STRAND 1 REVIEW

Working through these questions will help you assess your understanding of the learning outcomes listed here:

Level	All
Learning outcomes	– list outcomes of an experiment – apply the fundamental principle of counting – apply the principle that in the case of equally likely outcomes the probability is given by the number of outcomes of interest divided by the total number of outcomes (examples using coins, dice, spinners, urns with coloured objects, playing cards, etc.) – recognise that, if an experiment is repeated, there will be different outcomes and that increasing the number of times an experiment is repeated generally leads to better estimates of probability – use stem and leaf plots and histograms (equal intervals) to display data – evaluate the effectiveness of different displays in representing the findings of a statistical investigation conducted by others – use a variety of summary statistics to describe the data: central tendency – mean, median, mode; variability – range – evaluate the effectiveness of different displays in representing the findings of a statistical investigation conducted by others – recognise that correlation is a value from -1 to +1 and that it measures the extent of the linear relationship between two variables

Q. A group of people was asked "***What is your blood type***?" Here is the data they gave.

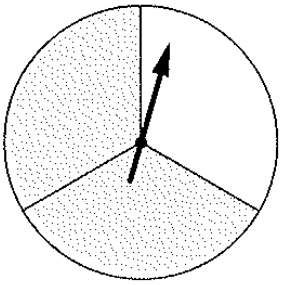
Type A	**Type B**	**Type O**	**Type AB**
50	65	70	15

If a person from this group is selected at random, what is the probability that this person has type **O** blood?

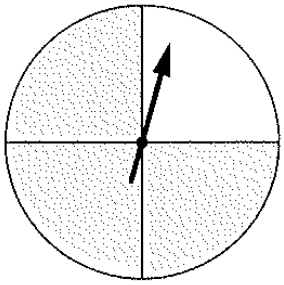
How many people answered the question?
How many people have type O blood?
Remember probability is always a number between **0** and **1.** This means it is a **fraction.** You should write fractions in their **lowest terms.**

Q. Five fair spinners are shown below.

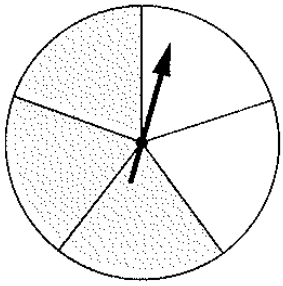
Each spinner is divided into equal sectors, which are coloured either grey or white.



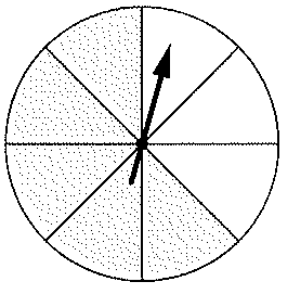
A



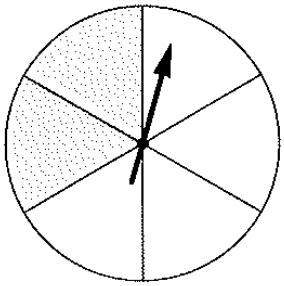
B



C



D



E

a) Identify the spinner for which the probability of spinning grey is ¾.

b) For two of the spinners, the probability of spinning grey is **more than 60%** but **less than 70%.** Which two spinners are these?

a) If the probability is ¾ what does this mean?

What does the 3 represent? What does the 4 represent?

Can you write ¾ in a different way?

Is 6/8 the same as ¾? Why? Why not?

If a student said the probability of spinning grey was 6/8 what might the spinner look like?

Would the student be correct in saying the probability of spinning grey was 6/8? Why? Why not?

b) Represent 60% and 70% as fractions.

Now work out the probabilities of spinning grey on each spinner.

Can you answer the question now?

Q Two coins are tossed. Complete the diagram to show all the **possible outcomes**.

a) What is the probability of getting 2 heads?

	H	T
T		
H		

b) Jennifer tossed the two coins 50 times and got a head and a tail 28 times.
Is there reason for Jennifer to think that one of the coins is not fair?
Explain.

c) Describe an experiment that would allow Jennifer to determine whether or not the coin was fair.

a) Can you make sense of the diagram? Does it help you to keep track of all the **possible outcomes**? How many possible outcomes are there?

b) Is it **more likely** that you get two heads than two tails? Why? Why not? Is it **more likely** that you get a head and a tail? Why? Why not? If you tossed the coins four times how many times would you expect to get a head and a tail? Why?

c) What would Jennifer have to do? How many times should she throw the dice – twice? 3 times? 100 times?

Remember the learning outcome

Students should be able to

- recognise that, if an experiment is repeated, there will be different outcomes and that increasing the number of times an experiment is repeated generally leads to better estimates of probability

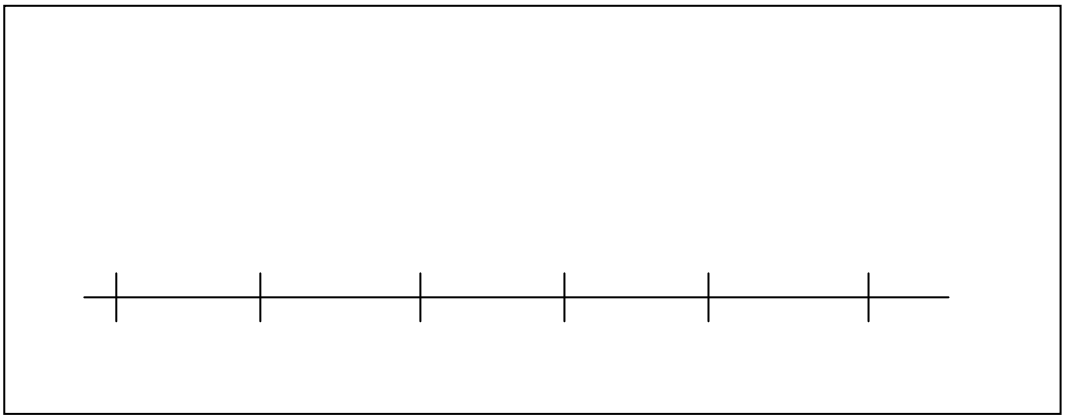
Try this experiment yourself

Q. The table below is a record of the number of texts sent by a group of students in one month.

No of texts sent	0 - 50	50 - 100	100 - 150	150 - 200	200 - 250
Number of students	10	15	25	18	8

a) How many students are in the group?

Illustrate the data on a histogram.



b) Using the table and /or histogram to help you estimate, complete this sentence:

On average these students send about _____ texts each month.

c) Sarah is in the group and she sends 210 texts every month. Describe in one sentence Sarah's text sending by comparison to the others in the group.

What is unusual about the way the data is displayed in the table?

If John sends 100 texts in a month where in the table would you enter his data?

What does a **histogram** look like?

What do you think is the **typical** amount of texts sent by students in this group? Does your histogram help you decide on what is the typical amount of texts sent by students in this group?

Q A teacher asked 21 students to estimate the height of a building in metres.

The stem-and-leaf diagram shows all 21 results

6	5	represents 6.5 m	

6	5	9				
7	0	2	6	8	8	
8	3	3	5	7	7	9
9	0	5	5	5		
10	4	8				
11	2	7				

a) What is the range of the estimated values?

b) What was the median estimated height?

c) The height of the building was 9.2m .How many people overestimated the height?

What other information can you get form the **stem and leaf plot**?

Is there any **evidence** to suggest that the group are good at estimating building height?

Q. Carol opened a new sandwich bar. She offers a lunch special consisting of a sandwich and a drink for €5.

The different choices available are shown below

Type of bread	Filling	Drinks
Brown	Salad	Tea
White	Egg	Coffee
Wrap	Meat	Hot Chocolate
Panini		Cold drink

All of the different combinations are possible. For example, you can order a salad sandwich on brown bread and a coffee.

How many different lunch specials are possible?

Think of a way to organise your thoughts. Can you write out all the possible combinations? Can you see a pattern as you write out all the combinations? Can you **generalise** this pattern that will help you to find out how many combinations there are without writing them all out?

Q. The lists of test results for two maths classes were posted on the college notice board. You do not know which of the lists is for your class.

List 1	List 2
75	92
80	85
83	87
46	91
35	85
27	81
95	89
84	88
65	87
76	88
15	90
100	92
23	87
20	6
15	0

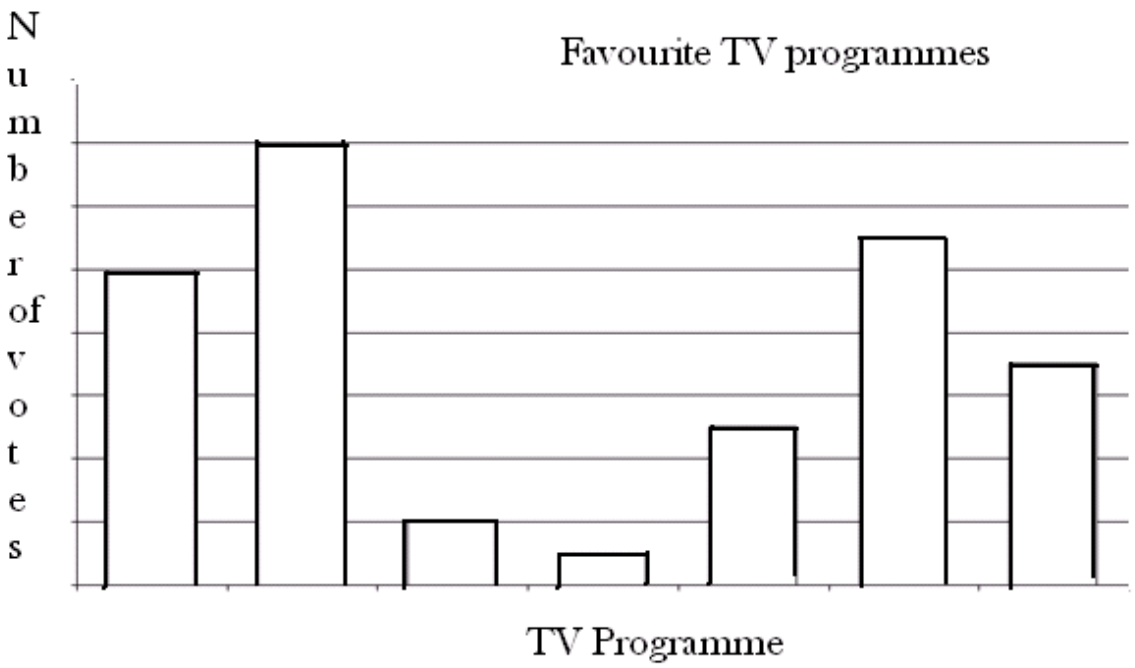
- Display the data from each list in stem and leaf plots.
- Give one reason why you would hope that list 1 is for your class and one reason why you would hope that list 2 is for your class
- Which list represents the better results? Give a reason for your answer.

Q.

Think about what mark you would be hoping to get in the test. Is it **likely** that you would get this mark if your class results were on list 1? List 2?

What is the **typical** mark on list 1? On list 2?

Would you like to get 100? How likely is it that you would get 100 if your class results were on list 1? On list 2? What does it mean to have the **better results**? Is there any evidence that list 1 has **better** results than list 2? Is there any evidence that list 2 has **better** results than list 1?



Clues

- Coronation Street was the most popular TV show
- Twice as many liked Coronation Street as Eastenders
- Fair City got 4 votes less than Coronation Street
- Casualty was the second most popular TV show
- Primetime got 4 votes more than Frontline
- 5 voted for Primetime
- Some people voted for Desperate Housewives

Use the information above to complete the frequency table

TV Programme	No of Votes

Place Coronation street first then Casualty

The bar representing Eastenders must be half the size of the bar representing Coronation Street. Why is this?

Can you locate the bar representing Eastenders?

How will you decide which are the bars representing Primetime and Frontline?

What about the bars representing Desperate Housewives and Fair City?

Task

In 1999 a university librarian put a number of measures in place to try to stop students “stealing” books from the library.

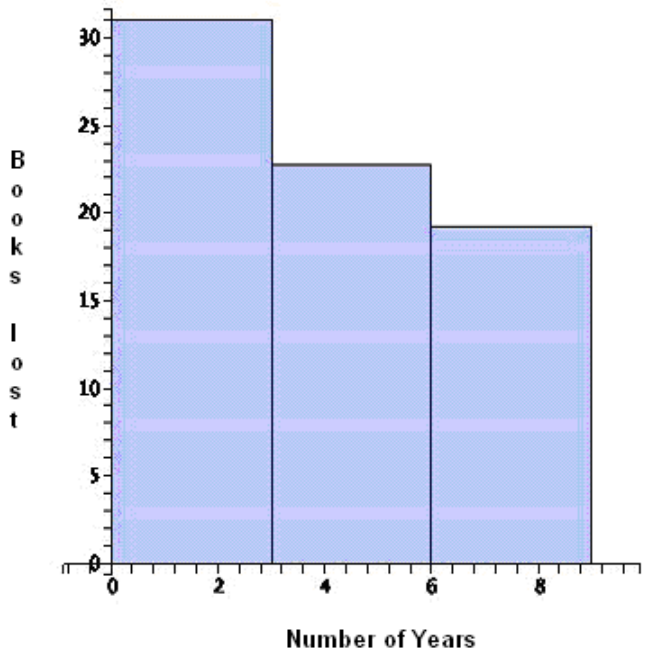
To see how effective these measures were she recorded the number of non-returned books over the next number of years.

The data is recorded below

Year	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
No of non-returned books	9	10	9	14	6	3	4	5	10

When asked to report to the budget committee on book loss she wrote:

Whilst the drain on resources due to lost books is significant, the histogram below shows that over the last nine years the number of books lost to the library is steadily decreasing, which suggests that the measures implemented to combat this practice are working.



The finance officer was not convinced that the measures were working.

Plot the same data in a histogram but, instead of using three year intervals like the librarian did, divide the data into nine intervals, one for each of the last nine years.

Now, use your histogram to write two statements about the trend.
Does your histogram support the librarian's view that the measures are working, or does it lend more support to the doubts of the finance officer?
Explain your reasoning.

Note to student

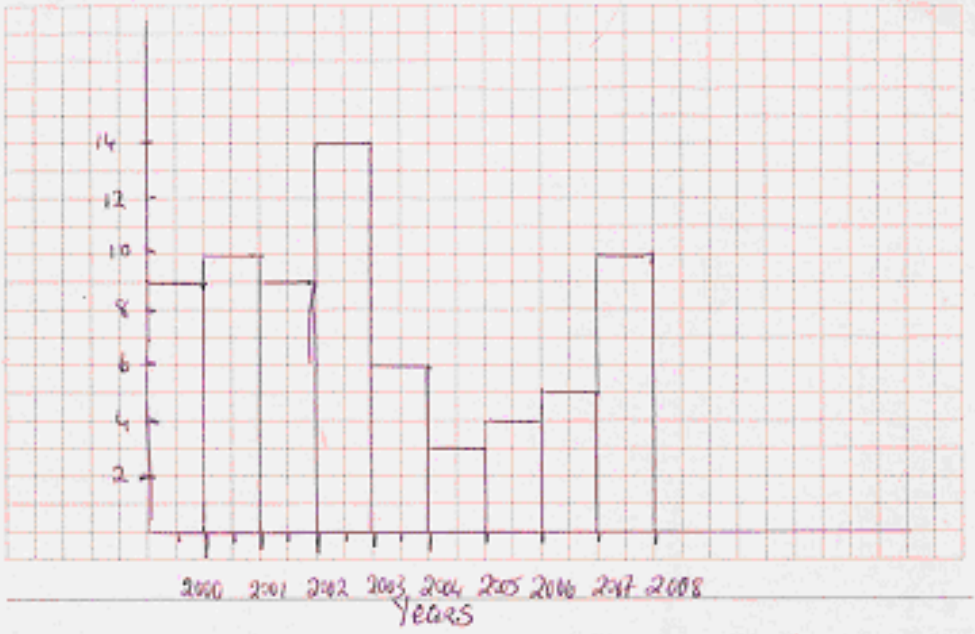
This question highlights the fact that the choice of interval length can **reveal** certain trends or **hide** others.

Look at the plot that is given. How many intervals are there? How many years are in each interval? What can you **conclude** about the number of lost books? Would you say that the measures taken to discourage non-return are working?

Now divide the data into nine intervals and plot the histogram. Is there a difference in the **trend**?

Examine the student work below. Compare this with your work

You may use this page for extra work



This shows a steady increase from 2004 to 2008 which is worrying. There was a good decrease from 2002 to 2004 although it had gone very high from 2001 to 2002

I think this histogram supports the finance officer's view that the measures haven't worked because even though there was a decrease from 2002 to 2004 there is now a steady increase over 4 years and the number of lost books is back to the same as it was in 2001 which was slightly up from when the measures came in

Remember:

Histograms can be cleverly designed to hide or highlight certain trends. Remember this when you are interpreting histograms.

Which type would give you more detail? Which type would give you less detail?

Why might you want to highlight or hide certain trends?

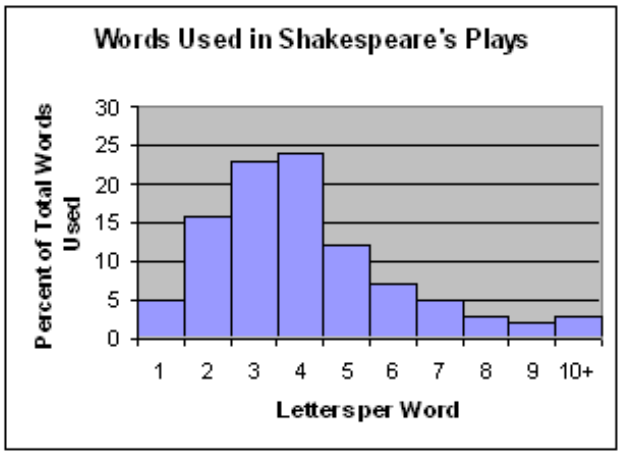
Learning outcome	This partial question gives you the opportunity to provide evidence that you can – recognise that correlation is a value from -1 to +1 and that it measures the extent of the linear relationship between two variables

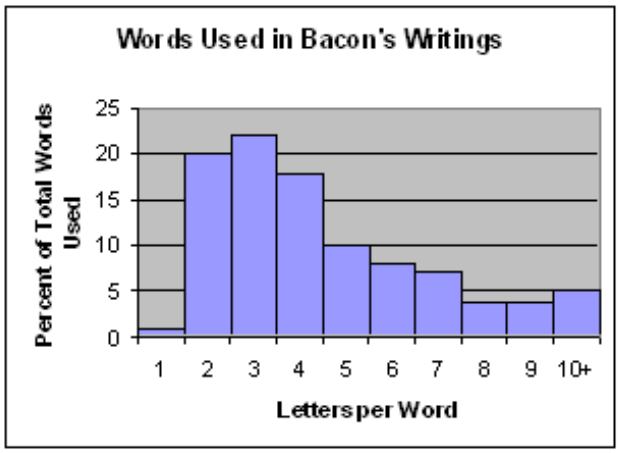
The points in the scatter plot below represent the ages of cars and their values.
Based on this scatter plot, which of the following is a reasonable conclusion?

(i) Age and value have a correlation coefficient that is equal to zero.
(ii) Age and value have a correlation coefficient that is greater than 0.5.
(iii) Age and value have a correlation coefficient that is less than zero.
(iv) Age and value have a correlation coefficient that is between zero and 0.5.

Is the correlation positive or negative? How do you know?
Is there a strong or weak correlation? How do you know?
What numbers represent a strong correlation?
What numbers represent a weak correlation?
Sketch graphs that could represent each of the other answer options.

Q. Some scholars think William Shakespeare was really just a **pen name** for Sir Francis Bacon. (A pen name is a 'fake' name used by another person when writing.) In order to determine if this was true, a researcher counted the letters in every word of Shakespeare's plays and Bacon's writing. The results are recorded in the histograms below.





Based on these histograms, do you think that there is any **evidence** to suggest that William Shakespeare was really just a pen name for Sir Francis Bacon? Explain.

There is a lot of information in these histograms that you could use to support either argument. Yes, William Shakespeare was a pen name for Sir Francis Bacon; or no, William Shakespeare was not a pen name for Sir Francis Bacon. Might there be another explanation?

Are the **distributions** similar? Describe each **distribution.** Use fractions and percentages.

What percentage of Shakespeare's words have 4 letters per word or less?

What % of Bacon's words have 4 letters per word or less?

What percentage of Shakespeare's words have 5 letters per word or less?

What % of Bacon's words have 5 letters per word or less?

Q. The data shows the head circumferences for a group of men and women.

(a) Display the data in a way that will allow you to compare the distributions of head circumferences for both men and women.

(b) Is there any evidence to suggest that men have larger heads than women? Explain your reasoning.

Gender	Head Circumference
F	522
M	580
M	552
F	531
M	563
F	546
F	545
M	545
M	545
M	568
F	560
M	613
F	555
F	573
M	577
F	584
M	600
M	595
M	593
F	590
M	594
F	564
F	536
M	586

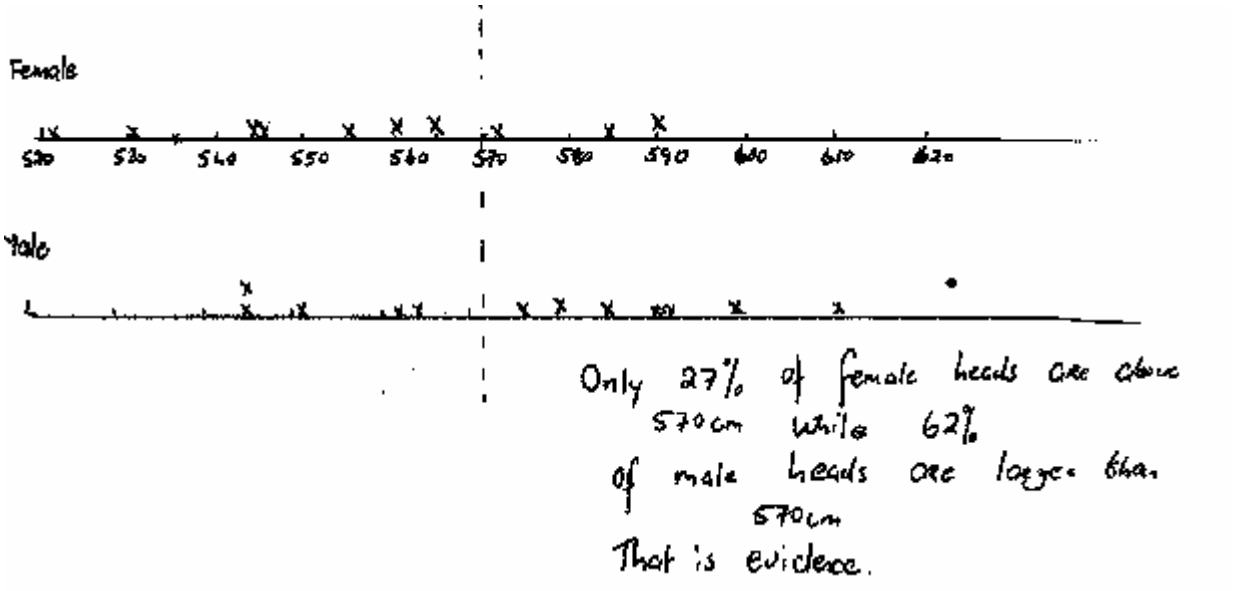
There are a lot of ways to display this data; a line plot, a back to back stem and leaf plot, or a histogram. Eyeball the data and think about how you would display it.

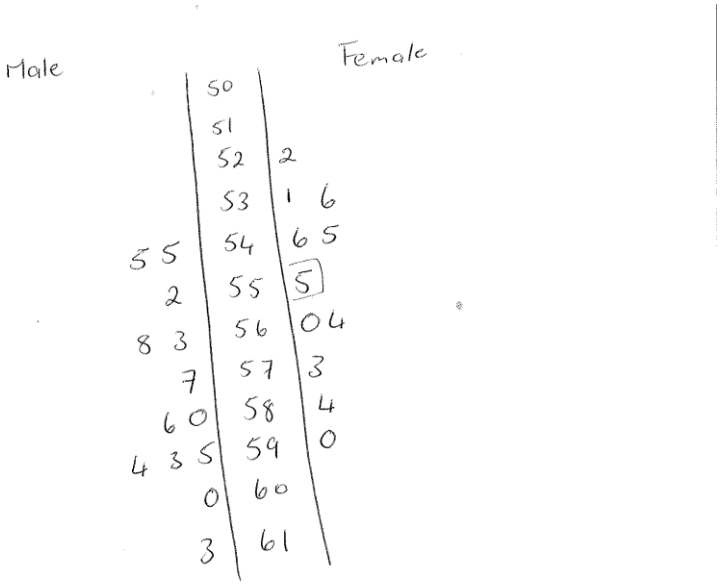
- What features are you looking at in the data?
- How are you deciding which display is most appropriate?

Once displayed you will be able to **comment** on the **distributions** and **draw conclusions** about the relative sizes of the heads of men and women.

Have a go at this and then examine the following examples of how other students displayed the data and drew conclusions from it.

Student work





Range of female is 590 - 522 = 68
Range of male is 613 - 545 = 68

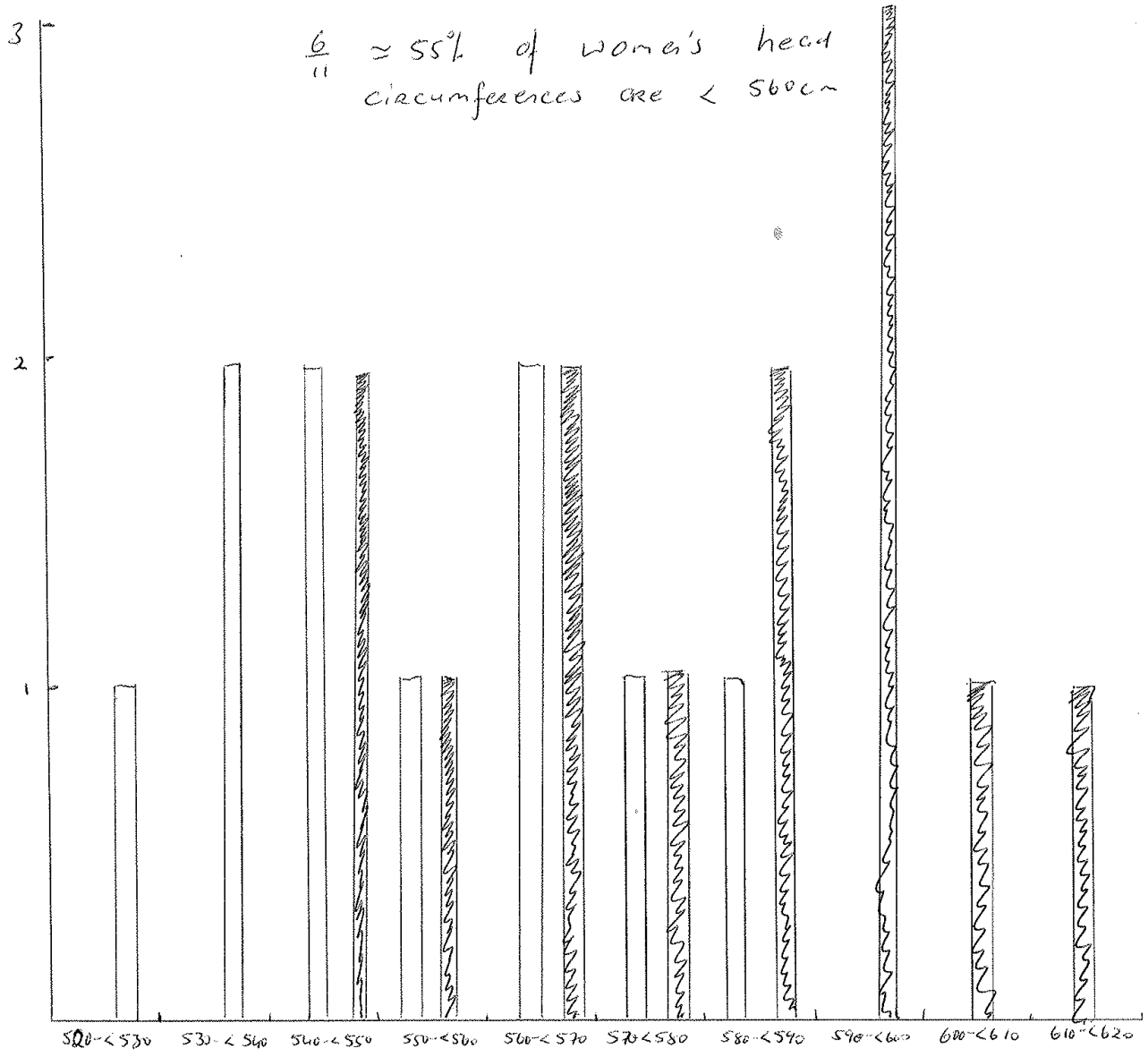
50% of the female head sizes lie between 522cm and 555cm while only 23% of the males lie in that Range

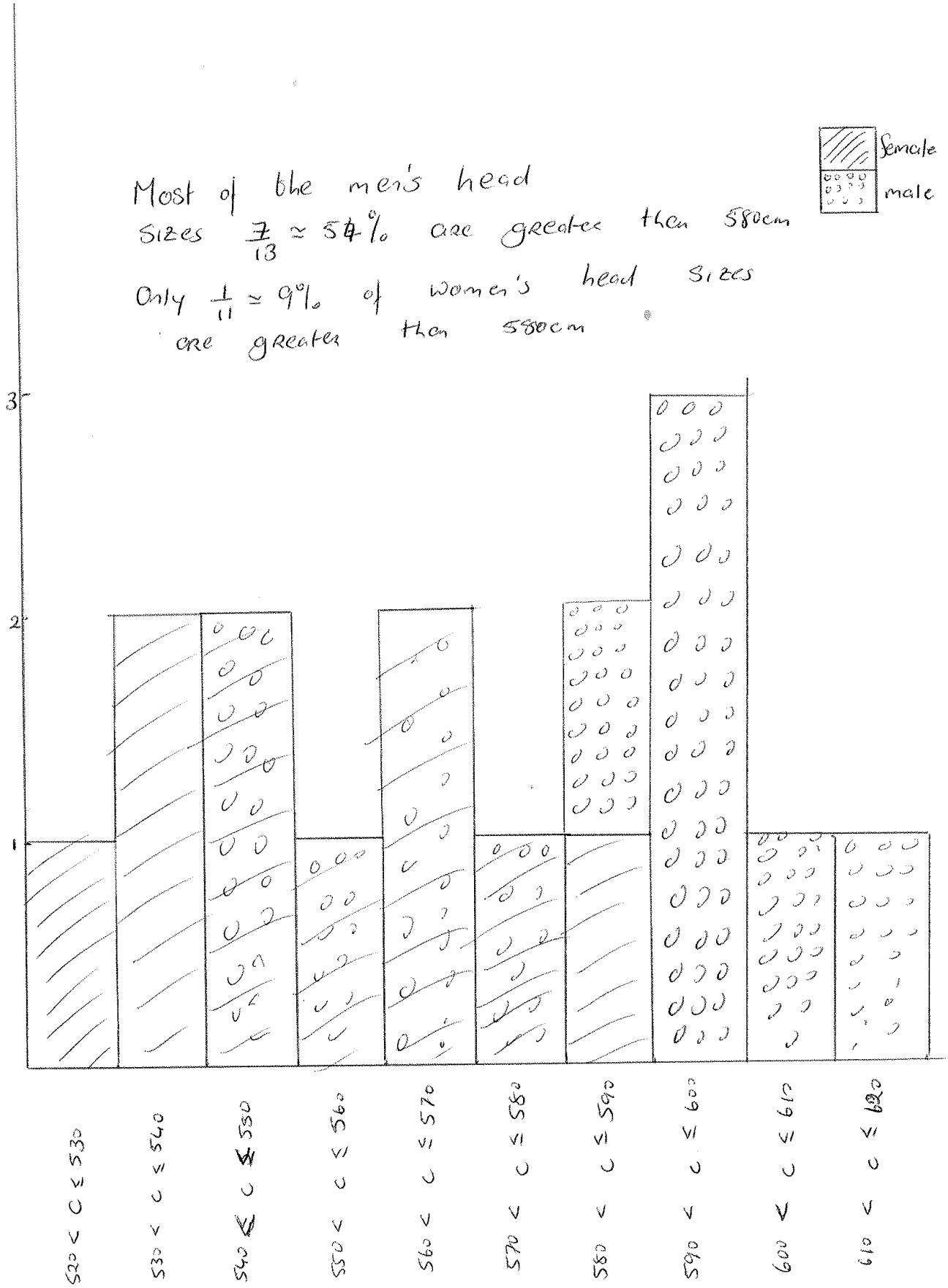
54% of male head sizes are between 580 and 613cm and only 18% of female head sizes are in this Range so yes there is evidence to support statement men have bigger heads than women.

	Male	Female
$520 < C \leq 530$	0	1
$530 \leq C \leq 540$	0	2
$540 < C \leq 550$	2	2
$550 < C \leq 560$	1	1
$560 < C \leq 570$	2	2
$570 \leq C \leq 580$	1	1
$580 < C \leq 590$	2	1
$590 \leq C \leq 600$	3	0
$600 < C \leq 610$	1	0
$610 < C \leq 620$	1	0

$\frac{7}{13} \simeq 54\%$ of men's head circumference is $> 580cm$

$\frac{6}{11} \simeq 55\%$ of women's head circumferences are $< 560cm$




Female
male
Most of the men's head sizes $\frac{7}{13} \approx 54\%$ are greater than 580cm
Only $\frac{1}{11} \approx 9\%$ of women's head sizes are greater than 580cm
3
2
1
520 < C ≤ 530
530 < C ≤ 540
540 < C ≤ 550
550 < C ≤ 560
560 < C ≤ 570
570 < C ≤ 580
580 < C ≤ 590
590 < C ≤ 600
600 < C ≤ 610
610 < C ≤ 620

Q. The 5th year and 6th year students in a local school were asked about the number of hours per week they spent playing on a games console. The results are shown below.

Number of hours spent playing on a games console	**Number of 5th year students**	**Number of 6th year students**
1		
2	1	1
3	2	3
4	1	1
5	1	2
6	5	2
7		3
8		
9	1	3
10		1
11		3
12		2
13	3	3
14	1	1
15	4	
16	4	3
17	2	1
18	4	2
19	4	4
20	3	2
21	2	
22	3	
23	1	
24		
25	1	4

Display the data in a way that allows you to **comment** on the **shape of the distributions**. Is there any **evidence** to suggest that 6th year students spend longer playing a games console than 5th year students?

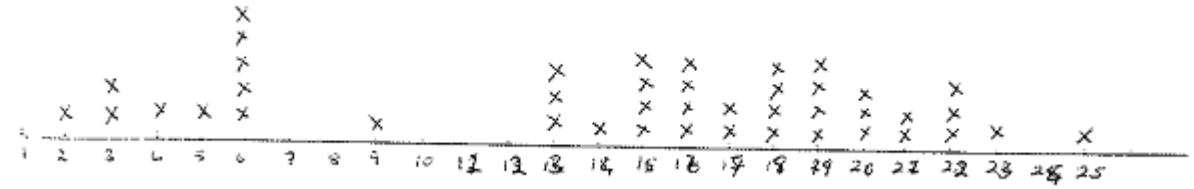
Note to Students

There are many ways you may choose to answer this question.
The data could be displayed in line plots, a back to back stem and leaf plot, or a histogram.
Once displayed you will be able to **comment** on the **distributions** and **draw conclusions** about the relative times spent by 5th and 6th year students on games consoles.

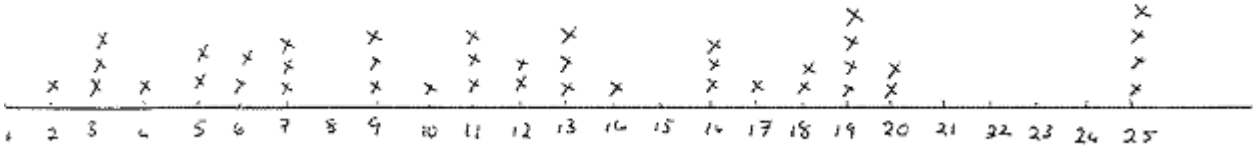
Have a go at this and then examine how student A below displayed the data and drew conclusions from it.
Now try to use a back to back stem and leaf plot and a histogram. Evaluate each display.

Student A

5TH Year



6th year

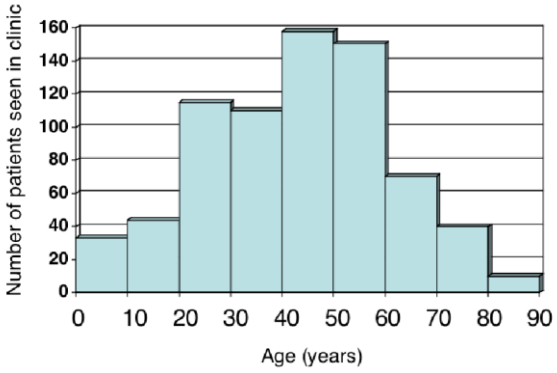


In the 5th year data there are two clusters: between 2 and 6 hours per week and 13-23 hours per week. 10 out of 43 or almost 25% of students play the console over the range of the first cluster. 31 of 43 or 72% are in the second cluster.
Only 1 out of 43 students uses the games console between 7 and 12 hours per week.

The data from 6th years are more evenly spread than the 5th year data there are no real clusters.
16/41 or 39% of students play the games console between 16 and 25 hours per week while ½ of the students play between 2 and 12 hours per week.
The range is the same as the 5th years.

The fact that 24 of 43 or approximately 56% of 5th year students play the console between 16 and 25 hours per week whilst only 39% of the 6th year students play for this length of time indicates that there Is no evidence to suggest that 6th year students spend longer playing a games console than 5th year students. In fact the evidence shows the opposite.

Q. The ages of the patients seen by a group of doctors in a clinic over the last month are shown in the histogram below.

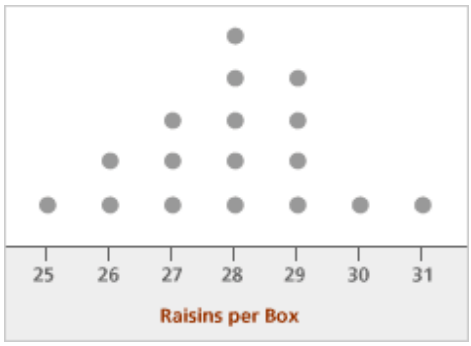


The clinic is about to begin a Swine Flu vaccination programme and must order the drugs they need from the HSE.

If $^1/_3$ of the 40-90 year olds, ½ of the 20-40 year olds, $^1/_5$ of the 10-20 year olds and all the 0-10 year olds who attended the clinic last month are likely to attend for vaccination, what is the minimum number of vaccinations that the clinic should order from the HSE?
Show your workings.

Use the histogram to decide how many of each age group visited the clinic over the last month.
Try to organise your work into a table
How many 40-90 year olds attended the clinic in the last month? What fraction of these is likely to attend for vaccination? How many of these are likely to attend for vaccination?
What about the 20-40 year olds? What fraction of these attended the clinic in the last month? How many of these are likely to attend for vaccination?

Q. Students were investigating the number of raisins contained in individual boxes of Sun-Maid raisins. They recorded their results in the diagram shown.



(a) If the students choose a box at random from all the boxes they surveyed what is the probability that the box contains 29 raisins?

(b) Four boxes were found after the students had completed the line plot above.
Jack, Sarah, Amy and Kevin were each given a box and asked to count the contents.
Jack said his contained 28 raisins. Sarah said hers contained 28 raisins also.
Another student said: “I bet Amy’s contains 28 raisins also.
Kevin said “Wait, Amy; don’t reveal the contents of your box yet.”

He and Amy whispered together and then Kevin said “I will tell you that if the contents of our two boxes are added to the data the mean number of raisins per box will be 28.

Give one possible value each for the number of raisins in Kevin’s and Amy’s boxes.
Justify your choice.

Is it possible that the student won the bet?
Explain your reasoning.

LC-HL only

The next day the students looked up the statistics from the Sun-Maid company.
The mean number of raisins per box is published as 25 and the standard deviation as 5.
If a normal distribution applies, what is the probability that a box chosen at random from the production line contains 29 or more raisins?

Compare this probability with that obtained from the class sample.

- How many boxes contained 29 raisins? How many boxes were there altogether?
- When the 4 extra boxes were found were you surprised that two of them contained 28 raisins? Why or why not? Is it possible that the final two boxes contained 28 each also? Why do you think this? How likely is it that both boxes contain 28 raisins? Is it possible they contain 24 raisins or 25 raisins each? Why, or why not?
- If the mean of the 21 boxes is 28 what does this mean? How many raisins are there altogether? How many are in the 19 boxes? How do you know? How many must be in the other two boxes if the mean stays at 28? How do you know? Is it possible that only one of the other two boxes contains 28 raisins? Justify your thinking.
- If the published data states that the mean number of raisins per box is 25 with a standard deviation of 5, would 29 raisins be unusual? How many standard deviations from the mean is 29? Is it a whole number of standard deviations? Can you use the tables to find the probability of obtaining 29 raisins? Compare this with the probability of getting a box with 29 in the sample described above.