

Galaxy Statistics

Key Stage 4

Topics covered: ICT, galaxies, cosmology, frequency bar charts, median, interquartile range

Teacher's Notes

This activity involves using the Galaxy Zoo website www.galaxyzoo.org to classify five samples of ten galaxies. The results are recorded in a table and presented in the form of a bar chart for each sample. These can be drawn on paper or plotted using Excel. Students then look at the differences between their graphs and think about what could affect their results such as sample size and personal bias. They also look at the distribution of different types of galaxies in the Universe.

Equipment: computer, MS Excel or graph paper, rulers, pencils

Questions to ask the class before the activity:

What is a galaxy?

Answer: a large collection of stars hundreds of thousands of light years across.

What do you think determines the shape and size of a galaxy?

Answer: the amount of matter in the galaxy (stars), the age of the galaxy and whether the galaxy has interacted with another galaxy.

Questions to ask the class after the activity:

Why do you think galaxies have different colours?

Answer: this depends on the type of stars they contain, old dying stars are cooler and appear red, young hot stars more massive than the Sun are blue/white in colour.

Did you find a pattern in the types of galaxies you discovered and their colour?

Answer: 'Smooth' galaxies (ellipticals) are yellow/red in colour, spirals are blue/yellow and clumpy galaxies are a mixture of both – these could be irregular galaxies that don't fit into either type or are just too distant to classify properly.

What was difficult about the project?

Answer: this is likely to be trying to classify faint galaxies with no apparent structure.

What do you think the 'clumpy' galaxies might be?

Answer: some of them might be spirals that are too distant to resolve, or they could be early galaxies that haven't developed yet.

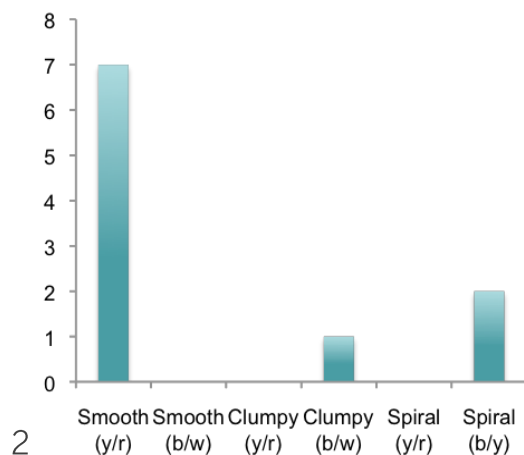
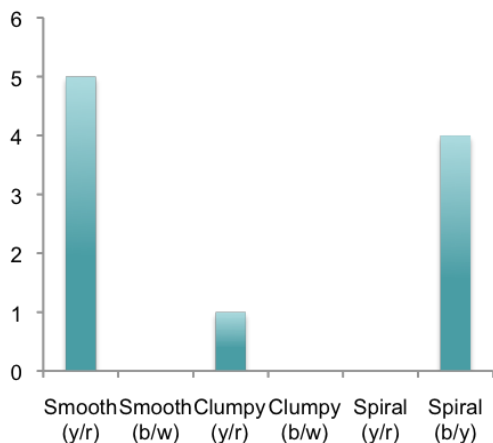
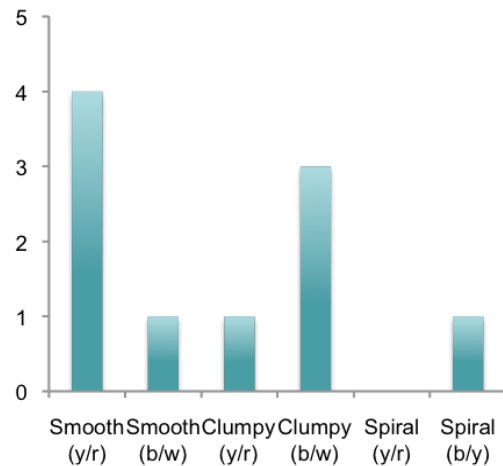
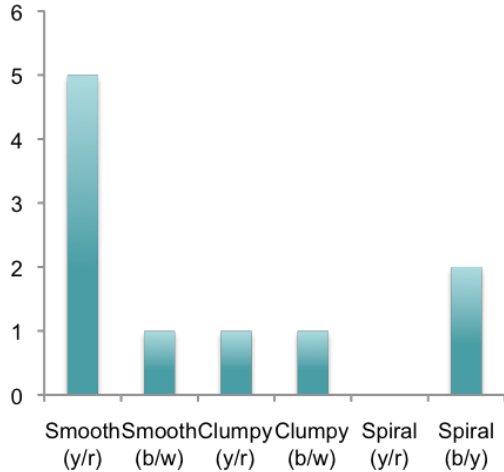
How do you think human error and personal bias is removed from the classification on the Galaxy Zoo website?

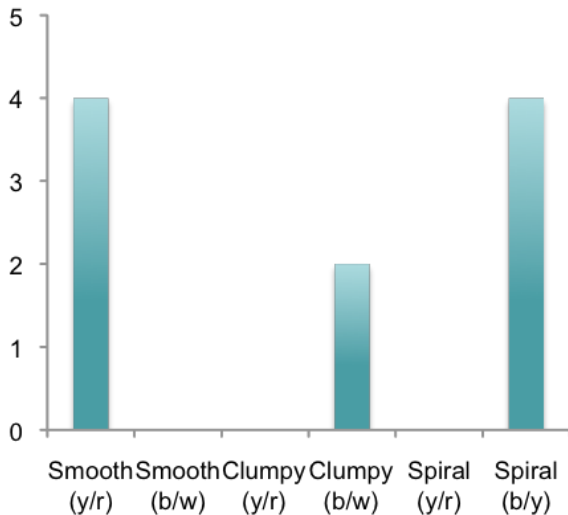
Answer: by giving the same galaxy to a number of people and finding the most common result.

Using Galaxy Zoo: Example Answers

(students will get different answers depending on their sample data)

Smooth (yellow/red)	Smooth (blue/white)	Clumpy (yellow/red)	Clumpy (blue/white)	Spiral (yellow/red)	Spiral (blue/yellow)
				0	
				0	
	0		0	0	
	0	0		0	
	0	0		0	
Total: 25	2	3	7	0	13





Plot a bar chart for each sample and compare the results.

Are there differences? Why do you think this is? What could you do to improve the accuracy of your results?

Answer: The sample sizes are small, only 10 galaxies per sample. Only 5 samples were taken giving a total of 50 galaxies. There will be a lot of intrinsic variation in small numbers. Comparing much larger samples would give a better view of galaxy distribution in the Universe.

What do your results tell you about the distribution of galaxies in the Universe?

Answer: Smooth galaxies are mostly red and spiral galaxies are mostly blue and yellow. There appear to be more smooth type yellow/red galaxies (elliptical galaxies) than the other types, however more classifications may reveal whether there are an equal number of spiral (blue/yellow) galaxies.

Higher Level

(students will get different answers depending on their sample data)

Taking samples allows you to find an average value, the mean, or the middle value, the **median**.

1. Find the range in the number of blue/yellow spirals over the 5 samples (maximum number – minimum number).

Answer: max = 4, min = 1, range = 3

2. Find the middle or median value using $(n+1)/2$ (n = number of values) for the numbers of spirals.

Answer: Values in order 1, 2, 2, 4, 4. The median is the 3rd value in the line $((5+1)/2 = 3)$, median = 2

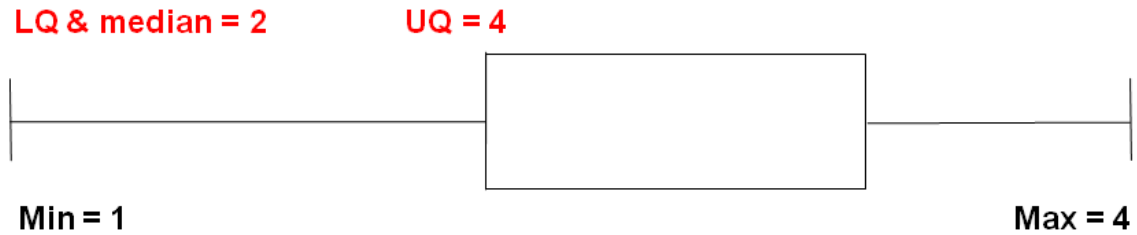
3. Find the lower quartile, LQ lies at $(n+1)/4$.

Answer: Values in order 1, 2, 2, 4, 4. LQ is the 2nd value in the line $((5+1)/4 = 1.5)$, therefore LQ = 2

4. Find the upper quartile, UQ lies at $3(n+1)/4$.

Answer: Values in order 1, 2, 2, 4, 4. UQ is the 4th value in the line $(3(5+1)/4 = 4.5)$, therefore UQ = 4

5. Draw a box plot for your spiral data on graph paper.



6. Write down the interquartile range (UQ - LQ). Answer: $4 - 2 = 2$

What do you think the interquartile range indicates?

Answer: the spread in the data.

What could you do to change the interquartile range?

Answer: increase the sample size and the number of samples. This might reduce the interquartile range.

What do your results tell you about the distribution of galaxies in the Universe?

Answer: Smooth galaxies are mostly red and spiral galaxies are mostly blue and yellow. There appear to be more smooth type yellow/red galaxies (elliptical galaxies) than the other types, however more classifications may reveal whether there are an equal number of spiral (blue/yellow) galaxies.

Activity: Galaxy Statistics (the Galaxy Zoo Project)

The Galaxy Zoo project is an opportunity for the public to contribute to scientific research. The project needs you to help classify millions of galaxies. Knowing the morphology of a galaxy tells us something about its mass, its star formation history and its dynamical and merger history and gives us an insight into the structure and history of the Universe.

The site is very easy to use, go to www.galaxyzoo.org, once you are registered you can begin classifying galaxies straight away. The images are taken by the Hubble Space Telescope and many are very distant galaxies that formed in the early Universe. Some of these are difficult to classify so it is recommended that you look at the tutorial page:

www.galaxyzoo.org/how_to_take_part

There are 6 categories of galaxy that you will be looking at, these are listed below:

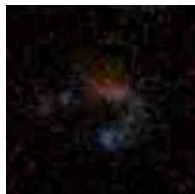
Smooth (yellow/red)



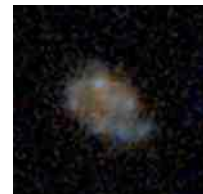
Smooth (blue/white)



Clumpy (yellow/red)



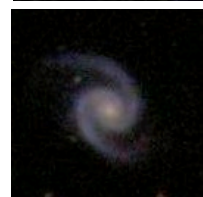
Clumpy (blue/white)



Spiral (yellow/red)



Spiral (blue/yellow)



Use the table on the next page to record your classifications of a total of 50 galaxies. Do these in 5 samples of 10. Make a tally of each type in each sample (the total should equal 10 for each row). If your galaxy is neither smooth, clumpy or spiral continue classifying and record the next one that fits one of these 3 categories.

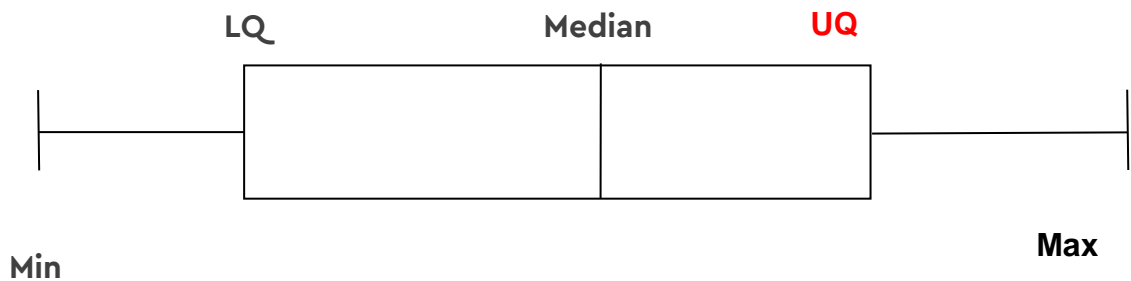
Plot a bar chart for each sample and compare the results. Are there differences? Why do you think this is? What could you do to improve the accuracy of your results?

Smooth (yellow/red)	Smooth (blue/white)	Clumpy (yellow/red)	Clumpy (blue/white)	Spiral (yellow/red)	Spiral (blue/yellow)
Total:					

Higher Level

Taking samples allows you to find an average value, the **mean**, or the middle value, the **median**. To find the median all values should be placed in order from smallest to largest. The middle value can be found by using $(n+1)/2$, where n = number of values. If $n = 3$, the median will be the 2nd value in the line.

1. Find the range in the number of blue/yellow spirals over the 5 samples (maximum number – minimum number)
2. Find the middle or median value using $(n+1)/2$
3. Find the lower quartile, LQ lies at $(n+1)/4$
4. Find the upper quartile, UQ lies at $(3n+1)/4$
5. Draw a box plot (below) for your spiral data on graph paper.



6. Write down the interquartile range (UQ – LQ).

What do you think the interquartile range indicates?

What could you do to change the interquartile range?

What do your results tell you about the distribution of galaxies in the Universe?