

# Displaying and describing the distributions of categorical variables

Friday, 1 February 2019 8:03 PM

By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Know the different way we can display and describe the distributions of categorical variables including:
  - The frequency table
  - The bar chart
    - Stacked or segmented bar charts
- Know what the mode is and what it is used for
- Know how to answer statistical questions involving categorical variables

## RECAP:

---

In the last lesson we looked at the differences between variables. We found that there were two main categories, which can then be subdivided into subcategories:

- Categorical Variables
  - Nominal variables
  - Ordinal variables
- Numerical Variables
  - Discrete Numerical
  - Continuous Numerical

We are going to take some time in looking at how to display and describe the distributions of categorical variables

## Why do we need to display data in different forms

---

Imagine I held a survey and asked 75 people lots of questions. One of them might have been, "What gender are you?" This might be the data we received:

M	F	F	F	F	M	M	M	F	M	F	M	F	F	M
M	F	F	F	M	F	M	M	M	F	M	F	F	M	F
M	F	F	F	M	M	F	M	F	F	M	M	M	M	F
F	F	M	M	F	F	F	F	F	M	M	M	F	M	F
M	F	M	M	M	F	F	F	F	F	M	M	F	M	F

Probably not the most useful way to display this. How does the above answer the following questions?

- How many males were asked?
- How many females were asked?
- What percentage of the survey was male?
- What percentage of the survey was female?

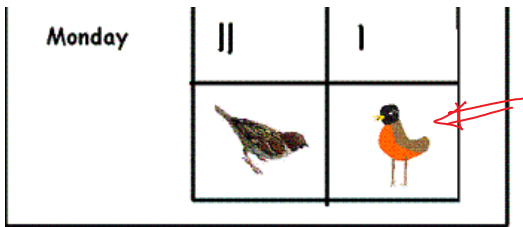
We can take the data above and place it into different forms. One of these is a frequency table. We want the data to be shown in a **compact form** and to show how the data is **distributed**.

## We love tallies!

---

Do you remember back in primary school when we had to keep drawing those wonderful tally charts?!

Bird Count		
	Sparrows	Robins
Friday		
Thursday		
Wednesday		
Tuesday		
Monday		



Sadly, we don't have the fun of drawing pretty pictures of birds, or the little marks.  
We can now count!

We use a table to help us display the data in a more useful way:

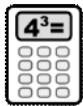
M	F	F	F	F	M	M	M	F	M	F	M	F	F	M
M	F	F	F	M	F	M	M	M	F	M	F	F	M	F
M	F	F	F	M	M	F	M	F	F	M	M	M	M	F
F	F	M	M	F	F	F	F	F	M	M	M	F	M	F
M	F	M	M	M	F	F	F	F	F	M	M	F	M	F

Sex	Frequency	
	Number	Percentage
Female	40	53.33
Male	35	46.67
Total	75	100%

NOTE: The order of the male/female was arbitrary.  
This is nominal data and hence no order can be inferred.

**How to calculate a percentage?**

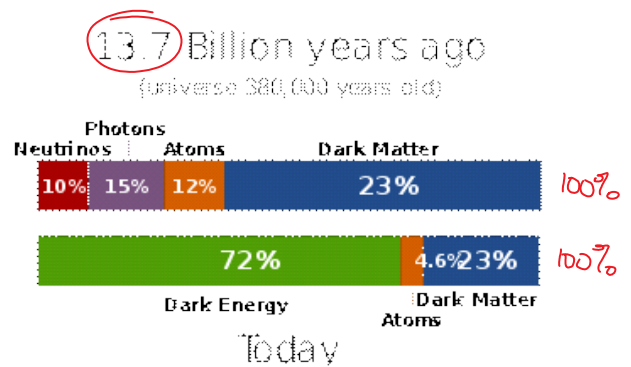
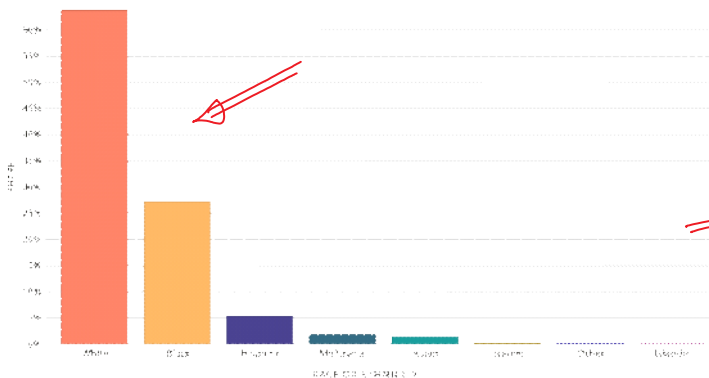
To find the percentage of anything you need to do the following:

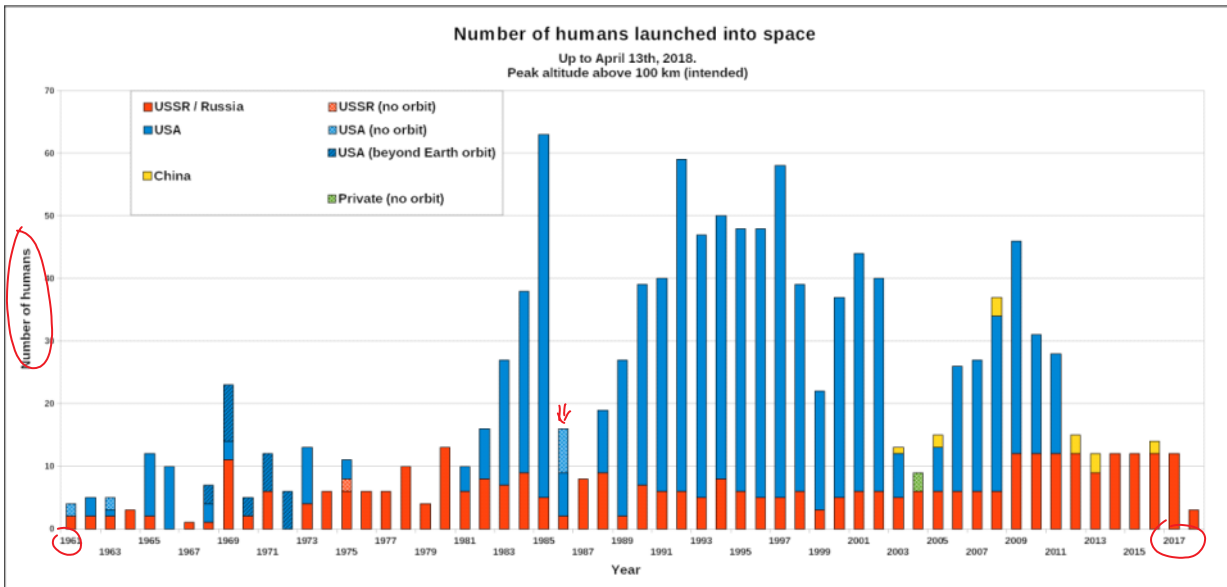


$$\text{Percentage} = \frac{\text{number of successes}}{\text{total number of trials}} \times 100$$

**The Bar Chart**

Who doesn't love a good old bar chart!  
The interesting thing, there are a number of different Bar Charts.  
**[Good old Barry - at it again]**  
Here are some examples:





Bar charts can give us a great way to show data.  
 We tend to use a bar chart to display data after it has been placed into a **Frequency Table**.

### Important things to note about a Bar Chart

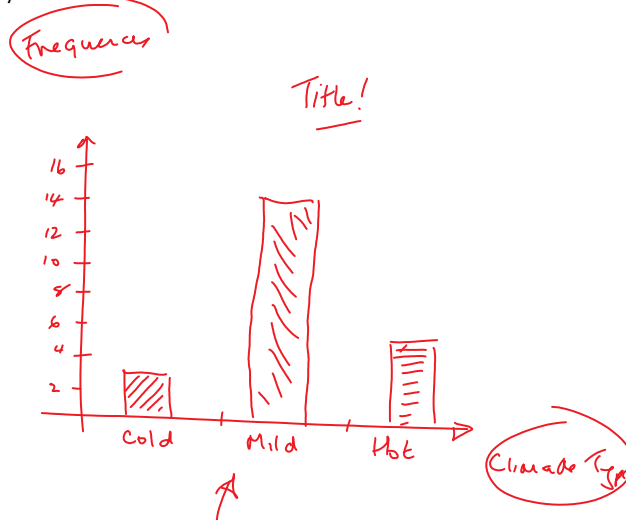
- There are ALWAYS gaps between the bars
- Frequency (or percentage frequency) always goes on the vertical axis
- The variable you're displaying goes on the horizontal axis.
- There is one bar per category!

### Example of drawing a bar chart:

*This question is taken from the Cambridge Essentials Textbook series*

The climate type of 23 countries is classified as 'cold', 'mild' or 'hot'. The results are summarised in the table opposite. Construct a frequency bar chart to display this information.

Climate type	Frequency	
	Number	Percentage
Cold	3	13.0
Mild	14	60.9
Hot	6	26.1
Total	23	100.0

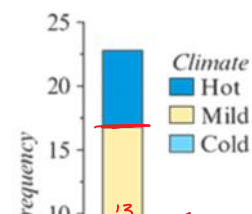


### Stacked or Segmented bar charts

An excellent example of a stacked bar chart is shown below.  
 Notice how the bars have sections.  
 Each of these sections stands for a different category.

In the example shown, we can see that the categories are Hot, Mild and Cold.  
 These are examples of Ordinal Variable - they have an implied order.

The frequency is shown on the vertical axis.



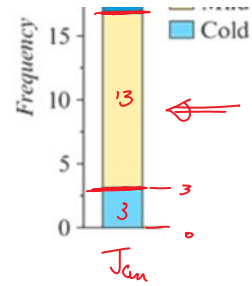
These are examples of Ordinal Variable - they have an implied order.

The frequency is shown on the vertical axis.

This might be showing the climate on a particular day.  
It might be showing the number of hours in that day when it was considered either Hot, mild or cold.

We could extend the graph and show the results from other days.

In the example, as this is only one day (for example) we do not need to place a label on the horizontal axis.



### Finding the MODE

I love the mode as it's the easiest thing to find out of all the measures we can take for data.  
What do you notice about their first few letters of the word Mode?  
Remember, Barry is here to try and confuse you with each new term.  
Look for the tricks and help with language.



To find the mode we simply see which category has the highest frequency.



### What is a statistical question?

This is a question which depends on data for its answer.  
Further Mathematics requires you to be able to analyse information (data) and then write a concise report which highlights the key findings from the data.

The great thing is, that in many cases, you can scaffold your answer in the same way and just "change the numbers".

Here is an example, extracted from the Cambridge Essentials Textbook:

In an investigation of the variation of climate type across countries, the climate types of 23 countries were classified as 'cold', 'mild' or 'hot'. The data are displayed in a frequency table to show the percentages.

Use the information in the frequency table to write a concise report on the distribution of climate types across these 23 countries.

Climate type	Frequency	
	Number	%
Cold	3	13.0
Mild	14	60.9
Hot	6	26.1
Total	23	100.0

Here is a suggested response:

The climate types of 23 countries were classified as being, 'cold', 'mild' or 'hot'. The majority of the countries, 60.9%, were found to have a mild climate. Of the remaining countries, 26.1% were found to have a hot climate, while 13.0% were found to have a cold climate. 4

To be considered as a good response we need to ensure the following:

- If there is a clear modal category that you mention it
- Include frequencies or percentages in the report. **Percentages are preferred.**

To be considered as a good response we need to ensure the following:

- If there is a clear modal category that you mention it
- Include frequencies or percentages in the report. **Percentages are preferred,**
- If there are lots of categories, do not mention them all, but make sure the modal one is **mentioned.**
- Ensure you mention reference to what the question was asking you. The context. Include the total number of items which were included in the survey.