

# NPA Data Citizenship

Learners' Guide to the  
National Progression Award

# Level 5 NPA Data Citizenship Notes 2025

Data Education in Schools

## Contents

|          |                                                                                                                 |           |
|----------|-----------------------------------------------------------------------------------------------------------------|-----------|
| 0.1      | Support and Resources                                                                                           | 5         |
| <b>1</b> | <b>Outcome 1 - Describe the use of data in society.</b>                                                         | <b>6</b>  |
| 1.1      | 1a - Describe reasons for the growth of data.                                                                   | 6         |
| 1.2      | 1b - Describe how data can be both used and misused, and its impact on individuals, organisations, and society. | 6         |
|          | The role of data in decision-making.                                                                            | 6         |
|          | Success Stories and Challenges in Data Usage                                                                    | 7         |
|          | Data Disasters                                                                                                  | 7         |
| 1.3      | 1c - Describe the rights and responsibilities of data subjects.                                                 | 8         |
|          | Personal Data                                                                                                   | 8         |
|          | Risks of Sharing Online                                                                                         | 9         |
|          | Benefits of Sharing Online                                                                                      | 9         |
|          | How Long Does Data Stay Online?                                                                                 | 9         |
|          | Can It Be Deleted?                                                                                              | 10        |
|          | GDPR                                                                                                            | 10        |
| 1.4      | 1d - Describe methods of keeping data secure.                                                                   | 11        |
| <b>2</b> | <b>Outcome 2 - Describe data literacy concepts.</b>                                                             | <b>12</b> |
| 2.1      | 2a - Describe reasons for poor quality data and its impacts.                                                    | 12        |
|          | What is high-quality data?                                                                                      | 12        |
|          | The benefits of using high-quality data                                                                         | 12        |
|          | Spotting Poor-Quality Data                                                                                      | 13        |
| 2.2      | 2b - Describe types of common data visualisations and state where each is most appropriate.                     | 15        |
|          | Frequency tables                                                                                                | 15        |
|          | Dot plot                                                                                                        | 15        |
|          | Bar Chart                                                                                                       | 15        |
|          | Line graph                                                                                                      | 17        |
|          | Pie Chart                                                                                                       | 17        |

|                                                                                              |           |
|----------------------------------------------------------------------------------------------|-----------|
| Histogram . . . . .                                                                          | 20        |
| Scatterplot . . . . .                                                                        | 20        |
| Bubble plot . . . . .                                                                        | 20        |
| Heat map . . . . .                                                                           | 22        |
| Time series graph . . . . .                                                                  | 22        |
| Stacked area chart . . . . .                                                                 | 22        |
| 2.3 2c - Describe how data visualisations can be interpreted and misinterpreted. . . . .     | 24        |
| 2.4 2d - State methods of gathering data using a survey. . . . .                             | 24        |
| <b>3 Outcome 3 - Interpret data.</b>                                                         | <b>25</b> |
| 3.1 3a - Create and carry out a survey to investigate a problem. . . . .                     | 25        |
| 3.2 3b - Interpret data visualisations to identify patterns and trends. . . . .              | 26        |
| 3.3 3c - Describe data visualisations in terms of quality and trust. . . . .                 | 26        |
| Examples of Graphical Crimes . . . . .                                                       | 27        |
| 3.4 3d - Describe data generated from AI in terms of quality and trust. . . . .              | 28        |
| 3.5 3f - Make recommendations based on conclusions and communicate findings. . . . .         | 29        |
| Tips for Drawing Conclusions and Communicating Findings . . . . .                            | 29        |
| Examples of drawing conclusions, communicating findings, and making recommendations. . . . . | 29        |

## Introduction

Welcome to the NPA Data Science Notes for 2025! These notes are designed to guide you through the content for your NPA Data Science qualification.

These notes have been written for the updated (2024) NPA Data Science specification.

This document is a summary document covering the core concepts that you will need to know in order to learn the content and undertake the assessments. It can be used by educators to introduce each topic, or for learners as they go through the course as a support resource.

Throughout the guides, you will come across links to videos, and lessons which relate to the content.

These notes are organised by learning outcome. At the beginning of each Outcome section, you will find links to the lessons related to that Outcome.

## Support and Resources

These guides have been written with the support of the University of Edinburgh's Data Education in Schools team. The Data Education in Schools project aims to work with schools and colleges that are delivering this course. To date, they have worked with every school delivering this qualification, providing professional learning, facilitating sharing of resources, and working together to review materials and share the development workload.

Visit [www.dataschools.education](http://www.dataschools.education) for more information about support materials.

For the NPA Lessons which were developed for the previous version of this course, visit [www.dataed.in/learndata](http://www.dataed.in/learndata). These lesson materials are also linked to throughout this guide in relevant sections.

Visit [dataed.in/NPADS](http://dataed.in/NPADS) for more information about the qualification on the SQA site.

This document covers the Level 5 Data Citizenship unit in particular. There are separate documents available for other levels. [Insert link to other documents here]



Scan the QR code or go to [dataschools.education/level-5-data-citizenship-lessons/](http://dataschools.education/level-5-data-citizenship-lessons/) for relevant lessons and resources for this unit, separated by outcome.

## 1 Outcome 1 - Describe the use of data in society.

### 1a - Describe reasons for the growth of data.

In recent years, data is needed more, and larger and larger amounts of data are being stored and processed. Reasons for this increase in the amount of data and how much it's used include factors such as:

- Increased number of data-generating devices
- Internet of Things (IoT) and cloud computing
- Faster and wider internet availability, including 5G
- Increased demand
- Increase in use of AI, Machine Learning, and Generative AI

### 1b - Describe how data can be both used and misused, and its impact on individuals, organisations, and society.

#### The role of data in decision-making.

Data plays an important part in lots of areas of society. Below are some key areas **where data plays a part in decision-making**:

##### 1. Home

- **Energy Usage:** Smart meters collect data on electricity and gas usage, helping families save money by identifying when to use appliances more efficiently.
- **Shopping Decisions:** Online retailers track buying habits and suggest products based on past purchases or popular trends.
- **Health Monitoring:** Smartwatches or fitness trackers collect data about steps, sleep, and heart rate, enabling people to improve their health routines.

##### 2. Community

- **Transportation:** Data from buses, trains, and traffic sensors can show where services need improvement, such as adding more buses on busy routes.
- **Public Safety:** Communities use data from crime reports to decide where to increase patrols or install security cameras.
- **Planning Events:** Surveys or attendance data from past events help organizers plan better events tailored to the interests of the community.

##### 3. Sport

- **Athlete Performance:** Wearable devices track speed, endurance, and recovery, helping athletes train smarter and avoid injuries.
- **Team Strategies:** Coaches analyze data from matches, like player positions or time spent with the ball, to improve tactics.
- **Fan Engagement:** Sports organizations use ticket sales and social media data to improve fan experiences, like offering discounts or organizing popular events.

## Success Stories and Challenges in Data Usage

Below are some examples of success stories from the use of data, as well as some examples of where the use of data has caused harm.

### Success Stories

#### 1. Using COVID-19 Data to Allocate Health Resources

- **Case:** Hospitals use patient data to identify health risks and personalize treatments. During the COVID-19 pandemic, data was used to track infection rates, allocate vaccines, and manage hospital capacity. ([Article on Gov Website](#))
- **Impact:** Lives were saved, and resources were used more efficiently.

#### 2. Liverpool FC Using Data to Improve Performance

- **Case:** In football, teams like Liverpool FC use data analytics to improve player performance and game strategies ([Article](#)). Their data-driven approach contributed to their Premier League win in 2020 after a 30-year gap. ([Article](#))
- **Impact:** Better team strategies and higher chances of success.

#### 3. Successful Recommendations for Shoppers by Amazon

- **Case:** Amazon uses customer data to recommend products, personalize the shopping experience, and optimize delivery times. ([Article](#))
- **Impact:** Increased sales and customer satisfaction, making Amazon a global leader in e-commerce.

#### 4. Using Data to Implement Congestion Tax in Stockholm

- **Case:** In Stockholm, Sweden, data from traffic sensors helped implement a congestion tax, reducing traffic volume by 25% and improving air quality. ([Article by IBM](#))
- **Impact:** Better urban living conditions.

### Data Disasters

#### 1. Facebook and Cambridge Analytica Scandal (2018)

- **Case:** Data from millions of Facebook users was harvested without proper consent and used for political campaigns. ([Article](#))
- **Impact:** Public trust was damaged, and Facebook faced fines and stricter regulations.

#### 2. UK COVID-19 Data Loss (2020)

- **Case:** The UK government lost around 16,000 COVID-19 test results due to a spreadsheet error. The system couldn't handle the large dataset. ([Article](#))
- **Impact:** Delayed contact tracing and potential spread of the virus.

#### 3. Target's Pregnancy Prediction Backlash (2012)

- **Case:** Target used customer data to predict pregnancies and sent related ads to customers. This accidentally revealed a teenager's pregnancy to her family. ([Article](#))
- **Impact:** Privacy concerns and public criticism.

## 1c - Describe the rights and responsibilities of data subjects.

### Personal Data

#### Personal Data

Any information relating to an identified or identifiable natural person.

#### Data subject

The identified or identifiable living individual to whom the personal data relates.

There is a sub-category of personal data called **sensitive personal data**, which is required to be treated even more stringently than personal data. This includes the personal data of children (anyone under 18). Sensitive personal data should not be collected or processed except under certain conditions and with an identified lawful basis for doing so.

This table gives examples of data that would be classed as personal or sensitive personal data.

| Personal Data                                                                                                                                                                                                                                                                     | Sensitive Personal Data                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Names</li> <li>• Addresses</li> <li>• Phone numbers</li> <li>• Identification numbers</li> <li>• Location data</li> <li>• Online identifiers</li> <li>• A combination of identifiers that together can identify an individual</li> </ul> | <ul style="list-style-type: none"> <li>• Racial or ethnic origin</li> <li>• Political opinions</li> <li>• Religious or philosophical beliefs</li> <li>• Genetic data</li> <li>• Biometric data, where used for identification</li> <li>• Health data</li> </ul> |

### What Data Is Shared When Going Online?

It is important to actively manage your privacy online, otherwise more information may be shared than necessary. The kind of information that is often being stored, and possibly shared, is more than just name and email addresses. It could be:

- Geographic location
- Web browsing habits
- Websites visited
- Products bought online
- Illnesses searched for online
- Devices used to connect to the internet
- Reading habits and history

- Food preferences
- Political views

## Risks of Sharing Online

When sharing data online, users should consider these risks.

- **Privacy Breaches:** Personal information (like your address, phone number, or location) can be exposed to strangers or unauthorized parties.  
**Example:** Posting a photo with a visible address or geotag can reveal your location to others.
- **Identity Theft:** Hackers can use your shared information (e.g., full name, birth date, or photos of documents) to impersonate you and commit fraud.  
**Example:** Sharing a photo of your new ID or credit card can lead to theft of your identity or financial details.
- **Reputation Damage:** Old posts, even if intended as jokes, can resurface and harm your reputation in the future, especially when applying for jobs or schools.  
**Example:** A controversial tweet from years ago could lead to public backlash.
- **Phishing and Scams:** Scammers may use shared information to trick you into revealing sensitive details or money.  
**Example:** Posting about a recent purchase might make you a target for fake refund scams.

## Benefits of Sharing Online

While there is risk in sharing data online, there are also benefits.

- **Connecting with Others:** Sharing updates, photos, and stories helps you stay connected with family, friends, and communities, even if they are far away.  
**Example:** Posting family photos can keep relatives updated on your life.
- **Sharing Knowledge and Ideas:** Online platforms allow you to share expertise, learn from others, and contribute to global conversations.  
**Example:** Writing blogs or creating tutorials helps others while showcasing your skills.
- **Raising Awareness and Advocacy:** Social media can amplify your voice and help bring attention to causes you care about.  
**Example:** Sharing information about environmental initiatives can inspire others to take action.
- **Expressing Creativity:** Sharing art, music, writing, or other creative content allows you to express yourself and gain feedback.  
**Example:** Posting your paintings on Instagram can help you build an audience and improve your skills.

## How Long Does Data Stay Online?

Personal data shared online can remain **indefinitely** because:

- Platforms store backups even after deletion.
- Shared content can be copied or reshared.
- Search engines cache old versions of web pages.

### Can It Be Deleted?

Removing personal data is sometimes possible in the following ways:

1. Delete Content: Remove posts and adjust privacy settings.
2. Request Deletion: Contact websites or use search engine tools like Google's "Remove Outdated Content."
3. Delete Old Accounts: Delete accounts that you are no longer using.

### GDPR

GDPR is an EU-wide law that applies to the processing of personal data either for activities carried out by processors established in the EU, whether or not the processing takes place inside or outside the EU. It also covers offering goods, services, or monitoring behaviour within the EU, whether or not the processor is based in the EU.

### Your Data Rights

Individuals have certain rights under GDPR. These are listed in the boxes below.

#### Informed

A privacy notice provides transparency about the use of their personal data. This should be in clear plain language and be age-appropriate if aimed at children.

#### Access

Individuals can ask to see what data is held on them. This is called a Subject Access Request. This should be free and should be dealt with within one month.

#### Rectification

Any data that is incorrect or incomplete should be fixed for free. This request should be dealt with within one month.

#### Erasure

Known as the right to be forgotten, individuals can ask for personal data to be deleted. This can be refused in certain circumstances such as crime prevention or public health reasons.

#### Restrict processing

Individuals can ask to limit how data is used. This could be especially relevant if awaiting rectification.

#### Portability

Individuals can move their data between different providers, however currently agreed standards for most data sharing does not exist.

#### Automated decision-making

Individuals can insist that decisions, such as applications for credit, are not made using automated algorithms and can request they be made manually.

#### Object

This is the right to object to their data being processed. This could be as simple as unsubscribing from marketing.

## GDPR Responsibilities

GDPR states **responsibilities** for those **holding data**:

- Provide **accurate** and **up-to-date** data
- Be responsible when using their GDPR rights and consider the **impact** of their request on organisations
- Be aware of **risks of sharing personal data**; secure devices and accounts from unauthorised access

### 1d - Describe methods of keeping data secure.

If data is private, it is critically important to both individuals and businesses to keep it secure. This will stop it falling into the wrong hands.

Keeping data safe is everybody's responsibility. Human beings are often unknowingly the weakest link in keeping data secure.

#### Personal data

information that relates to an identified or identifiable individual

Strategies for keeping personal data secure might include methods such as:

- **Strong passwords:** a combination of letters, numbers and special characters that are difficult to guess by a person or program.
- **Password manager:** a software that securely stores passwords that a user has for online accounts.
- **Anti-virus software:** Software designed to detect and destroy computer viruses.
- **Using encryption:** A way of scrambling data so that it can only be decoded by the intended recipient.

Some more advanced ways that users can protect their data are:

- **Multifactor Authentication:** Adds an extra layer of security by requiring multiple forms of verification before granting access.
- **Biometrics:** Uses unique physical characteristics, like fingerprints or facial recognition, to verify identity.
- **Wiping Drives Prior to Disposal:** Permanently erases data from storage devices to prevent unauthorised access.
- **Firewalls:** Monitors and controls incoming and outgoing network traffic to block unauthorised access to the network.
- **VPNs (Virtual Private Networks):** Encrypts internet connections to protect data and maintain privacy online.
- **Software Upgrades:** Keeps systems secure by updating to the latest versions, which often include security patches.



## Benefits of high-quality data

Figure 1: The benefits of using high-quality data: improved customer experience, reduced risk, competitive advantage, increased revenue.

## 2 Outcome 2 - Describe data literacy concepts.

### 2a - Describe reasons for poor quality data and its impacts.

#### What is high-quality data?

High-quality data refers to data that correctly represents the real-world constructs that it is referring to. High-quality data is fit for the analytical purpose which it is being used for.

[This video](#) from IBM describes simply the different factors that come into data quality.

#### The benefits of using high-quality data

All analysis is only as good as the data it is carried out on. Therefore, the quality of the underlying data is critical to any analysis. There are benefits in using high-quality data. For businesses, the benefits of high-quality data include:

- **Improved customer experience:** For example, if high-quality data has been used to train recommendation systems, customers are likely to receive better recommendations.
- **Reduced risk:** Using high-quality data reduces the risk of inaccurate predictions, which could be potentially harmful.
- **Competitive advantage:** Companies using high-quality data can make better predictions, leading to happier customers, meaning they have a competitive advantage over other companies.
- **Increased revenue:** The improved customer experience leads companies to have increased revenue as they have more paying users.

# Reasons for Poor-Quality Data



Figure 2: Reasons for poor-quality data.

## Spotting Poor-Quality Data

Poor-quality data can significantly impact decision-making and outcomes. It is important to identify and address issues such as:

- **Out of Date:** Data that is not current may lead to decisions based on outdated trends or information.
- **Inaccurate:** Errors or inconsistencies in data can result in incorrect conclusions and misguided strategies.
- **Incomplete:** Missing data points can skew analysis and provide an unreliable picture of reality.
- **Gathered from a Small Sample Size:** Limited data can lead to biased results that do not accurately represent the larger population.

Regular reviews and data validation processes are essential to ensure data quality and reliability.

## Reasons for Poor Quality Data

There are a number of ways that data quality can be poor. These include:

- **Duplicate data (Uniqueness):** If data entries are duplicated, this might skew calculations about the distribution of data such as averages.
- **Inaccurate data (Accuracy):** Inaccurate data means that any predictions we make will be less accurate, and any claims made about findings are less plausible.

## Impacts of Poor-Quality Data



Figure 3: The impacts of poor quality data.

- **Outdated information (Timeliness):** Outdated information means that data is less applicable to any current work that is being done.
- **Missing values (Completeness):** This makes the overall spread of our data less accurate, particularly if missing data is concentrated in one specific area.
- **Non-standardised data (Consistency):** This means data lacks a consistent format or structure, which makes it difficult to compare or analyse.
- **Data security and privacy (Validity):** If data security is poor, unauthorised changes could be made to it, making it inaccurate or corrupted.

### Impacts of Poor-Quality Data

There are negative impacts of using poor-quality data in training or analysis for a company. These include:

- **Analysis rework:** Having to redo parts of an analysis because defects in the data were discovered too late.
- **Organisational inefficiencies:** Waste, delays and duplication that spread across teams and processes when data isn't fit for purpose.
- **Customer dissatisfaction:** Less accurate predictions and insights will lead a customer to be dissatisfied.
- **Opportunity cost of missed sales:** The value of sales that could have been achieved but weren't because of poor quality data.
- **Reputational costs from loss of trust:** If errors are made due to poor quality data, as a company, your reputation may be damaged.
- **Compliance costs or fines from incorrect reporting:** In some cases, if inaccurate claims are made, fines will have to be paid for inaccurate reporting.

| Mark | Tally | Frequency |
|------|-------|-----------|
| 4    |       | 2         |
| 5    |       | 2         |
| 6    |       | 4         |
| 7    |       | 5         |
| 8    |       | 4         |
| 9    |       | 2         |
| 10   |       | 1         |

Figure 4: A frequency table

## 2b - Describe types of common data visualisations and state where each is most appropriate.

When analysing data, it's useful to have a visual representation of that data so that we can more intuitively understand it. We call a visual representation of data a **graph**. When looking at a graph, it is often easier to spot patterns in the relationships between different variables in the data.

Graphs, charts, plots, visualisations, diagrams – these terms all mean roughly the same thing and are often used interchangeably

In this section, we look at the different types of graphs that you should know about, and when they are most suitable to use.

### Frequency tables

Frequency tables (Figure 4) provide a structured way to display how often each value in a dataset occurs. They are particularly useful for summarizing categorical data and identifying patterns.

A typical frequency table lists categories alongside their corresponding counts or frequencies. They offer a clear, concise overview of data distribution, making it easier to spot trends and outliers. Frequency tables are often used as a preliminary step before creating more complex visualizations like bar charts or histograms.

### Dot plot

In a dot plot (Figure 5, 6), each dot represents a single observation. For example, this dot plot records the month each child in a class of children was born. The dots can also be swapped for icons or images for a more visually appealing graphics.

### Bar Chart

Bar charts (Figure 7) use rectangular bars to compare values in different categories. The bars normally show the counts or sizes of categorical data. Since there is no connection between the bars, they are normally shown not touching.

A horizontal bar graph (Figure 8) is often a good option when there are many categories, or the category labels are long. It is also possible to reorder the bars, which makes it easier to see the

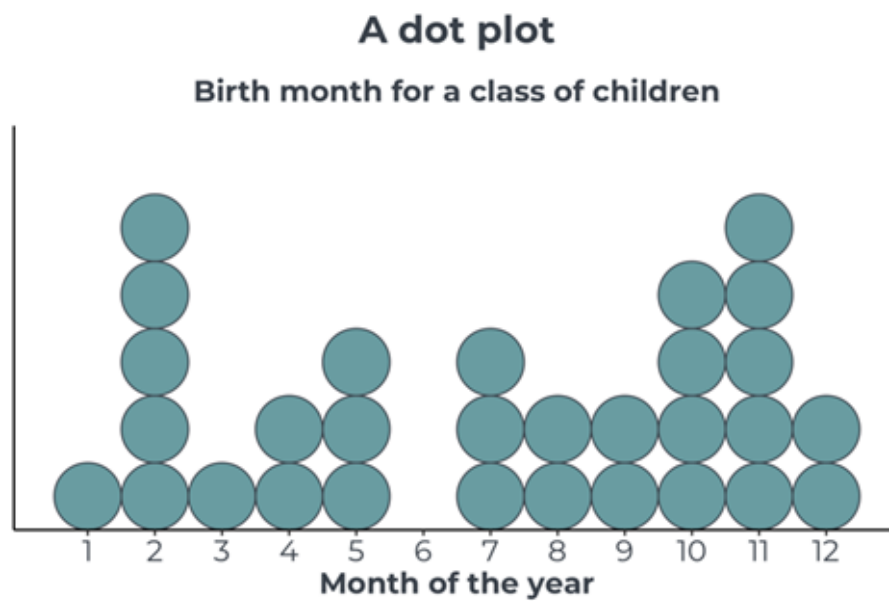


Figure 5: A dot plot.

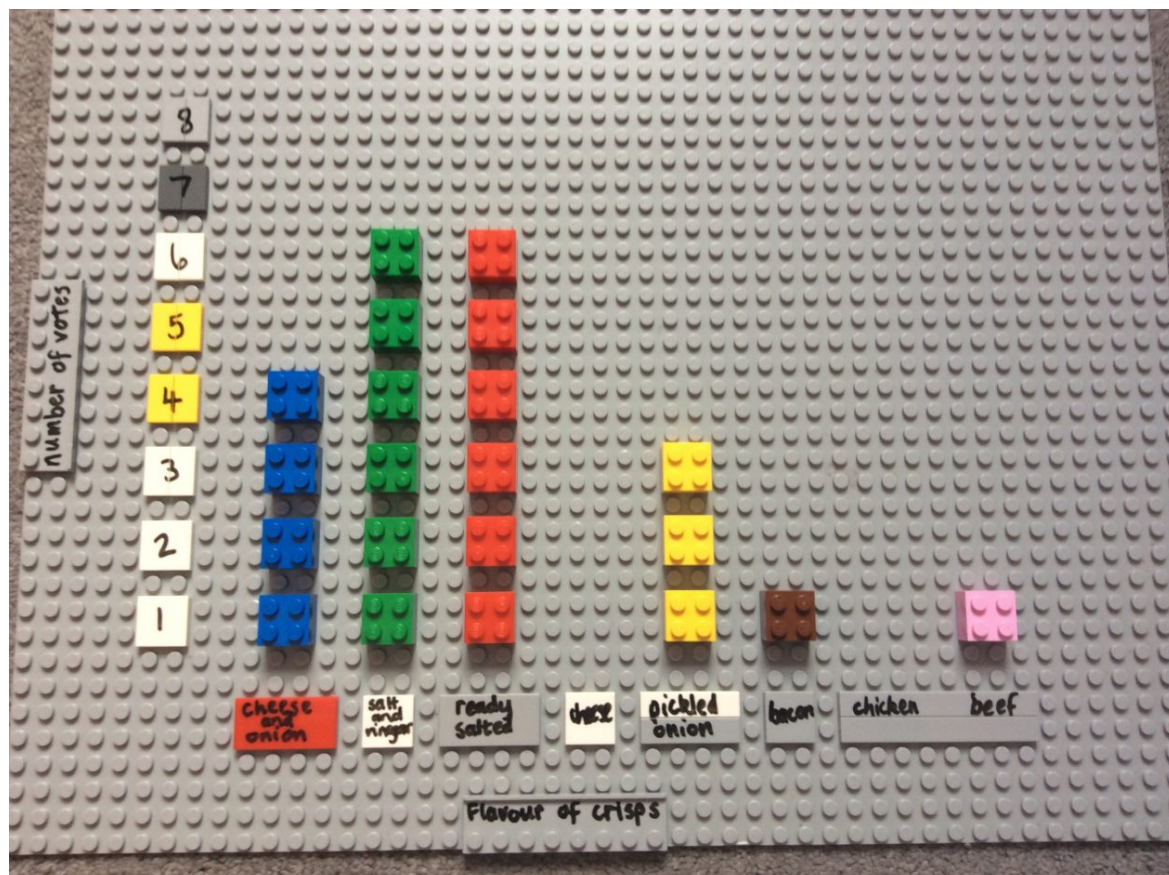


Figure 6: A dot plot made out of Lego.

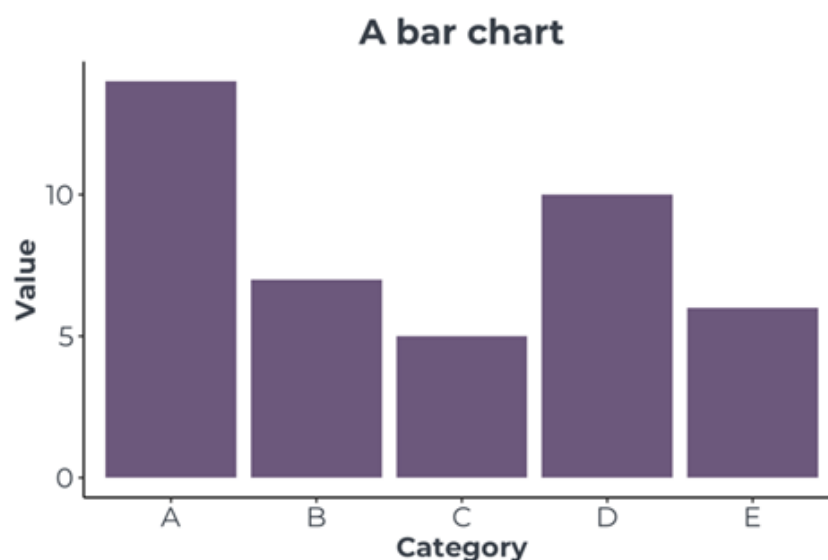


Figure 7: A bar chart.



Figure 8: A horizontal bar chart.

smallest and largest categories. These can also be highlighted by using different colours.

### Line graph

Line graphs (Figure 10) are used to show the change, or evolution of a numerical variable as another quantity varies. Both the x-axis and y-axis are numeric, with the x-axis containing the varying quantity. This is often time but could be another varying quantity such as temperature or distance. The data points in a line graph are joined sequentially by lines.

### Pie Chart

Pie charts (Figure 12) show the proportion of a whole. The total of the pie must add up to 100%. Although popular, pie charts are often not the best choice of graph to use, since it is much more difficult for human brains to estimate relative angles, or segments of the chart.

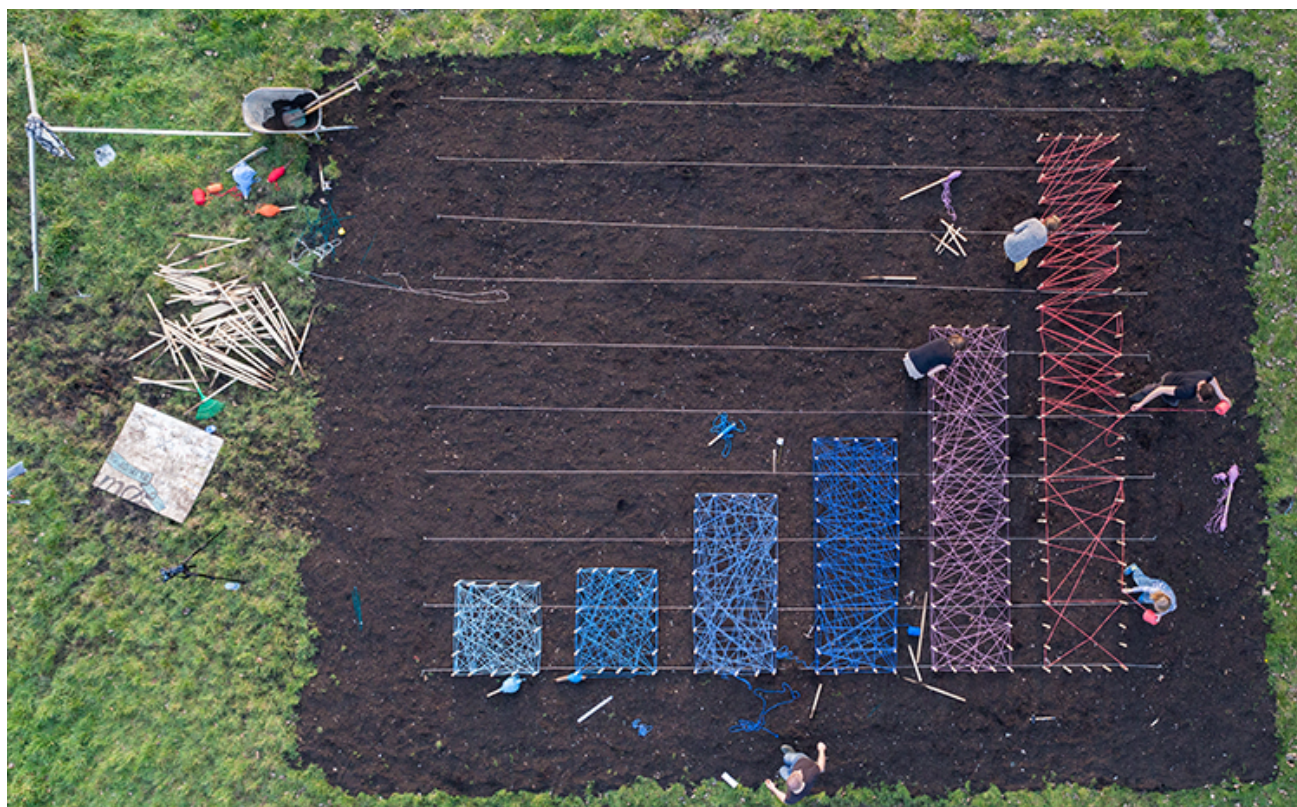


Figure 9: A vertical bar chart being made in a garden.

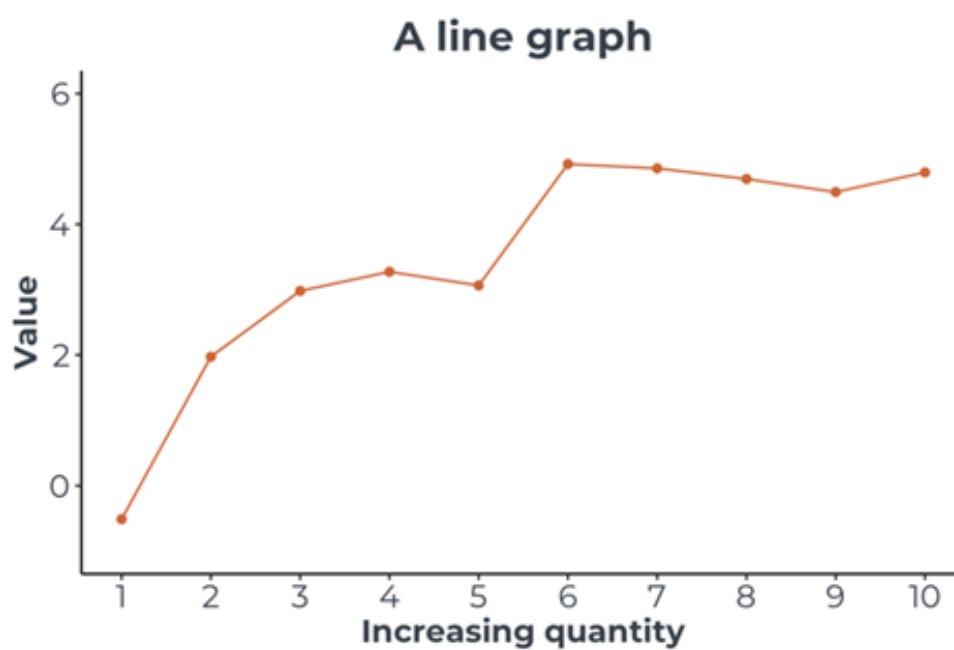


Figure 10: A line graph.

## SUBSTANTIAL Netflix Keeps Pulling Ahead Of The Competition

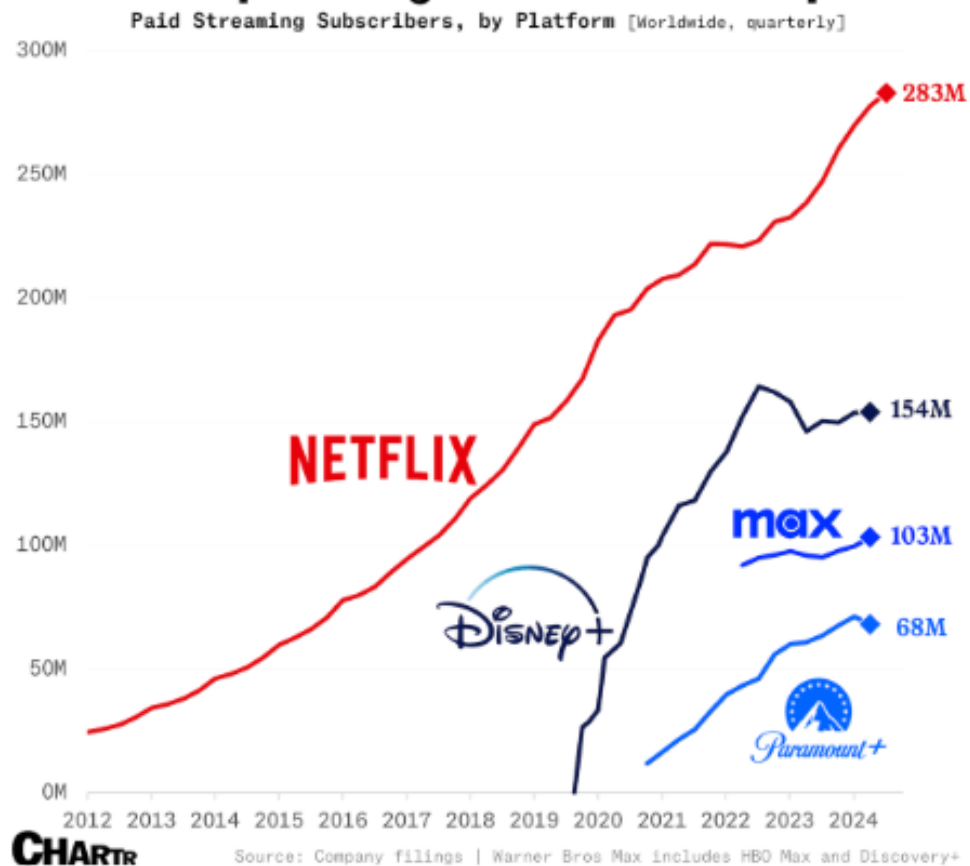


Figure 11: Line chart showing streaming service subscribers by platform.

## A pie chart



Figure 12: A pie chart.



Figure 13: Pie chart showing the amount of pie eaten versus the amount not (yet) eaten.

When they are used with more than two or three segments, it isn't easy to pick out slivers or compare relative segment sizes. A bar chart (Section 2.2) can always be used in place of a pie chart and is much clearer to read.

## Histogram

A histogram might look very similar to a bar chart, but it is fundamentally different since it is plotting numerical rather than categorical data.

Histograms (Figure 14) are used to examine the distribution of a numerical variable. The x-axis contains the value of the numerical variable, which is then binned into ranges, and the frequency of points in the range is displayed on the y-axis. The bars on a histogram should always be displayed as touching, since the variable is continuous.

## Scatterplot

Scatterplots (Figure 15) are used to show the relationship between two numerical variables. Both the x-axis and y-axis contain numerical quantities. There is often a line of best fit added to demonstrate the relationship between the two variables.

## Bubble plot

A bubble plot (Figure 16) is like a scatter plot, but with extra information provided by the bubble size, and in this case colour as well. They are a simple way of adding an extra dimension to a chart.

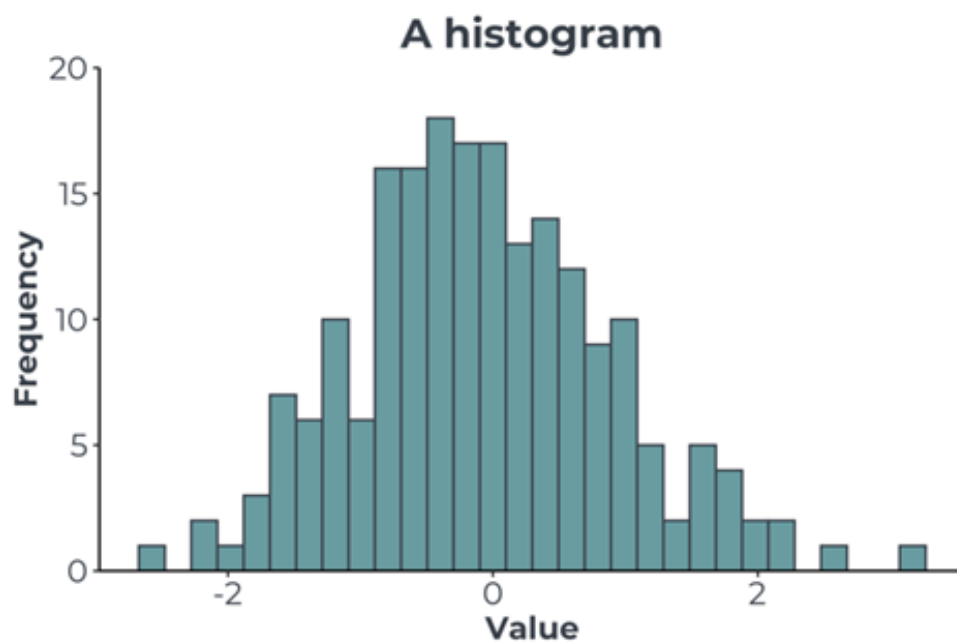


Figure 14: A histogram.

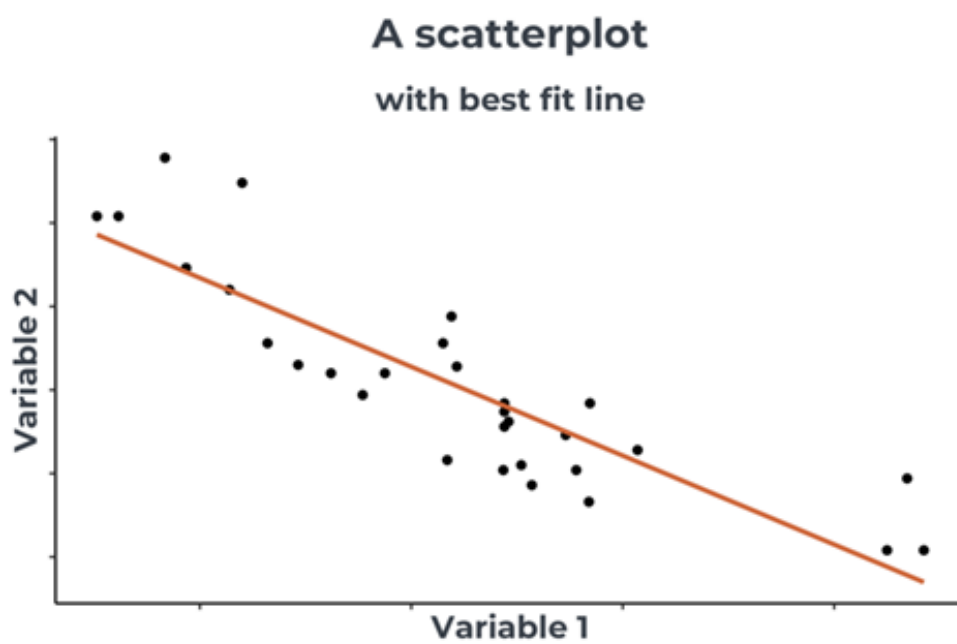


Figure 15: A scatterplot

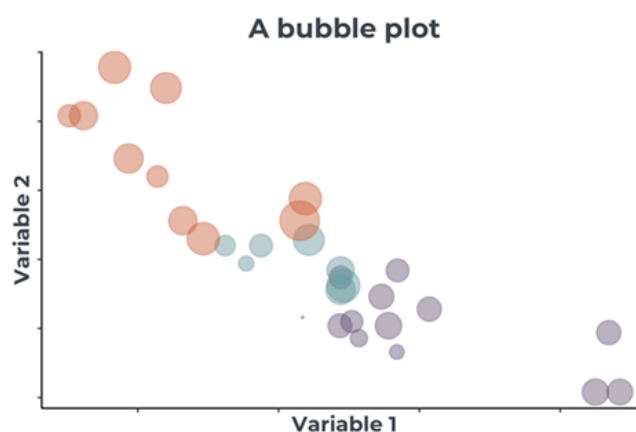


Figure 16: A bubble plot

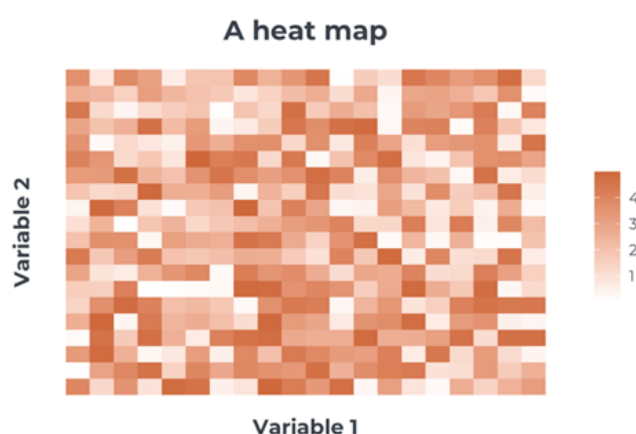


Figure 17: A heat map

## Heat map

A heat map (Figure 17) is able to show patterns between three variables. The first two variables are demonstrated spatially, and the third variable utilises a colour scale. Heat maps are best for identifying spatial patterns rather than reading off accurate values.

## Time series graph

A time series graph (Figure 18) is a special type of line graph, with time on the x-axis and regular repeated measurements of a variable on the y-axis. Time series are good for spotting long term trends, a regular seasonal variation, or even a cyclical variation that doesn't align with the seasons.

## Stacked area chart

An area chart (Figure 19) or a stacked area chart highlights proportions changing over the varying quantity. As well as raw data volumes, this can also be done as proportions of the whole, which is useful for example in demonstrating the change in survey results over time.

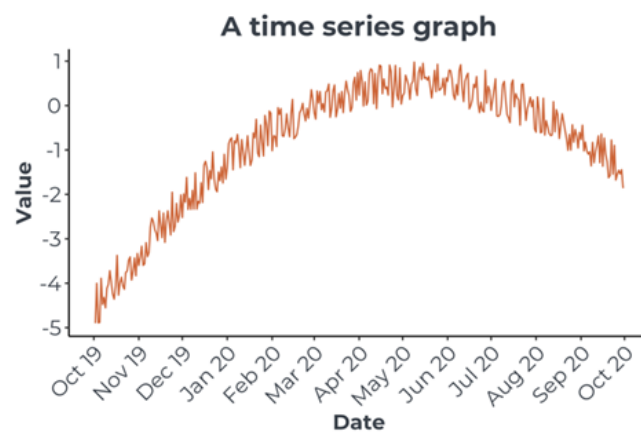


Figure 18: A time series graph.

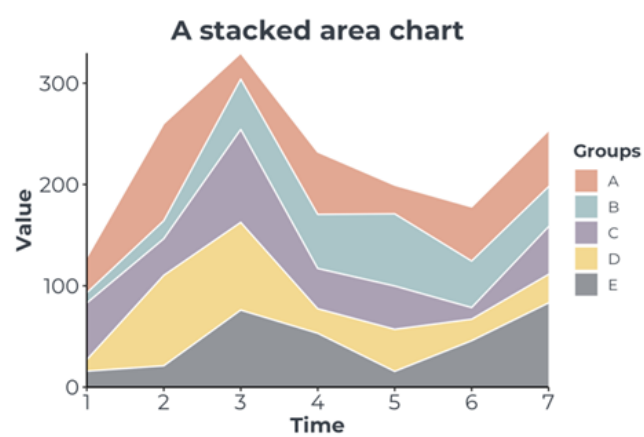


Figure 19: A stacked area chart

## **2c - Describe how data visualisations can be interpreted and misinterpreted.**

Some key elements of reading plots which you should be able to do include:

- Identify the name of the plot
- Interpret the axis
- Identify information
- Look for trends

### *Identifying Factors in Misleading Graphs*

Sometimes data visualizations can be misleading. Some factors that could contribute to this are:

- Proportions not adding up to 100%
- Axes not starting at zero
- Missing data points
- Too many colours or segments

## **2d - State methods of gathering data using a survey.**

- Manual data capture: Face-to-face, telephone, postal, internet.
- Online data gathering tools (such as Google Forms, Microsoft Forms, SurveyMonkey, or Typeform).

### 3 Outcome 3 - Interpret data.

#### 3a - Create and carry out a survey to investigate a problem.

When carrying out a survey, we can use the PPDAC cycle. Below are the steps involved.

##### **Step 1: Identify the Problem**

Before you design your survey, decide on a problem or question you want to investigate. This could be something relevant to your school, community, or daily life. Some examples include:

- What are the most common methods of transport to school?
- How much exercise do students get each week?
- What types of food do students eat at break time?

##### **Step 2: Plan Your Survey**

Once you have a clear problem to investigate, plan how you will collect the data. Consider:

- What information do you need? Think about what questions will give you useful answers.
- Who will you ask? Will you survey your classmates, students from different year groups, or your whole school?
- How will you collect the responses? You can use online tools like Microsoft Forms or Google Forms, or you can collect data manually with paper surveys or tally charts.
- Which questions should you ask? Your survey questions should be clear, concise, and unbiased. Use a mix of:
  - Multiple-choice questions (e.g., "How do you usually travel to school? - Walking, Cycling, Bus, Car, Other")
  - Scale or rating questions (e.g., "On a scale of 1-5, how much do you enjoy school lunches?")
  - Short-answer questions (e.g., "What improvements would you like to see in school lunches?")

##### **Step 3: Collect Data**

- If using an online survey tool such as Microsoft Forms or Google Forms, share the link with your target audience. These tools are useful as they automatically generate visualisations based on the data collected!
- If collecting data manually, make sure to record responses accurately.
- Aim to collect enough responses to make your findings meaningful.

### **Step 4: Analyse Data**

Once you have collected your responses, review the results. You can view the built-in graphs and charts created in Microsoft Forms or Google Forms.

### **Step 5: Draw Conclusions**

- Look for trends in your data. For example, if most students travel to school by bus, what does that tell you about transport options?
- Think about what your data means and how it might be used to make improvements or inform decisions.

### **Present Your Findings**

After collecting your results, you should present your findings to others, such as peers in your class. Some ways you could choose to present your findings are:

- Create a presentation (PowerPoint, Google Slides)
- Write a short report summarising key findings
- Design a poster with key statistics and graphs
- Make a video or audio recording explaining your findings

### **3b - Interpret data visualisations to identify patterns and trends.**

In order to best describe a graph, the following should be done:

- **What is being measured?** - Identify the variables represented on the axes.
- **Quantitative description** - Use numbers and percentages to describe key data points.
- **Descriptive vocabulary** - Use terms like increase, decrease, peak, trend, correlation.
- **Consistency with the data** - Ensure your description accurately reflects the information in the visualization.

[add in the variability which should be thought about in level 5]

### **3c - Describe data visualisations in terms of quality and trust.**

Graphs are everywhere, on the news, on the internet, in reports and publications. Not all graphs are good graphs though. Good graphs convey their message at a glance, whilst bad graphs can be either deliberately misleading or just hard to decipher.

When describing a data visualisation in terms of quality, mentioning whether or not the following features are present can be useful:

### A pie chart that doesn't add up

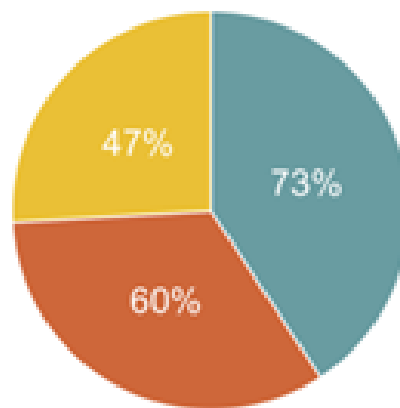


Figure 20: A pie chart that doesn't add up.

- The axes are visible, labelled, and scaled correctly
- Units of measurement are given
- The data is plotted accurately
- There is a legend present
- The graph is overall neat and legible
- There is a title or caption
- There is a trend line shown, if required
- Graph helps answer the question

### Examples of Graphical Crimes

#### Proportions not adding up to 100% (Figure 20)

When plotting proportions of a whole the numbers must always add up to the whole or 100%.

#### Axes not starting at zero

Many graphics types such as bar graphs are interpreted by the reader by comparing the lengths of the different bars. If the bars do not start from zero, then the length comparison is distorted, and patterns can be made to appear that don't actually exist.

#### Missing data points

By choosing only data that fits the creator's objective the reader will not see the full picture. In Figure 21 only half a year of data is shown to imply a trend that doesn't exist in the second half of the year.

#### Too many colours and segments

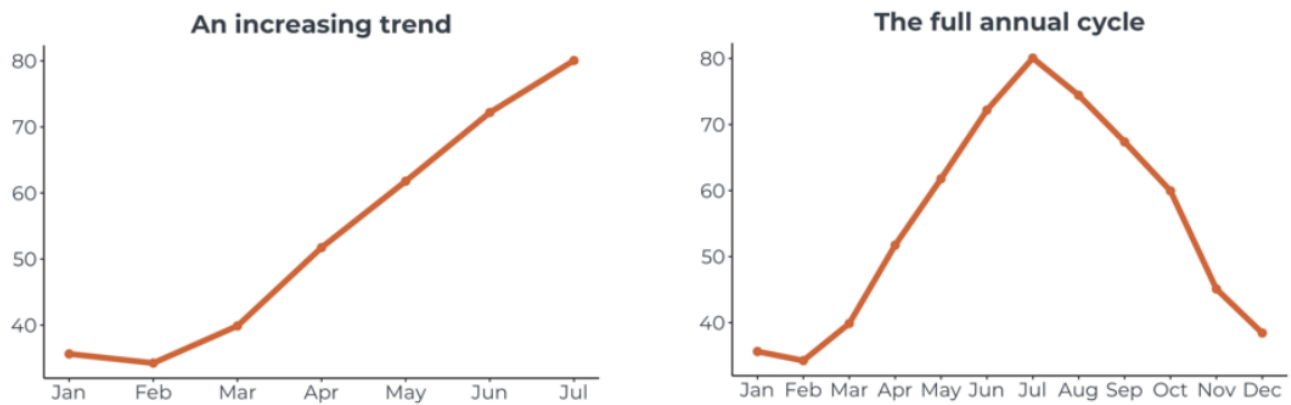


Figure 21: Two line charts. The one on the left has less datapoints than the one on the right.

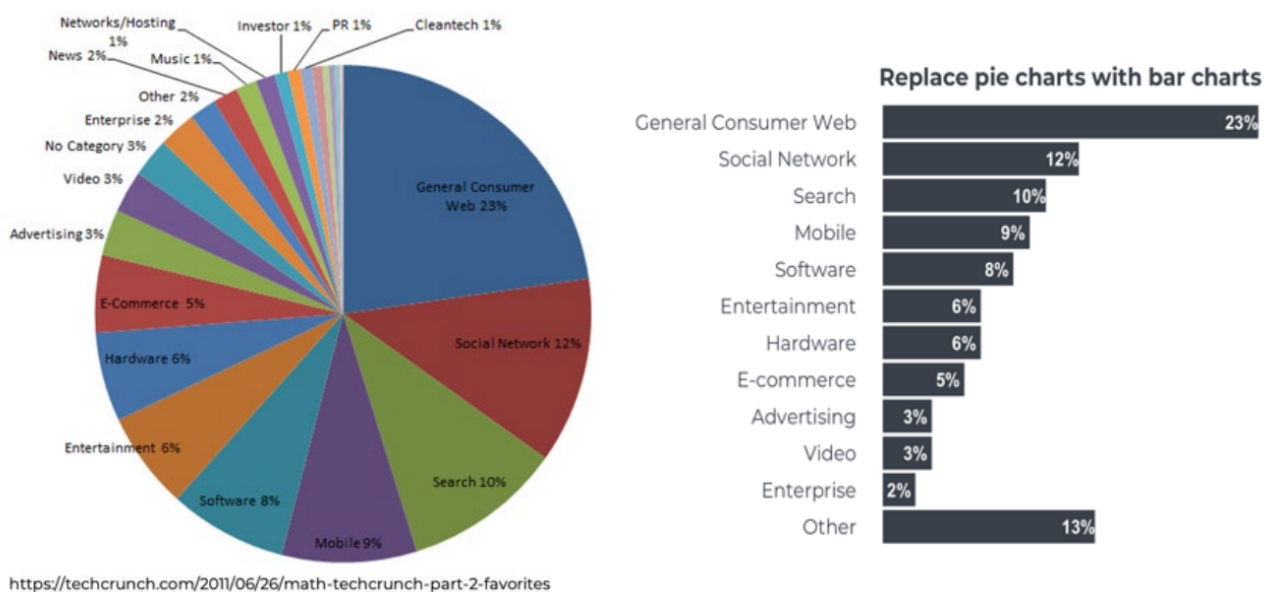


Figure 22: On the left, a pie chart with too many segments. On the right this has been replaced with a bar chart.

Although vibrant, too many colours make a visual that is very hard to interpret. It is best to stick to one or two colours and make use of grey to de-emphasise unimportant patterns.

Pie charts should ideally be replaced with bar charts. If used, they should never have more than 2 or 3 segments. In Figure 22, all the small categories have been merged together.

### 3d - Describe data generated from AI in terms of quality and trust.

When generating output using AI, it's important to be able to think about the output critically, and to be aware that there can be mistakes and inaccuracies in it.

When describing data generated by AI, there are a number of features that can be commented on. Some of these include:

- **Inaccuracies:** For example, in images: the wrong number of fingers in people, other strange artefacts.
- **Relation to Prompt:** If the image or text is not fitting the criteria of what was requested.

- **Misinformation:** Sometimes generative AI will come out with factually wrong things. If unsure, compare the information with another trusted source such as well known news outlets.
- **Bias and Fairness:** Outputs can be biased due to bias in data that the model was trained on.
- **Ethical Considerations:** Sometimes outputs can reinforce harmful stereotypes or messages.

### 3f - Make recommendations based on conclusions and communicate findings.

#### Tips for Drawing Conclusions and Communicating Findings

In order to draw conclusions from data, we should make a claim about what a graph is showing in response to a question or issue. The reasoning used to reach a claim should be clear and logical.

When communicating findings, we should present with an audience in mind (such as peers, family, school management, or community) with a purpose, such as to inform or persuade.

1. Use values (e.g., “The average test score was 75%”).
2. Use visualisations to support findings (e.g., “The bar chart shows that football is the most popular sport among students”).
3. Answer a question based on the data (e.g., “What is the most common age group in the survey?”).

#### Examples of drawing conclusions, communicating findings, and making recommendations.

**Example 1: Figure 23.** The bar chart shows that football has the highest participation with 120 students, while chess has the lowest with 30 students. We can conclude that football is the most popular extracurricular activity among students, indicating a strong interest in team sports. We might suggest that school resources could be allocated to support more football events, and initiatives can be developed to increase interest in less popular activities like chess.

**Example 2: Figure 24.** We can observe that there seems to be a positive correlation between study time and exam performance. This leads us to the potential conclusion that if students study for longer they will achieve a higher score. We might consider studying longer if we would like to achieve a higher grade.

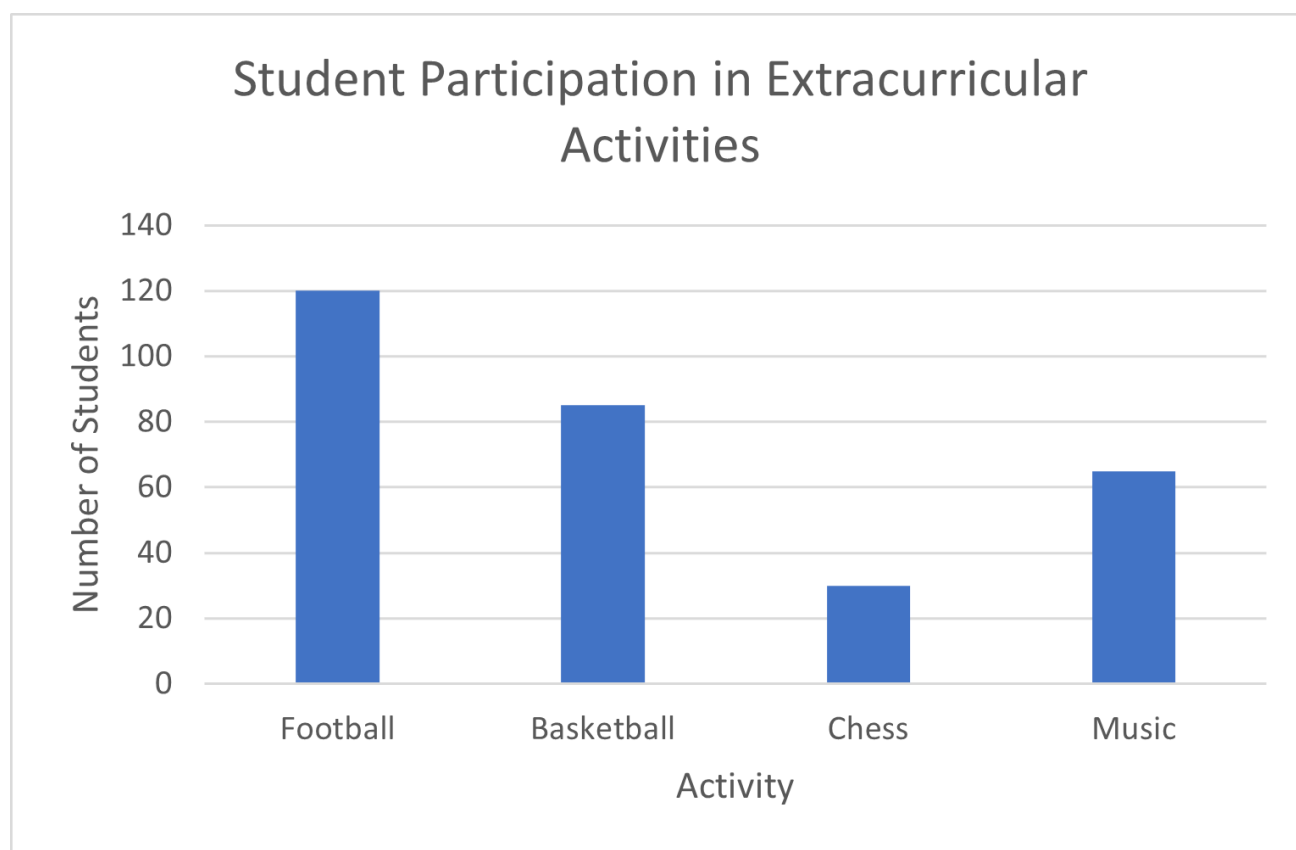


Figure 23: A bar chart showing how many students partake in different extracurricular activities.

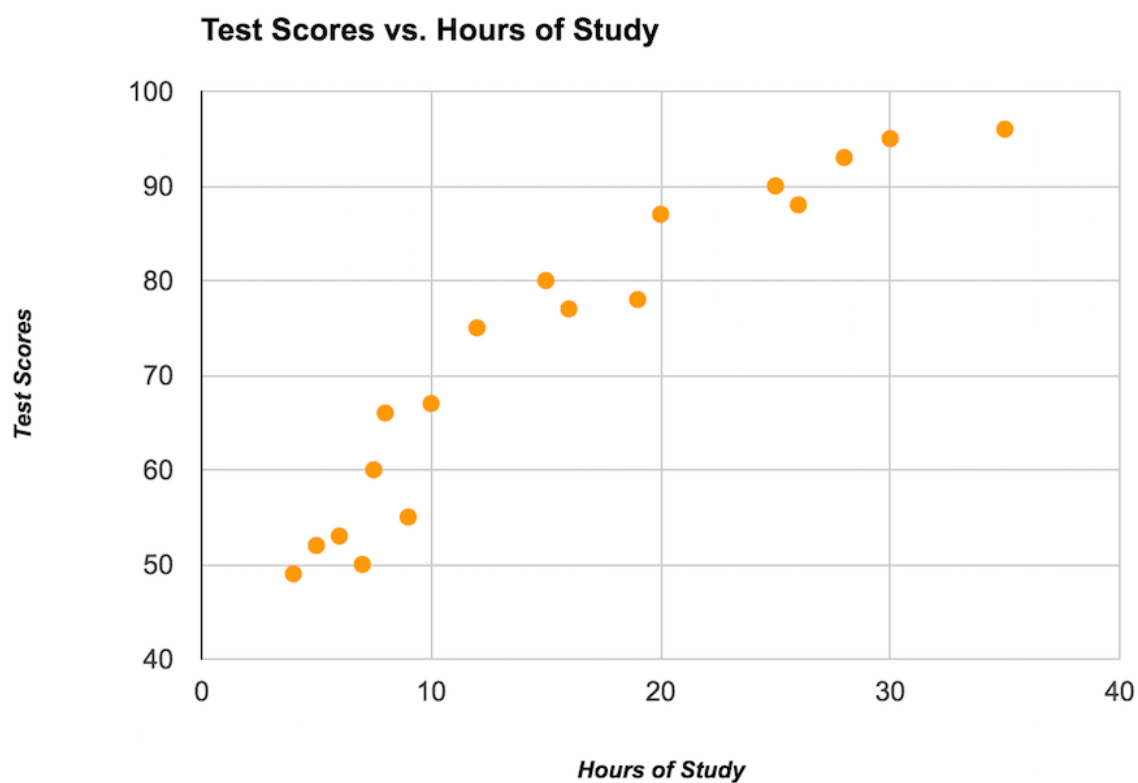


Figure 24: Graph of time studied against exam performance.