"February is the driest month."

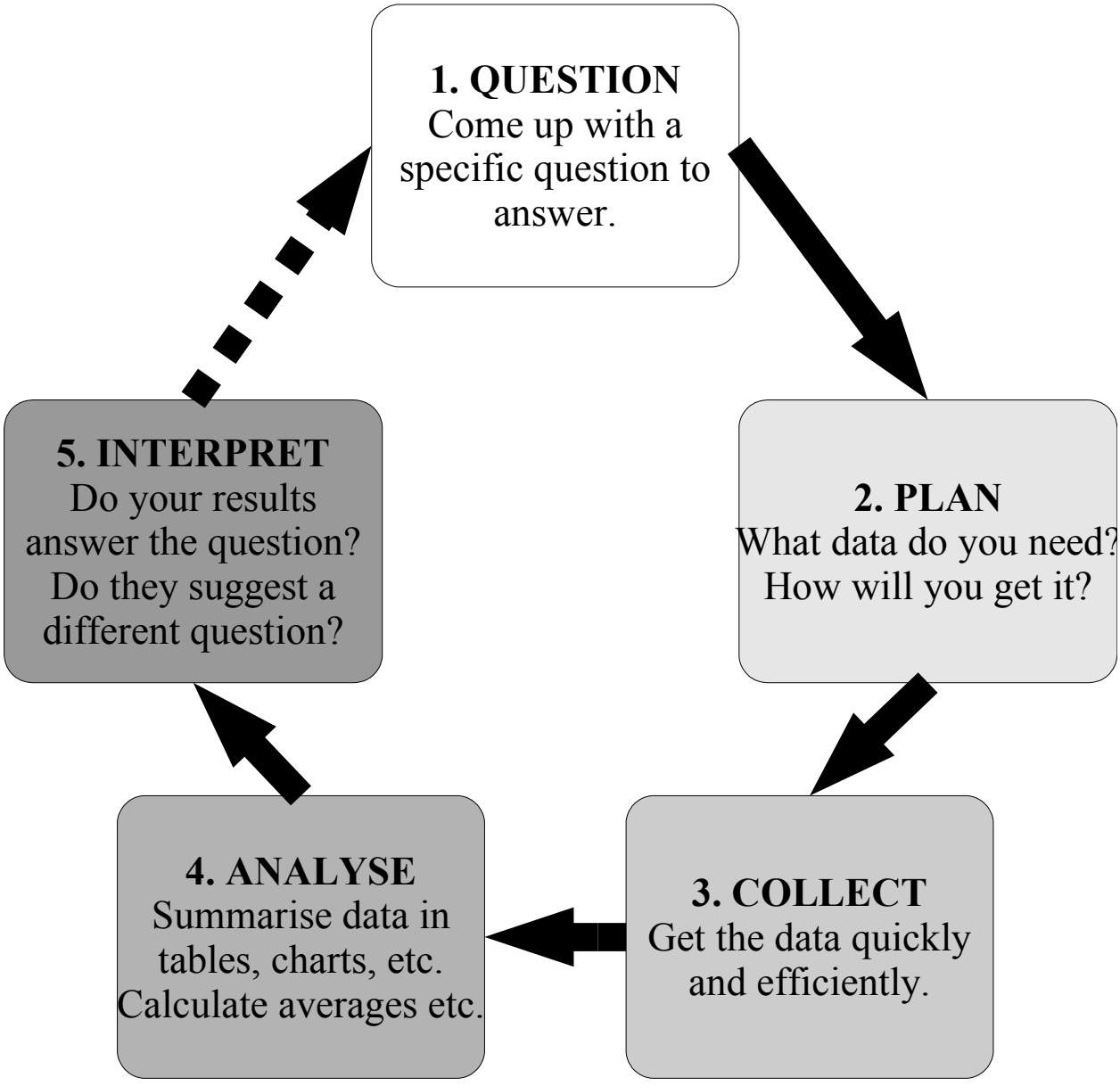
"Whenever I buy a lottery ticket my numbers never come up — I'm sure some numbers come up more often than others."

"Girls are better than boys at estimating someone's height."

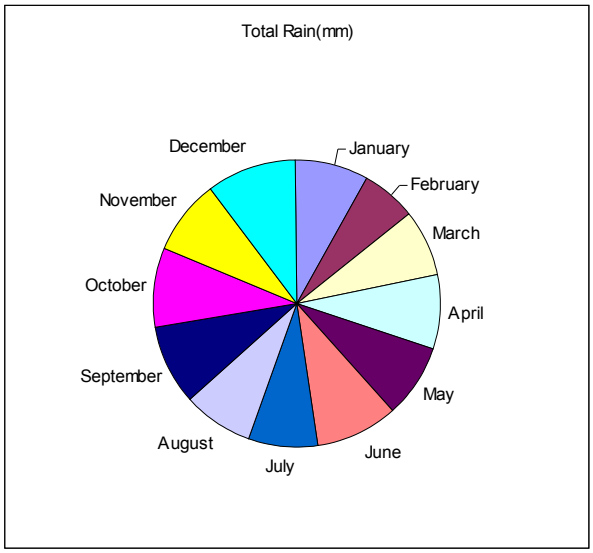
How can you decide if these statements are true?

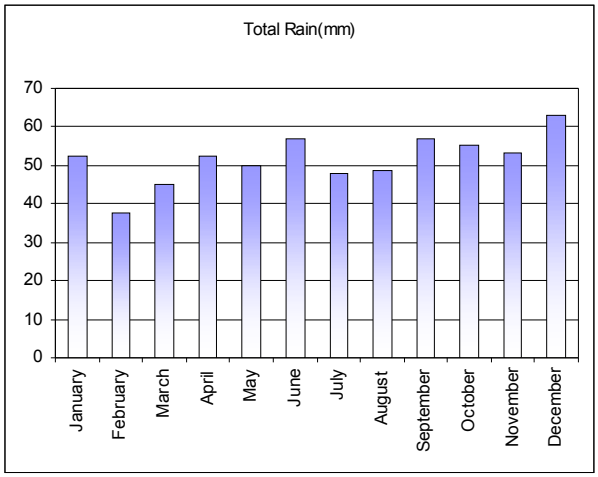
How can you convince someone else?

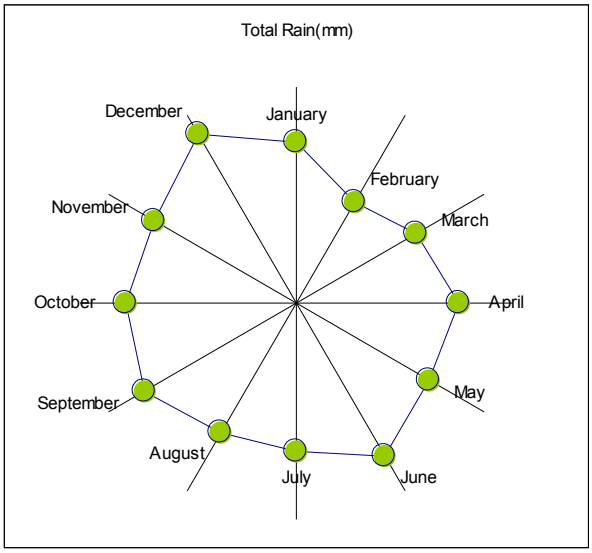
The Data Handling Cycle

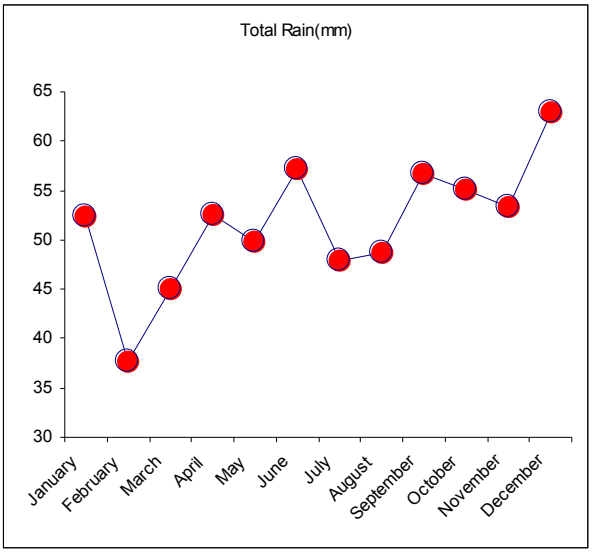


	Mean Temp(°C)	Rain Total (mm)
January	4.2	52.3
February	4.5	37.7
March	7.2	45
April	8.7	52.5
May	12.4	49.8
June	14.6	57.1
July	16.9	47.9
August	16.3	48.7
September	13.7	56.7
October	10.3	55.1
November	6.9	53.3
December	5.3	62.9









Which chart is clearest?

Source: Pitsford Hall Weather Centre, Northampton. Data for 1971-2000.

Aim

This is the general aim of the project.

"I will try to find out if girls' hands are smaller than boys. I will also try to find out if there is a link between a person's height and their hand span."

Hypothesis (plural: Hypotheses)

A specific statement or prediction that you can show to be true or false.

- *"The mean hand span of girls in Year 11 is smaller than the mean for boys."*
- *"There is correlation between the heights and hand spans of people in Year 11."*

Plan

- **What** data do you need?
- **Who** will you get it from?

"I will need to find out the heights, hand spans and shoe size of people in my class. I might need to get more data from another class to get a bigger sample..."

- **How** will you collect it?

"I will get together with other people in my class so we can collect it efficiently. Eric will measure people's hand spans, I will..."

- **How** will you record it?

"I will record the results in a table like this..."

- **How** will you make sure the data is reliable?

"The hand span was measured from the tip of the thumb to the tip of the little finger. The measurement was done twice and if the measurements were different the mean was worked out."

- **Why?** Give reasons for the choices you made.

"We made everyone take their shoes off when we measured the height because..."

Analysis

This is where you do the calculations and draw charts.

- The mean and median are **averages** — they give you a single number to compare groups. The median is less affected by **outliers**. The mode is rarely useful.
- The **range** is a measure of how spread out a group is.
- To see if there is a link between two measurements from the same people (for example height and show size) draw a **scatter graph**.
- **Histograms**, **frequency polygons** and **stem-and-leaf diagrams** are good for comparing two groups (for example boys and girls). They also give you the opportunity to **group** data in a **frequency table** (tally table).
- You can do the calculations and charts on a computer, but you should do one of each type by hand (to prove you can).
- If you make a mistake **leave it in**. This lets you show you have checked your work and corrected it.

Conclusions

- Do your results agree with the hypothesis? How confident are you?
- *"The mean for girls' hand span was lower than the boys, but they were close so I can't be sure there is a real difference."*
- What went wrong? How did you deal with it?
- *"We couldn't get the data for 3 people, but they looked average so I don't think it will affect the results."*
- What would you do differently if you did the experiment again?
- *"I would take more care to measure the height by..."*

Aim

Rewrite in your own words:

"I will try to find out if girls' hands are smaller than boys. I will also try to find out if there is a link between a person's height and their hand span."

Hypothesis (plural: Hypotheses)

Rewrite in your own words:

- *"The mean hand span of girls in Year 11 is smaller than the mean for boys."*
- *"There is correlation between the heights and hand spans of people in Year 11."*

Plan

- **What** data do you need?
- **Who** will you get it from?

"I will need to find out the heights, hand spans and shoe size of people in my class. I might need to get more data from another class to get a bigger sample..."

- **How** will you collect it?

"I will get together with other people in my class so we can collect it efficiently. Eric will measure people's hand spans, I will..."

- **How** will you record it?

"I will record the results in a table like this..."

- **How** will you make sure the data is reliable?

"The hand span was measured from the tip of the thumb to the tip of the little finger. The measurement was done twice and if the measurements were different the mean was worked out."

- **Why?** Give reasons for the choices you made.

"We made everyone take their shoes off when we measured the height because..."

Analysis

This is where you do the calculations and draw charts.

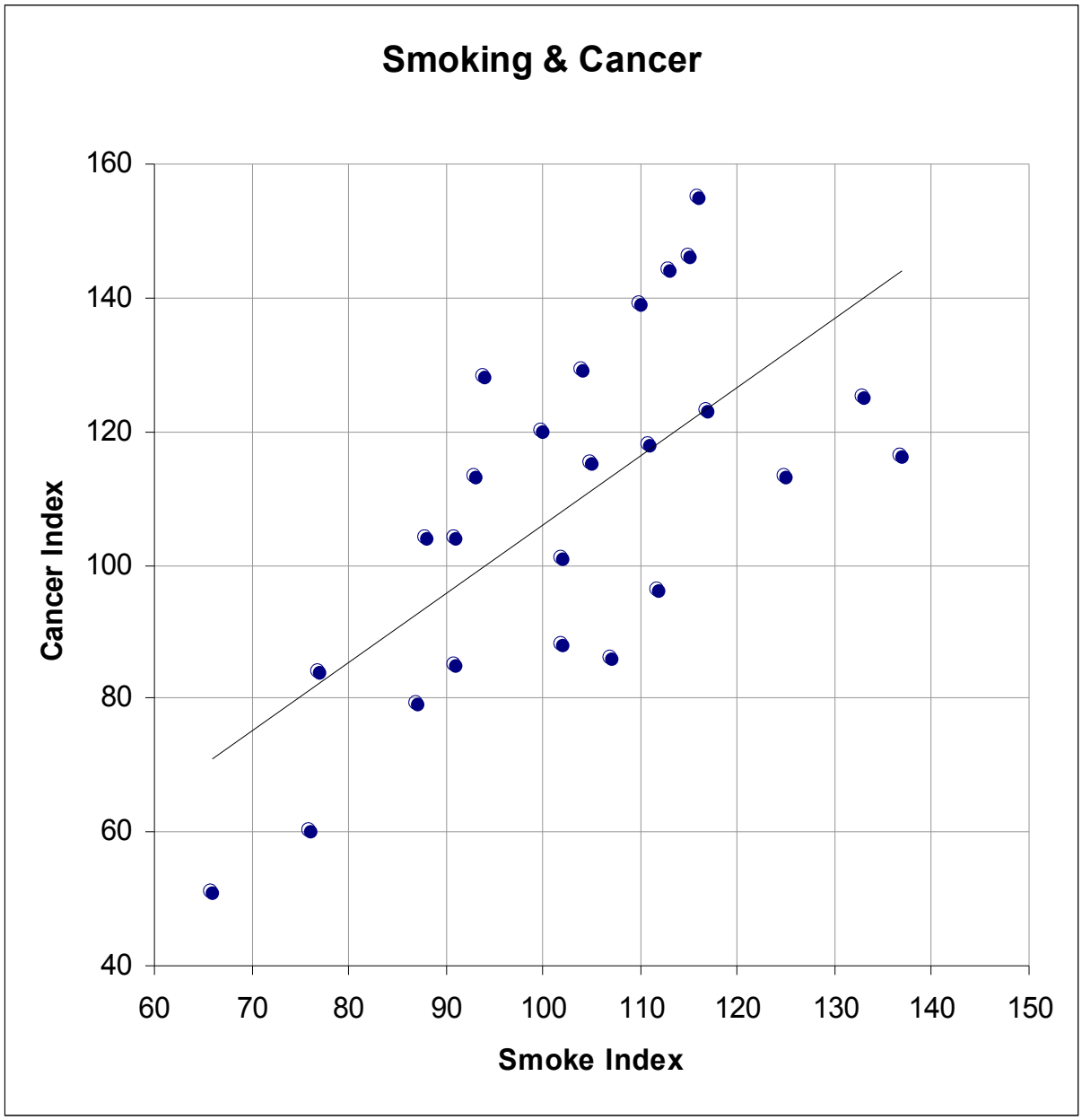
- The mean and median are **averages** — they give you a single number to compare groups. The median is less affected by **outliers**. The mode is rarely useful.
- The **range** is a measure of how spread out a group is.
- To see if there is a link between two measurements from the same people (for example height and show size) draw a **scatter graph**.
- (optional) **Histograms**, **frequency polygons** and **stem-and-leaf diagrams** are good for comparing two groups (for example boys and girls). They also give you the opportunity to **group** data in a **frequency table** (tally table).
- You can do the calculations and charts on a computer, but you should do one of each type by hand (to prove you can).
- If you make a mistake **leave it in**. This lets you show you have checked your work and corrected it.

Conclusions

- Do your results agree with the hypothesis? How confident are you?

"The mean for girls' hand span was lower than the boys, but they were close so I can't be sure there is a real difference."

- What went wrong? How did you deal with it?
- *"We couldn't get the data for 3 people, but they looked average so I don't think it will affect the results."*
- What would you do differently if you did the experiment again?
- *"I would take more care to measure the height by..."*


Smoking & Cancer
Cancer Index
160
140
120
100
80
60
40
60
70
80
90
100
110
120
130
140
150
Smoke Index

Bias

Bias is anything that distorts your data.

Continuous Data

This is data that can take *any* value in a range. Example: A person's height is continuous. It can be measured as accurately as you like. A height of 1.732455432m is possible.

Discrete Data

Data that can take only certain values. Example: A person's shoe size is discrete. It goes up in jumps. You could have a shoe size of 6 or 6½ but not 6.4325454.

Hypothesis

This is the question you are trying to answer. Your investigation will **test the hypothesis**.

Population

The group that your question is about. Example: If you were investigating if fish in a lake had a lot of disease then the population would be all the fish in the lake.

Primary Data

This is data you have collected yourself.

Protocol

A detailed description of how the data is collected.

Reliable

Data is **reliable** if you would get much the same results if you repeated the experiment.

Sample

If you wanted to find out if men were taller than women it would be impractical to measure everyone in the world. Instead you would only measure some of them — this is a sample. It is important to make sure the sample is **representative**.

Secondary Data

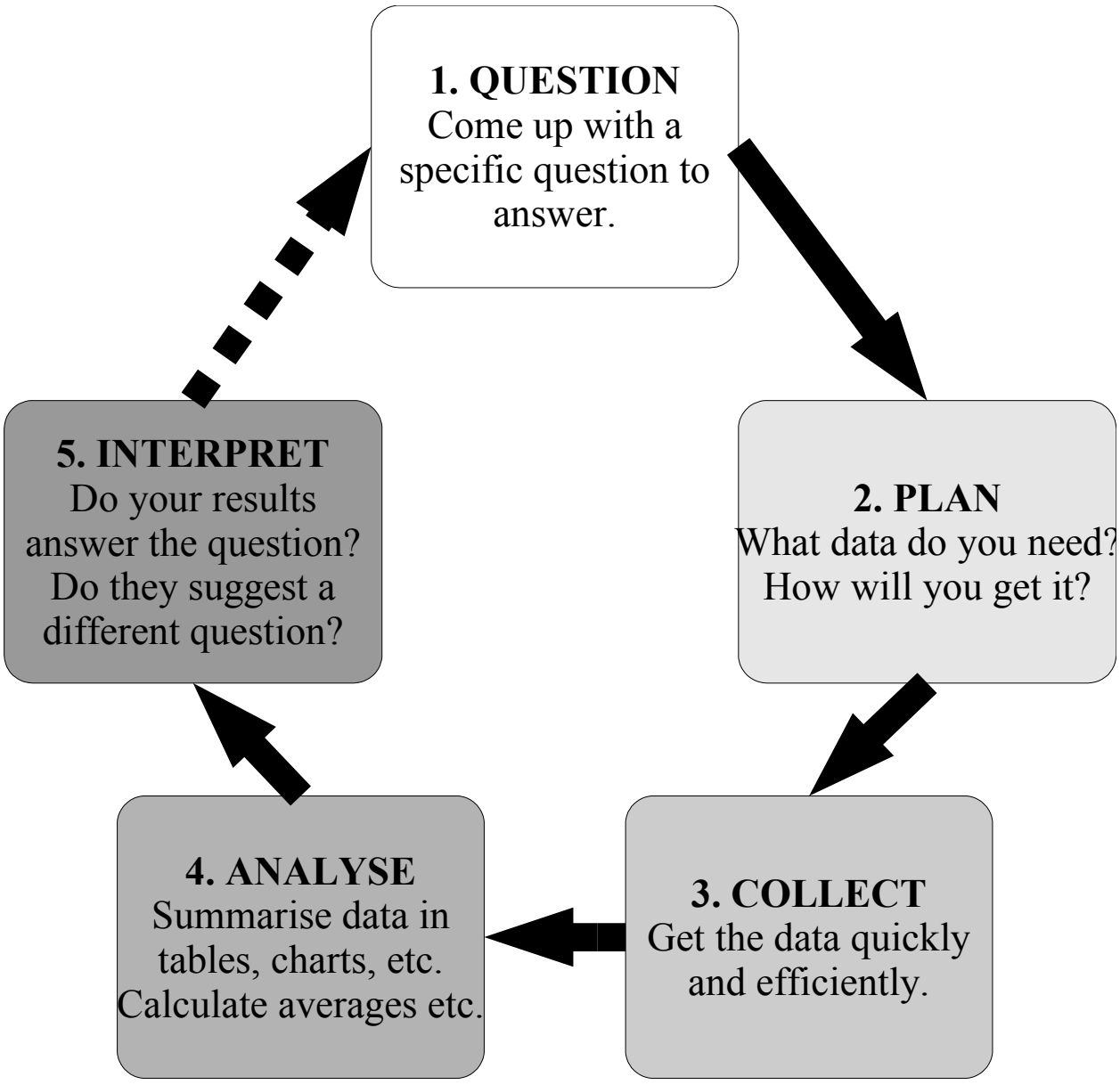
This is data that someone else has collected. You might get it from a reference book, a newspaper, a web site (or your teacher).

Valid

Data is **valid** if it truly represents what it claims to represent.

- "February is the driest month."
- "Whenever I buy a lottery ticket my numbers never come up — I'm sure some numbers come up more often than others."
- "Girls are better than boys at estimating someone's height."

These questions could each be investigated by statistical means.



- "Whenever I buy a lottery ticket my numbers never come up — I'm sure some numbers come up more often than others."

1.Come up with a **specific** question that can be answered.

2.What data do you need. Where will you get it?

3.Collect the data.

4.Analyse the data. Calculations and diagrams.

5.Interpret the data. Does it answer the question?

Secondary Data is data that you have not collected yourself. This is an exercise in using secondary data.

Evidence that smoking is linked to cancer first started to appear in the 1950's. This table is from one of the first research studies. Can you show the link?

Occupation	Smoke Index	Cancer Index
Farmers, foresters, and fisherman	77	84
Miners and quarrymen	137	116
Gas, coke and chemical makers	117	123
Glass and ceramics makers	94	128
Furnace, forge, foundry, and rolling mill workers	116	155
Electrical and electronics workers	102	101
Engineering and allied trades	111	118
Woodworkers	93	113
Leather workers	88	104
Textile workers	102	88
Clothing workers	91	104
Food, drink, and tobacco workers	104	129
Paper and printing workers	107	86
Makers of other products	112	96
Construction workers	113	144
Painters and decorators	110	139
Drivers of stationary engines, cranes, etc.	125	113
Labourers not included elsewhere	133	125
Transport and communications workers	115	146
Warehousemen, storekeepers, packers, and bottlers	105	115
Clerical workers	87	79
Sales workers	91	85
Service, sport, and recreation workers	100	120
Administrators and managers	76	60
Professionals, technical workers, and artists	66	51

- What calculations might you do?
- What charts might you draw?

- "Whenever I buy a lottery ticket my numbers never come up — I'm sure some numbers come up more often than others."

1.Come up with a **specific** question that can be answered.

2.What data do you need. Where will you get it?

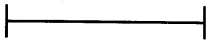
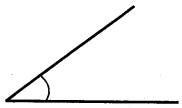
3.Collect the data.

4.Analyse the data. Calculations and diagrams.

5.Interpret the data. Does it answer the question?

Choose one of these tasks. You should try to extend the task if you can.

Sarah asked a sample of people to estimate

- the length of this line 
- the size of this angle 

Sarah then said that people estimate the length of lines better than the size of angles.

- Write down a hypothesis to test how well people estimate
- Design and carry out an investigation to test your hypothesis

Grandad told Simon that some people have slower reactions than other people.

Simon decided to test the reaction times of some of his friends.

- Write down a hypothesis for him to test
- Design and carry out an investigation to find out different ways in which reaction times can be affected

Not everyone has the same pulse rate – and pulse rate can be affected by a number of different things.

- Write a hypothesis about how someone's pulse rate can be affected
- Design and carry out an investigation to show different ways in which pulse rate can be affected

Charlie thinks that girls have smaller hands than boys.

- *Write down a hypothesis for him to test.*
- *Design and carry out an experiment to test your hypothesis.*

- "Whenever I buy a lottery ticket my numbers never come up — I'm sure some numbers come up more often than others."

1.Come up with a **specific** question that can be answered.

2.What data do you need. Where will you get it?

3.Collect the data.

4.Analyse the data. Calculations and diagrams.

5.Interpret the data. Does it answer the question?

Bonita thinks that the girls in her Year 10 class have big hands.

She measures their hand span. Here are her results (in cm):

14.1	13.8	16.2	13.9	14.3
18.2	17.0	15.8	14.8	16.3

Bonita says:

"The mean hand span is 14.2cm. I was right. The girls have big hands."

Do you agree with Bonita?

Task

Design and carry out a survey to investigate hand spans. You may want to extend the investigation.

- Do boys have bigger hands than girls?
- Is hand span related to age?
- What other things might be related to hand span?
- What data do you need to collect?
- How can you make sure the data is collected fairly?
- Will you need to cooperate with others to get the data?
- How will you make the measurements?
- How will you record the data?

Advice on planning

To get good marks you need to:

- Record in detail *everything* you will do.
- Give reasons for your decisions.
- Show you have though about any problems that might come up and how you will cope with them.
- Have clear **aims** and **hypotheses**.

Aim

This is the general aim of the project.

"I will try to find out if girls' hands are smaller than boys. I will also try to find out if there is a link between a person's height and their hand span."

Hypothesis (plural: Hypotheses)

A specific statement or prediction that you can show to be true or false.

- *"The mean hand span of girls in Year 11 is smaller than the mean for boys."*
- *"There is correlation between the heights and hand spans of people in Year 11."*

Plan

- **What** data do you need?
- **Who** will you get it from?

"I will need to find out the heights, hand spans and shoe size of people in my class. I might need to get more data from another class to get a bigger sample..."

- **How** will you collect it?

"I will get together with other people in my class so we can collect it efficiently. Eric will measure people's hand spans, I will..."

- **How** will you record it?

"I will record the results in a table like this..."

- **How** will you make sure the data is reliable?

"The hand span was measured from the tip of the thumb to the tip of the little finger. The measurement was done twice and if the measurements were different the mean was worked out."

- **Why?** Give reasons for the choices you made.

"We made everyone take their shoes off when we measured the height because..."

Analysis

This is where you do the calculations and draw charts.

- The mean and median are **averages** — they give you a single number to compare groups. The median is less affected by **outliers**. The mode is rarely useful.
- The **range** is a measure of how spread out a group is.
- To see if there is a link between two measurements from the same people (for example height and show size) draw a **scatter graph**.
- **Histograms**, **frequency polygons** and **stem-and-leaf diagrams** are good for comparing two groups (for example boys and girls). They also give you the opportunity to **group** data in a **frequency table** (tally table).
- You can do the calculations and charts on a computer, but you should do one of each type by hand (to prove you can).
- If you make a mistake **leave it in**. This lets you show you have checked your work and corrected it.

Conclusions

- Do your results agree with the hypothesis? How confident are you?
- *"The mean for girls' hand span was lower than the boys, but they were close so I can't be sure there is a real difference."*
- What went wrong? How did you deal with it?
- *"We couldn't get the data for 3 people, but they looked average so I don't think it will affect the results."*
- What would you do differently if you did the experiment again?
- *"I would take more care to measure the height by..."*

General Comments

Each student is required to produce two pieces of coursework — one on *Using and Applying Mathematics* (this will be done in Year 11) and this one on *Handling Data*. Each piece of coursework contributes 10% of the final GCSE marks. The maximum mark for each piece of coursework is 24. Students are expected to plan, conduct and analyse a statistical investigation. The deadline for this piece of coursework is _____. (Two weeks of lesson time plus homework — the exam board recommends 10 hours should be spent on each coursework.)

Lesson 1 (Planning)

At the end of lesson 1 you should have:

- Written down clear aims for your investigation
- **Come up with one or more precise hypotheses**
- Started to make notes on the data you need and how you will collect it

Lesson 2 (Planning)

At the end of lesson 2 you should have:

- Written down what data you need
- **Have a clear plan for collecting and recording the data** — possibly data collection sheets
- Possibly conducted pre-tests or a pilot study, or chosen a **sample**.

Lesson 3 (Data Collection)

At the end of lesson 3 you should have:

- **Collected the data you need**
- Recorded the data in a form that can be shared with others

Lessons 4-6 (Analysis)

You should spend these lessons **analysing the data** — in order of difficulty (and marks!):

- Calculate averages and range
- Group the data to produce histograms or pie charts
- Calculate averages from the grouped data
- Stem-and-leaf diagrams to compare groups
- Produce accurate scatter graphs
- Produce cumulative frequency tables and graphs
- Use cumulative frequency graphs to estimate the median, quartiles and interquartile range
- Produce box plots

Lessons 7-8 (Conclusions & Reflection)

You should spend these lessons using the results of the analysis to complete your investigation:

- Comment on your results:
  - **What do your calculations and charts show?**
  - Is there correlation in the scatter graphs?
  - **Is there enough data for you to be confident?**
  - Have you compared different groups?
  - **Do your results support the hypotheses?**
- Is your report set out in a clear and logical way? Have you checked for errors?
- **What problems did you face?** How did you cope with them?
- What would you do differently if you repeated the investigation?
- What new questions are raised by your results?

⚠ You should spend time carefully planning the investigation — this is where many people lose a lot of marks.

General Comments

Each student is required to produce two pieces of coursework — one on *Using and Applying Mathematics* (this will be done in Year 11) and this one on *Handling Data*. Each piece of coursework contributes 10% of the final GCSE marks. The maximum mark for each piece of coursework is 24. Students are expected to plan, conduct and analyse a statistical investigation. The deadline for this piece of coursework is _____. (Two weeks of lesson time plus homework — the exam board recommends 10 hours should be spent on each coursework.)
It is hoped that each class will measure and record (at least) these data from their class: hand span, height, shoe size and sex to be shared with the other classes.

Lesson 1 (Planning)

Resources: worksheet CW1, and possibly rulers & calculators.
Go through Bonita's experiment with the class and invite discussion on whether Bonita was right. Some points to consider:

- Why did she calculate the mean? Has she done so correctly?
- What does "big hands" mean? To what are the girls' hand spans being compared?
- Is there enough data to come to a meaningful conclusion? Is her class typical?
- How was hand span measured? What equipment was used? Why did she record results to the nearest 0.1cm rather than the nearest cm?
- What other information would she need?

They should then consider their own investigation, and should be encouraged to look beyond hand spans. It is essential that two or more variables are involved if they are to get good marks. Students should then come up with the aims of their investigation (e.g. "To find out if girls have bigger hands than boys", "Tall people have bigger feet") and crystallise these into more precise hypotheses (e.g. "The mean hand span of girls is greater than that of boys", "There is positive correlation between height and hand span"). You may want to introduce worksheet DATA8 to give a structure to their investigation.
Outcomes: Students should have written down their aims and hypotheses, and started making notes on how they will collect the data they need.

Lesson 2 (Planning)

Students should spend this lesson coming up with a detailed plan for data collection. They should consider any problems that might arise in the data collection and describe the measures they would take to minimise them (e.g. "When the height is measured we will make sure everyone takes their shoes off").
Students should consider how to record the results they got — tables are a good idea.
They may also conduct *pre-tests* (where they test their techniques with a few people).
Outcomes: Students should have a clear plan. They should have described in detail what data they need and how it would be collected. They may have designed a data collection sheet or have conducted pre-tests or a pilot survey.

Lesson 3 (Data Collection)

Students collect and record the data. Efficiency & teamwork are key here.

Lessons 4 Onwards (Analysis and Interpretation)