, insightful and impactful. When developing a visualisation, you aim to represent data trends, patterns, and relationships effectively, enabling your audience to grasp critical insights at a glance. Think of your visualisation as a storytelling tool, guiding your audience through the narrative that your data presents.

**Step 2:** identify and define your scenario.

Choose a specific scenario that aligns with your task requirements. For instance, you could focus on an enterprise, social, or ethical issue. It is essential to clearly define:

* Who is your audience?
* What do you want them to learn or understand from your visualisation?
* What problem or question are you addressing?

Having clear answers to these questions will shape how you select, analyse and present your data.

**Step 3:** select relevant data and tools.

To ensure your visualisation is both meaningful and accurate, research, source, and organise data that is relevant to your chosen scenario. Look for datasets that are comprehensive, reliable and reflective of the trends or relationships you aim to highlight. Remember to:

* use appropriate tools like Microsoft Excel, Canva (Flourish) or Power BI for your data analysis and visualisation
* ensure your dataset is of adequate size to reflect the trends or patterns clearly.

**Step 4:** analyse your data for patterns, trends and relationships.

Once you have your data, use analytical techniques to identify key patterns or relationships. This is where you compare datasets and look for emerging trends. Consider asking questions like:

* What trends do I observe over time?
* Are there significant correlations between variables?
* What predictions can I make based on these trends?

Make use of statistical and analytical tools within your chosen software to validate your observations and predictions. Predictive analysis involves using historical data to forecast future trends. For instance, if you are examining sales data, you might use predictive algorithms to estimate future sales.

**Step 5:** create your visualisation with clarity and purpose.

Design your visualisation by applying best practices:

* Choose the right chart type: use bar charts to compare categories, line charts for time-series trends or scatter plots to show correlations.
* Avoid visual clutter: keep your design simple and focused. Avoid unnecessary 3D effects, overly bright colours or complex layouts.
* Add clear labels and legends: ensure that each part of your visualisation is easy to interpret.

Incorporating big data means acknowledging that your datasets may be extensive and complex. Break your visualisation into multiple views if necessary to avoid overwhelming the audience.

**Step 6:** reflect on your visualisation’s effectiveness.

After creating your visualisation, evaluate its clarity and effectiveness:

* Does it tell a compelling story?
* Does it accurately represent the trends or relationships identified?
* Is the key message clear without needing additional explanations?

Seek feedback from peers to refine your visualisation and improve its clarity and accuracy.

## Investigate and implement methods to maintain data security

Data visualisation involves transforming raw data into visual formats like graphs, charts and dashboards to help people understand complex information easily. However, when working with data, especially sensitive or confidential information, it is essential to ensure that the data is secure. This section will discuss methods to maintain data security, focusing on cybersecurity and data backup.

Maintaining data security is essential when working with data visualisations. By implementing strong cybersecurity measures and ensuring regular data backups, organisations and individuals can protect sensitive information from cyber threats and data loss. These practices not only secure the data but also ensure the reliability and continuity of data visualisation efforts.

### Cybersecurity

**What is cybersecurity?**

Cybersecurity refers to the practices and technologies used to protect computers, networks and data from unauthorised access, attacks, or damage. When creating or sharing data visualisations, especially online or in a networked environment, it is crucial to ensure that the data is secure from cyber threats.

**Protecting data during visualisation**

When data is being visualised, it is often pulled from databases or data warehouses. Ensuring that these sources are secure is the first step in maintaining data security. This can involve encrypting the data, which means turning it into a code that only authorised users can read. Encryption helps protect data from being intercepted or accessed by hackers during transmission or storage.

**Access controls**

Another important aspect of cybersecurity is controlling who has access to the data. Organisations often use access controls, which restrict access to certain data based on the user’s role or identity. For example, a company might allow only specific employees to view or modify financial data. This ensures that sensitive information is only accessible to those who need it.

**Using secure platforms**

When sharing or collaborating on data visualisations, it is important to use secure platforms. These platforms often have built-in security features, such as user authentication, encryption and activity monitoring, which help protect the data from unauthorised access.

**Activity 23:** high school scenario

Imagine your school is creating a visualisation of student performance data. The visualisation includes grades, attendance records and participation in extracurricular activities. This data will be shared with teachers and administrators to help improve student outcomes.

If this data were to fall into the wrong hands, it could be misused. By applying cybersecurity measures like encryption and access controls, the school can ensure that only authorised teachers and administrators can view or modify the data.

**Identify the risks**

List 3 potential risks if the student performance data falls into the wrong hands. Consider who might misuse the data and how it could negatively impact students.

|  |
| --- |
| **Sample answer:**  If student performance data falls into the wrong hands, potential risks include:   1. Privacy violations: sensitive information about students' academic performance and personal records could be exposed, leading to embarrassment or harm. 2. Identity theft: unauthorised access to personal student data could be used for identity theft or fraud. 3. Discrimination: misuse of data might result in unfair treatment of students based on their performance or background, impacting their opportunities and self-esteem. |

**Encryption importance**

****Explain why encryption is important when transmitting student performance data from the school database to the data visualisation tool. How does encryption protect the data?

|  |
| --- |
| **Sample answer:**  Encryption is crucial because it ensures that even if someone intercepts the data during transmission from the school’s database to the visualisation tool, they cannot read or misuse it. Encryption converts the data into a code that can only be deciphered by those with the correct decryption key, protecting students' sensitive information from unauthorised access. |

**Access control design**

****Design a basic access control system for this scenario. Who should have access to the data and what kind of access should they have (for example view-only, edit, or full control)? Justify your choices.

|  |
| --- |
| **Sample answer:**  **Access control system:**   * **Teachers**: view-only access to data for their students to monitor progress without being able to alter records. * **Administrators**: edit access to manage and update records as needed. * **IT Staff**: full control to manage the database and ensure its security.   This system limits access based on role responsibilities, reducing the risk of data being altered or misused while ensuring those who need to manage the data can do so. |

**Secure platform choice**

****Research and recommend a secure platform for sharing the data visualisation with teachers and administrators. Explain why this platform is a good choice in terms of cybersecurity features.

|  |
| --- |
| **Sample answer:**  **Recommended platform:** Google Workspace for Education  **Justification:** Google Workspace offers robust security features, including encryption, two-factor authentication and access controls. It also provides easy sharing options with specific permissions, making it a good choice for securely sharing the data visualisation with teachers and administrators. |

**Role-playing exercise**

****Split into groups and role-play a situation where a teacher accidentally shares a student's performance data with an unauthorised person. Discuss how this breach could have been prevented and what steps should be taken afterward.

|  |
| --- |
| **Sample answers:**  **Prevention**: the breach could have been prevented by ensuring that the teacher understood the importance of data security and by setting strict access controls that prevent unauthorised sharing.  **Response**: after the breach, the school should notify the affected students and parents, investigate the incident and review their data security policies to prevent future occurrences. |

### Data backup

**Why is data backup important?**

Data backup is the process of making copies of data and storing them in a secure location. This is crucial because it ensures that the data can be recovered in case it is lost, corrupted or compromised due to a cyberattack, hardware failure or accidental deletion.

**Backup methods**

There are several methods for backing up data. One common method is using cloud storage, where data is saved on remote servers accessible via the internet. Cloud storage providers often have robust security measures in place, such as encryption and multi-factor authentication, to protect the data. Another method is using external hard drives or other physical storage devices to create local backups. These should be kept in a secure location, and it is essential to ensure they are regularly updated with the latest data.

**Automated backups**

Many organisations use automated backup systems that regularly back up data at specific intervals, such as daily or weekly. This helps ensure that the most recent data is always protected without relying on manual processes.

**Importance for data visualisation**

For data visualisation, maintaining up-to-date backups is important because visualisations often rely on large datasets. If the data source is lost or corrupted, having a backup ensures that the visualisations can be re-created without starting from scratch. It also protects against data loss that could disrupt important business decisions or research.

**Activity 24:** company sales dashboard

**Scenario**

Consider a company that has created a dashboard showing sales trends. If the company’s database is compromised, the visualisation could be lost. However, if the data is backed up regularly, the company can quickly restore the database and continue to use the dashboard without significant disruption.

A company has created a dashboard that visualises sales trends over the past year. The data is stored in a database that is regularly backed up.

One day, the database is compromised due to a cyberattack.

**Impact of data loss**

Describe the potential consequences for the company if they did not have a backup of the sales data. How would this affect their business operations and decision-making?

|  |
| --- |
| **Sample answer:**  Without a backup, the company could face several consequences:   1. **Business disruption:** loss of sales data could halt operations as the company struggles to recover lost information. 2. **Financial loss:** inaccurate or missing data might lead to poor business decisions, affecting revenue. 3. **Reputation damage:** clients and stakeholders may lose trust in the company's ability to manage data securely, damaging its reputation. |

**Backup method evaluation**

Compare and contrast cloud storage and external hard drives as backup methods. Which method do you think would be better for this company’s sales data and why?

|  |
| --- |
| **Sample answer:**  **Cloud storage versus external hard drives:**   * **Cloud storage**: offers automatic backups, accessibility from anywhere and strong security features like encryption and disaster recovery options. * **External hard drives**: provide a physical backup option, but are vulnerable to theft, damage and require manual updates. * **Recommendation**: cloud storage is better for the company’s sales data due to its automation, security and ease of access, ensuring that data is consistently backed up and can be quickly restored. |

**Automated backup planning**

Create a backup plan for the company’s sales data. How often should backups be made and where should the backups be stored? Explain your reasoning.

|  |
| --- |
| **Sample answer:**  **Backup plan:**   * **Frequency:** daily backups to ensure minimal data loss. * **Storage location:** cloud storage with a reputable provider, ensuring data is stored securely and can be accessed quickly if needed. * **Reasoning:** daily backups minimise the risk of losing significant amounts of data, while cloud storage ensures that the backups are secure and easily retrievable in case of an emergency. |

**Cyberattack response**

****Outline the steps the company should take immediately after discovering that their database has been compromised. Include both short-term actions (for example restoring the backup) and long-term actions (for example improving security measures).

|  |
| --- |
| **Sample answer:**  **Immediate steps:**   1. Isolate the affected systems to prevent the spread of the attack. 2. Restore the database from the most recent backup to minimise data loss.   **Long-term actions:**   1. Investigate the breach to understand how it occurred and to prevent future incidents. 2. Upgrade security measures, such as enhancing encryption, implementing stronger access controls and conducting regular security audits. |

**Visualising data security**

Using a data visualisation tool, create a simple chart that shows the frequency of backups over the last month. Discuss how visualising this information can help the company monitor the effectiveness of their data security strategy.

|  |
| --- |
| **Sample answer:**  Backup frequency chart:this chart shows that backups were made daily over the last month, with no gaps in the schedule.  Visualising backup frequency helps the company monitor their backup strategy’s effectiveness, ensuring that data is consistently protected. Any gaps in the chart would indicate a failure in the backup process, prompting immediate action to address the issue. |

**Activity 25**: reflection questions

How do cybersecurity and data backup practices contribute to the overall reliability and trustworthiness of data visualisations?

|  |
| --- |
| **Sample answer:**  Cybersecurity and data backup practices ensure that data visualisations are based on accurate and protected data, maintaining their reliability. Without these practices, the data could be compromised, leading to incorrect conclusions and decisions. |

****Why is it important to consider data security when creating visualisations that will be shared with others?

|  |
| --- |
| **Sample answer:**  Data security is essential when creating and sharing visualisations because it protects sensitive information from unauthorised access and misuse. Ensuring data is secure helps maintain trust, integrity and compliance with privacy regulations. |

**Activity 26:** investigate and implement methods to maintain data security

Data security is very important. Examine this [data visualisation](https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/) of data breaches from across the world.

Answer the following questions on methods to maintain data security.

What are some potential security risks when collecting, storing, or visualising data?

|  |
| --- |
| **Sample answer:**  Potential security risks include unauthorised access to data, data breaches due to weak passwords, insecure storage methods like unencrypted databases, and risks from third-party services or tools used in visualisation. For example, an unencrypted database can expose sensitive information if accessed without permission. |

How would you secure sensitive data in an enterprise environment?

|  |
| --- |
| **Sample answer:**  To secure sensitive data, I would implement strong password policies, multi-factor authentication, and role-based access controls. I would also use encryption for both stored and transmitted data and regularly update software to address vulnerabilities. For instance, encrypting data at rest and in transit helps to prevent unauthorised access. |

What data security practices would you implement in a visualisation project?

|  |
| --- |
| **Sample answer:**  In a visualisation project, I would implement practices such as data anonymisation to protect personal information, using HTTPS protocols for secure communication, and ensuring access controls to restrict data access to authorised users only. For example, anonymising personally identifiable information before visualising it would help in maintaining privacy. |

How would you evaluate the effectiveness of the security measures used in protecting data integrity?

|  |
| --- |
| **Sample answer:**  I would evaluate the effectiveness by conducting regular security audits, penetration testing, and reviewing incident reports. Additionally, I would gather user feedback to identify any access issues or security loopholes. For instance, if penetration tests reveal no vulnerabilities and user feedback is positive, it suggests that the implemented security measures are effective. |

# References

This resource contains NSW Curriculum and syllabus content. The NSW Curriculum is developed by the NSW Education Standards Authority. This content is prepared by NESA for and on behalf of the Crown in right of the State of New South Wales. The material is protected by Crown copyright.

Please refer to the NESA Copyright Disclaimer for more information [https://educationstandards.nsw.edu.au/wps/portal/nesa/mini-footer/copyright](https://aus01.safelinks.protection.outlook.com/?url=https%3A%2F%2Feducationstandards.nsw.edu.au%2Fwps%2Fportal%2Fnesa%2Fmini-footer%2Fcopyright&data=05%7C01%7CCaitlin.Pace1%40det.nsw.edu.au%7C9c2c1a9f59c94d2df30708dafa7edb23%7C05a0e69a418a47c19c259387261bf991%7C0%7C0%7C638097720042599463%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=BzQh0UsffVZE3eO22b2Xba3p0VMOBZSHfS21FGHXtZM%3D&reserved=0).

NESA holds the only official and up-to-date versions of the NSW Curriculum and syllabus documents. Please visit the NSW Education Standards Authority (NESA) website <https://educationstandards.nsw.edu.au/> and the NSW Curriculum website [https://curriculum.nsw.edu.au/home](https://curriculum.nsw.edu.au/).

[Enterprise Computing 11–12 Syllabus](https://curriculum.nsw.edu.au/learning-areas/tas/enterprise-computing-11-12-2022/overview) © NSW Education Standards Authority (NESA) for and on behalf of the Crown in right of the State of New South Wales, 2022.

[Enterprise Computing 11–12 course specifications](https://curriculum.nsw.edu.au/learning-areas/tas/enterprise-computing-11-12-2022/overview" \l ":~:text=30-,Enterprise%20Computing%20course%20specifications,-Enterprise%20Computing%20Course) © NSW Education Standards Authority (NESA) for and on behalf of the Crown in right of the State of New South Wales, 2022.

365 Data Science (5 June 2020) ‘[What Is a Data Warehouse?’ [video]](https://www.youtube.com/watch?v=AHR_7jFCMeY), 365 Data Science, YouTube, accessed 29 October 2024.

Altexsoft (25 August 2022) ‘[Enterprise Data Warehouse: EDW Components, Key Concepts, and Architecture Types](https://www.altexsoft.com/blog/enterprise-data-warehouse-concepts/)’, Altexsoft, accessed 29 October 2024.

Canva (30 August 2022) [‘Storytelling with Data: Infographics and data visualization | Canva for Journalists Episode 5 of 6’ [video]](https://www.youtube.com/watch?v=SuFcF2uS1zY), Canva, YouTube, accessed 29 October 2024.

Canva (n.d.) [Workbook](https://canva.me/data-workbook-desktop) – Charts and Tables -Journalist Essentials [website], accessed 29 October 2024.

CareerFoundry (26 November 2021) [‘Data Analytics – The 9 Essential Tools! (2024)’ [video]](https://www.youtube.com/watch?v=jgXp1EE4Wms), CareerFoundry, YouTube, accessed 29 October 2024.

COURSE PLEASE (31 July 2023) [‘OLAP Cubes: Unleashing Multi-Dimensional Data Insights’ [video]](https://www.youtube.com/watch?v=s8UDqSRXgbg), COURSE PLEASE, YouTube, accessed 29 October 2024.

Commonwealth of Australia (2020) [*Coronavirus (COVID-19) at a glance – 25 November 2020*](https://www.health.gov.au/resources/publications/coronavirus-covid-19-at-a-glance-25-november-2020), The Department of Health and Aged Care website, accessed 14 November 2024.

eagereyes – Robert Kosara (15 March 2021) ‘[Minard's Famous "Napoleon's March" Chart – What It Shows, What It Doesn't’ [video]](https://www.youtube.com/watch?v=hlb1uM_SOcE), eagereyes – Robert Kosara, YouTube, accessed 29 October 2024.

Gaslowitz L (7 July 2017) [‘How to spot a misleading graph’ [video]](https://www.youtube.com/watch?v=E91bGT9BjYk), TED-Ed, YouTube, accessed 29 October 2024.

IBM Technology (3 April 2021) [‘Software as a Service (SaaS) Explained in 5 mins’ [video]](https://www.youtube.com/watch?v=20QUNgFIrK0), IBM Technology, YouTube, accessed 29 October 2024.

Kalodikis C (16 August 2024) [‘Enterprise Computing Year 12 Unit 2: Data visualisation’ [video]](https://www.youtube.com/watch?v=g4biYeAfkmU), Christopher Kalodikis, YouTube, accessed 29 October 2024.

Klein L, Sharma T, Varner J, Li S, Adams M, Yang N, Jutan D, Fu J, Mola A, Fang Z, Li Y, and Munro S (2024) ‘[Visualization as Argument, William Playfair’s Time-Series Charts](https://dataxdesign.io/chapters/playfair)’, in Data by Design: An Interactive History of Data Visualization, 1789-1900, accessed 29 October 2024.

Loehrke J (9 February 2023) ‘[Graphics: Super Bowl pizza sales reach into the millions. Here are the toppings favored most in the US](https://www.usatoday.com/in-depth/graphics/2023/02/09/national-pizza-day-super-bowl-favorite-toppings-graphics/11201105002/)’, USA TODAY, accessed 29 October 2024.

Microsoft Learn (2024) [*Get data with Power BI Desktop*](https://learn.microsoft.com/en-gb/training/modules/get-data-power-bi/)[online learning], Microsoft Learn, accessed 29 October 2024.

MyOnlineTrainingHub (12 March 2024) [‘From Excel to Power BI in 12 Minutes (Beginner Essentials)’ [video]](https://www.youtube.com/watch?v=gOs7EC-FebE), MyOnlineTrainingHub, YouTube, accessed 29 October 2024.

Nielson, J (15 February 2024) [How I Developed the 10 Usability Heuristics](https://www.uxtigers.com/post/usability-heuristics-history), *UX Tigers*, accessed 29 October 2024.

NSW Government (n.d.) ‘[Transport for NSW Open Data](https://opendata.transport.nsw.gov.au/dataset/)’, Datasets, NSW Government website, accessed 29 October 2024.

nullQueries (5 May 2021) [‘Using Design Techniques for Clear and Appealing Data Visualization’ [video]](https://www.youtube.com/watch?v=0Smgm2UTUSo), nullQueries, YouTube, accessed 29 October 2024.

Salesforce, Inc (2024) [*Data Is Beautiful: 10 Of The Best Data Visualization Examples From History & Today*](https://www.tableau.com/learn/articles/best-beautiful-data-visualization-examples), Tableau website, accessed 29 October 2024.

Smith D and Wynne E (10 October 2024) ,[New car buyers driven to white or grey, with bright colours rarer than ever, data shows](https://www.abc.net.au/news/2024-10-10/new-car-colours-more-monochrome-than-ever-before-data-shows/104451484)’, ABC News, accessed 29 October 2024.

Statista (2024) [*Most popular types of pizza ordered among consumers in Australia as at February 2023*](https://www.statista.com/statistics/1410088/australia-popular-pizza-orders/) [data set], Statista website, accessed 29 October 2024.

Surfline\Wavetrak (2024) [Surfline](https://www.surfline.com/), *Surfline*, accessed 29 October 2024.

The Royal Society (20 April 2023) ‘[Mapping data visualisations meteoric rise from Victorian London to today’ [video]](https://aeon.co/videos/mapping-data-visualisations-meteoric-rise-from-victorian-london-to-today), The Royal Society, Aeon Media Group, accessed 29 October 2024.

Visme (20 May 2022) [‘Data Visualization in 2024 | The Ultimate Guide’ [video]](https://www.youtube.com/watch?v=loYuxWSsLNc), Visme, YouTube, accessed 29 October 2024.

**© State of New South Wales (Department of Education), 2024**

The copyright material published in this resource is subject to the *Copyright Act 1968* (Cth) and is owned by the NSW Department of Education or, where indicated, by a party other than the NSW Department of Education (third-party material).

Copyright material available in this resource and owned by the NSW Department of Education is licensed under a [Creative Commons Attribution 4.0 International (CC BY 4.0) license](https://creativecommons.org/licenses/by/4.0/).

[](https://creativecommons.org/licenses/by/4.0/)

This license allows you to share and adapt the material for any purpose, even commercially.

Attribution should be given to © State of New South Wales (Department of Education), 2024.

Material in this resource not available under a Creative Commons license:

* the NSW Department of Education logo, other logos and trademark-protected material
* material owned by a third party that has been reproduced with permission. You will need to obtain permission from the third party to reuse its material.

**Links to third-party material and websites**

Please note that the provided (reading/viewing material/list/links/texts) are a suggestion only and implies no endorsement, by the New South Wales Department of Education, of any author, publisher, or book title. School principals and teachers are best placed to assess the suitability of resources that would complement the curriculum and reflect the needs and interests of their students.

If you use the links provided in this document to access a third-party's website, you acknowledge that the terms of use, including licence terms set out on the third-party's website apply to the use which may be made of the materials on that third-party website or where permitted by the *Copyright Act 1968* (Cth). The department accepts no responsibility for content on third-party websites.