Royal Geographical Society with IBG

Advancing geography and geographical learning

A Student Guide to the A Level Independent Investigation (Non-examined Assessment—NEA)

View or download at www.rgs.org/nea

Follow us on Twitter
@RGS_IBGschools
www.rgs.org/schools

Funded by the Nuffield Foundation

Nuffield Foundation

A Level Independent Investigation Guide

Royal Geographical Society with IBG

Advancing geography and geographical learning

The new geography A Level specifications raise exciting challenges for students, one of which is the introduction of the Independent Investigation. Its place in the A Level course prepares you more thoroughly for research-based learning at university level, as well as providing opportunities to develop key investigative skills for the workplace.

The A Level Investigation will allow you to develop your interests in a chosen aspect, or aspects, of geography independently. It will develop your research skills as well as help you to develop a stronger knowledge and understanding of geography, which previously you may have only experienced in the classroom. For those thinking about studying geography at university, your Independent Investigation will provide you with an important focus for your UCAS application.

For many students, this will be their first encounter with the enquiry process and this guidance is designed to make your journey through the investigation as clear as possible. From the initial decisions to be made over the focus of your research, to analysing and evaluating investigative techniques, the guidance will cover each stage of the research framework. You will find suggestions and points for consideration at every stage of the research, allowing you to remain in control of your Independent Investigation at all times, while simultaneously stretching and challenging you to get the most from this key part of your A Level course.

Regardless of your ability level or the specification your school or college follows, these guidelines are designed to give you the confidence to engage in what might be your first steps into geographical research.

Nuffield Foundation

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Contents

Before you start

Undertaking geographical research may initially sound daunting, especially if you have not covered this style of working in your GCSE studies. However, being able to carry out research at A Level is not difficult if you plan your study effectively and make good use of the support around you – such as the advice provided by your teacher, and by reading this guide.

New geographical research is becoming increasingly important in a fast-paced and changing world; through this research, geographers try to find answers, create models and design management plans. Whilst you may not yet be at a stage where your research is published, the Independent Investigation provides you with important grounding in the broad research process.

Universities and employers are also interested in research you may have undertaken at A Level. Being able to discuss a personal piece of independent research, whether for academic or employment purposes, can enhance your potential as a geographer. Whether you plan to continue studying geography beyond A Level or not, the skills you will learn whilst completing your Independent Investigation will benefit you in many other subjects in higher education, and indeed, in your life in the workplace.

You will need to fully consider health and safety arrangements according to the policies in place at your school. This document does not offer specific guidance in that respect, but it is an essential element of your planning. Discuss this further with your teacher.

You will also need to consider any cultural, heritage, ethical and political issues associated with the work you plan to undertake. Discuss this further with your teacher.

Choosing a Research Topic: The key questions to consider

Which broad area of the specification have I enjoyed and found interesting?

What in particular do I like about this topic and why?

What questions do I want to find an answer to - and are these questions geographical?

What do geographers already know about this topic? And how can I find out what they think?

How am I going to make my research unique and different?

How much time do I have to complete the research?

In which location might I be able to study this topic?

Where am I going for my school fieldwork and what locations do I have available to me?

What data could I collect or use?

The starting point for your Independent Investigation is to think about which area of geography you would most like to study. It may be a good idea to look at the areas of your exam board specification which have most interested you so far, or indeed ones looking ahead that you think sound most interesting. You may want to read around your particular subject area so you have a good idea of what researchers have already studied and how they went about it. You may also want to make a note of any models or theories which will inform your ideas further as you start the research journey.

Think about your locality and what questions you might have about it in relation to your subject area. There may be a new geographical feature, such as a flood management scheme or the building of a new public amenity that is drawing lots of local attention. These kinds of plans, and their associated local impacts, may not reach the academic heights of university research, but they can make excellent bases for comprehensive studies at A Level. It is worth remembering that it is relatively easy to broaden a small topic area once you have it in mind, and much harder to scale a large project down.

Most research begins with a broad aim. From this, researchers often develop more unambiguous research questions. These questions have to be quite specific in order for precise data to be collected and meaningful conclusions to be drawn. Research questions may be replaced by or used in addition to hypotheses. A hypothesis is a statement that the researcher aims to either prove or disprove by carrying out the research. It is important to make sure it is actually possible to answer your research questions or prove/disprove your hypotheses: very broad research questions or meaningless hypotheses can make this very difficult.

Your teacher will be able to give you clear deadlines for the completion of your study as well as information about how much of their class time or fieldwork may be given over to supporting you to undertake your investigation. Your exam board may also provide a range of support for this aspect of your A Level. It is also important to remember that your research area may have to hold your attention for quite a large section of your A Level year: it is wise to make sure you are not bored of the subject area before you begin.

Your teacher will be able to give you advice about the nature of your school or college's fieldwork and whether it may be able to accommodate you collecting data on your chosen topic(s). You may decide to modify your area of research slightly to fit with the nature of your fieldwork or you may decide to collect data independently and in your own time, possibly working with others – both are valid ways of conducting your data collection for your Independent Investigation. In this regard, 'independent' does not necessarily mean isolated; there will be issues of safety in the field to consider.

Once you have considered all these questions, and think you have the answers, you are ready to complete a planning form. It does not have to be exact but it should provide a fairly clear idea of the focus of your research. Your teacher may then be able to provide further guidance, such as increasing the scope of the study to make it more challenging, or scaling it back to make it more manageable. Your teacher will know your personal capability well and is best placed to offer support, within the limits set by your exam board.

Example One: Thinking through your research before you start

Which area of the specification have I enjoyed and found interesting?

Population and cultural diversity

What do I particularly like about this topic?

It's interesting that international migrants are now increasingly found in rural areas rather than cities.

Having done some background reading about this topic, what kind of research questions might be possible?

I would like to know why EU migrants are moving to rural areas – what the push and pull factors are for this movement. I am also thinking about what their impact has been on the rural area and what challenges they might face. I guess I should start by looking at what kind of number of EU migrants are actually moving to rural areas.

Where can I carry out my data collection into this topic? Where is my school going for their A Level fieldwork and will that help?

My class is going to Swansea for a week and there may be the chance to interview agricultural workers in the Valleys. However, here where I live in Lincolnshire there are lots of farmers who employ workers from the EU so it may make more sense to collect data in my own time and locality.

What kind of data might I need to collect and how might I use it?

Interviews and questionnaires might be a good idea for primary data collection and I can look at census data for Lincolnshire to see how EU migration has changed over time. If I get enough questionnaire responses I can look for trends within the data.

My first, rough thoughts for my research questions are...

1. How has EU migration to Lincolnshire changed in the last fifty years?
2. What have been the primary reasons for this movement to Lincolnshire?
3. What opportunities (social, economic and environmental) has this migration created?
4. What challenges (social, economic and environmental) has this migration created?

Example Two: Thinking through your research before you start

Which area of the specification have I enjoyed and found interesting?

The water and carbon cycles

What do I particularly like about this topic?

How the water and carbon cycles can have an impact on human well being and how we live our lives.

Having done some background reading about this topic, what kind of research questions might be possible?

I enjoyed reading about environmental Kuznets Curves and how socio-economic status might have an impact on management of air pollution in particular. So I would want to see if the Kuznets Curve model is true for water courses too and research whether the prosperity of different places is connected to the pollution level in their rivers , lakes and streams etc.

Where can I carry out my data collection into this topic? Where is my school going for their A Level fieldwork and will that help?

My class is going to the Lake District for a week and on the way, as well as while we are there, we will pass through lots of towns and villages with rivers running through them. I will be able to collect samples from these and test them later. I can also collect samples from places I visit in my summer holidays.

What kind of data might I need to collect and how might I use it?

The water samples can be tested using simple litmus or phosphorous strips which indicate numerically the level of pollution in the river. I can then use secondary data such as unemployment levels as an indicator for the strength of the local economies - I can then compare the results with the classic Kuznets Curve.

My first, rough thoughts for my hypotheses are...

1. The lowest levels of pollution will be found in areas that have both the lowest and highest levels of socio-economic status.
2. The highest levels of pollution will be found in areas that have the median levels of socio-economic status.
3. Anomalies may come about due to intensive agriculture in an area or an unusual history of higher than average unemployment in the area.

It is important to remember that even though your teacher, class and school may undertake fieldwork investigations together, the exact focus of your research and write-up of your study needs to be individual and independent. That means that your study cannot be copied from work that has already been published. This ‘copying’ is known as plagiarism and if found in your Independent Investigation will be dealt with very strictly by your exam board. To avoid plagiarism, it is important to reference the sources of information from which you get your ideas.

i – Student Planning Form

This form can be used by students to gather ideas before completion of a formal proposal form.

Student Name	
Proposed Independent Investigation title	
Link to exam board specification	
Broad aim of the research	
Research questions / Hypotheses	1.
	2.
	3.

<table>
<tr><td colspan="2"></td><td colspan="2">4.</td></tr>
<tr><td colspan="2">Proposed location(s) for data collection</td><td colspan="2"></td></tr>
<tr><td colspan="2">Safety or ethical issues and concerns</td><td colspan="2"></td></tr>
<tr><td rowspan="3">What data will be collected?</td><td></td><td>Qualitative</td><td>Quantitative</td></tr>
<tr><td>Primary</td><td></td><td></td></tr>
<tr><td>Secondary</td><td></td><td></td></tr>
</table>

			Who?	Where?	What equipment?	How?	When? / How often?	Sampling technique?
	Overview of qualitative data collection methods to be used	Primary Data						
		Secondary Data						
	Overview of quantitative data collection methods to be used	Primary Data						
		Secondary Data						

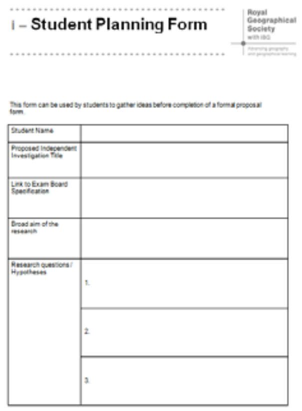
					Why?
What information might I be missing?					
What are my 'known unknowns'?					
Are there any 'unknown unknowns'?					
Possible statistical analysis methods to be used and why					
Possible presentation techniques for data	1				
	2				
	3				
	4				
	5				
	6				
Data Presentation: Numerical Spatial Cartographic Graphical					
Use of GIS (if appropriate) in data presentation					

ii – A Guide to
Writing a Research Plan

Royal Geographical Society with IBG

Advancing geography and geographical learning

Entering into a long-term research investigation is not something that can be done well without proper planning. A planning sheet can be a good way of focussing on the different stages of the research process that lie ahead. With sections in the chronological order in which the Independent Investigation can be written, the planning sheet can act as both a rough guide and a diary for your personalised research procedures. Showing a research planning sheet to a teacher can also help them to see what you intend to do and how you intend to do it. In some cases, they may be able to offer you guidance before you go too far down the wrong path in your thinking. Show your teacher the document again if you make changes.



i – Student Planning Form

This form can be used by students to gather ideas before completion of a formal proposal form.

Student Name	
Proposed Independent Investigation Title	
Link to Exam Board Specification	
Broad aim of the research	
Research questions / Hypotheses	1.
	2.
	3.

Spending time completing a planning sheet can really save you time and effort in the long term and it is worth considering each of the questions carefully. However, the sheet is also just a rough guide: you should not aim to write in too much detail at this stage, as specific descriptions and explanations should be saved for the actual write-up of the Independent Investigation. Instead the planning sheet can be a set of notes and ideas – some might not even be firmly thought through at this stage, but instead be seen as a starting point for future thought.

The completed planning sheet will help you to identify any gaps in your knowledge of research methods and identify areas for future background reading. As you understand how to find certain information or carry out certain techniques you can add to your planning sheet.

The planning sheet can also be viewed as a working document: it is highly likely that as you plan your research in more detail your ideas will change and you will refine your strategies for tackling different parts of the process. Consider editing your ideas as you think through your research. It is important to remember that a good researcher is one who adapts their research to new ideas and information. Most researchers at university level will change their research focus many times before they actually go into the field to collect data – and in some cases after this point too.

In this sense the planning sheet also acts as a diary and you can use it to record observations and make notes on what goes well about your research and indeed what you would change if you were to carry out your Independent Investigation again: something on which you should aim to comment in the final stages of your write-up. You may also be able to use your planning sheet as a guide for when you write-up the earlier stages of your Independent Investigation. To have a working document like this makes this process a lot easier than if you were to try to undertake it from cold memory, sometimes months after you have carried out your data collection, for example.

iii – A Guide to Effective Background Reading

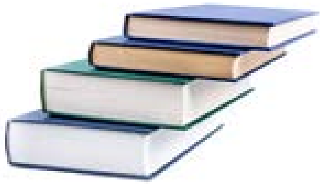
Royal Geographical Society with IBG
Advancing geography and geographical learning

Researchers use background reading from the very start of their investigative process. It can inform their ideas, lead their thinking in a certain direction and tell them where a gap in the body of knowledge lies; a gap that hopefully the researcher will be able to fill through their own enquiry.

For your Independent Investigation, background reading is also important. As well as gaining you access to the higher grade levels, it will also make the write-up and planning of your investigation a lot easier.

What to read

Generally your reading will be focussed around four sources of information. Good research looks to use a variety of these sources so try not to get too caught up with, for example, only using the internet as a supply of information. It is worth noting that not everything you read will automatically find its way into your final Independent Investigation write-up – you are likely to read about a third more than you actually quote from. However, none of your reading should be viewed as a waste of time: even if you do not end up using a particular idea in your write-up, it will have informed your thinking or steered you away from the wrong direction.



Books are a good starting point for background reading. They will often give a broad overview of a topic and give you an idea of which areas of your investigation to look into next. Avoid using just your school textbook at any stage of your background reading – it not only shows a lack of real research endeavour, but as a text it is unlikely to be designed for the depth needed for your Independent Investigation. Instead, it is a good idea to ask your teacher for good books on a topic, or your school librarian. Most local libraries can also order books for you if there is something specific you are looking for.

Journals are academic magazines that contain a collection of recently written research from researchers working in universities. They will contain highly specific details about a certain topic, often quoting real life examples from around the world. Reading and understanding the whole of a journal article may seem a little daunting at first so it is a good idea to start by reading the abstract and seeing if it covers the particular idea you are looking for.
Some geographical journals are available online for open download, while others are hosted through special servers. Your school librarian may also be able to provide advice about any subscriptions to journals that the school holds.

The ease with which you can use an internet search engine to find information makes **websites** common sources of information. However, this ease comes with a catch and the internet can be notorious for being a source of wrong and misleading 'facts'. It is important to consider authorship: while books and journal articles have to be reviewed many times before they can be published, the internet is an open forum on which anyone can write their opinion. If used with caution, and by using websites from well-known organisations, the internet can open up a wealth of information which you may otherwise not have access to.

Newspapers can provide easy access to relevant and up-to-date information on a range of contemporary geographical topics. However, like the internet, they should be used with caution. Most newspapers subscribe to a particular political viewpoint and so may only present one side of a particular argument. Newspaper editors may over-sensationalise an individual story as they know it may sell more copies, ignoring or underplaying some of the more contextual, but less entertaining, facts about the piece. Local newspapers may be a good starting point for finding out about issues that are affecting your local area.

How to read

Before you start reading anything make sure you are ready to record the information you find. Then note the following information:

- Who wrote the article
- The year of writing
- The name of the book / journal / newspaper / website etc.
- The name of the article under which the information appears
- Who published the article and the place of publication
- The page numbers where the article appears
- Other specific information such as a web address

You may wish to visually scan the article first for key words and ideas and then go through in more detail noting the main points. Aim not to write down everything the author is saying, but instead only what is relevent or of interest to your study – in the case of a journal article, this may only be two or three sentences summarising the main points. Remember to note the specific page numbers of any diagrams that you may wish to reproduce or refer back to.

It is tempting to photocopy large sections of books or print out journal articles without actually reading them and in doing so create your own library which you will read once it is compiled. Excessive photocopying can be a distraction from the process of actually reading the information and can end up being a waste of time. It may be far better to read articles as you find them – each article you read will direct your thinking in a certain direction and thus reduce the chance of reading the same thing many times from many sources.

Many researchers have also fallen into the trap of using their background reading as a distraction from the harder task of actually carrying out the research itself. Try to gain a sense of when to stop reading and remember the marker of your Independent Investigation does not necessarily need to know absolutely everything in the history of your chosen topic.

iv – A Guide to **Referencing**

Royal Geographical Society with IBG

Advancing geography and geographical learning

Before geography researchers carry out any fieldwork they will always read around the subject matter and make notes on what kind of research has already gone before them on that topic. When they write-up their overall findings, many will include a **Literature Review** or a summary of what they have read during this time and what it has told them about their area of research.

Plagiarism:

Using someone else's ideas or wording and making them appear to be your own.

In your Independent Investigation you may also decide to write a Literature Review; you might quote others who have studied your chosen topic. In order to avoid plagiarism, and to commend the work done by these researchers, it is important that you **reference** these writers – this means you acknowledge in writing from where exactly you have obtained certain bits of information.

In most formal research there is a standard method for referencing known as **Harvard Style Referencing** – and this is the method you should adopt for your Independent Investigation. This method, along with an accompanying **Bibliography**, allows the reader to find the exact source from which you found the idea.

Harvard Style Referencing

- Referencing to show credit for ideas

  You may read about a particular idea in a book, journal article or on a website and wish to talk about this idea in your Independent Investigation write-up. This can be done in two ways; by quoting the surname of the author and the year the piece was written in brackets after you have spoken about the idea, or alternatively, by using this information within the flow of a sentence.

… and in some regions the need for food security may actually create justification for land grabbing (Nally, 2015).

or

Though as Nally (2015) explores, in some regions the need for food security may actually create justification for land grabbing.

If the article is written by more than one person, the latin 'et al.' (meaning 'and others') is used after the principal writer.

Despite European regulations for urban sustainability, cities like Sarajevo have actually developed unsustainably due to the urgent need to deal with vulnerable groups (such as homeless people) (Martin-Diaz et al., 2015).

or

Despite European regulations for urban sustainability, some cities have still developed in an unsustainable way. As found by Martin-Diaz et al. (2015), the urgent need to deal with vulnerable groups (such as homeless people) was given priority over sustainability targets.

- Referencing direct quotes

  Whilst reading you may wish to directly quote from a book, journal article or website. This is usually done within the flow of a sentence, using quotation marks where necessary. However, it is worth noting that copying out long passages from an article or website, even when correctly referenced, is simply not good academic practice – in all cases you should aim to use your own words as far as possible, as few marks can be gained from simply copying out text.

…and as Nally points out

> *"food security supplies a moral sanction for land grabs (2015)"*

or

The same cannot be said of all cities. In Sarajevo,

> *"the built environment of the city has moved in an increasingly unsustainable direction as a result of the need to deal with vulnerable groups in the population" (Martin-Diaz et al., 2015).*

- Completing the referencing system by using a bibliography

  At the end of your Independent Investigation you should include a bibliography showing the full reference for any sources of information you have used. These sources should be listed in alphabetical order by author surname.

  For journal articles the following format is used:

  Surname, Initial. (Year) Article title, Journal title, Issue number, Page numbers

Martin-Diaz, J. Nofre, J. Oliva, M. and Palma, P. (2015) Towards an unsustainable urban development in post-war Sarajevo, Area, 47:4, p376-385

Nally, D. (2015) Governing precarious lives: land grabs, geopolitics, and 'food security', The Geographical Journal, 181:4, p340-349

For books the following format is used:

Surname, Initial. (Year) Book title, Publisher, Place of publishing, Page numbers

Lovelock, J. (2000) Gaia: A New Look at Life on Earth, Oxford University Press, Oxford, p100-114

When using websites, it is not always easy to find the exact name of the person who has written the article. In these cases, an organisation name can be used.

Organisation name (Year) Website article title, Website address, Date website was accessed

Defra (2005) Climate Change: What is Climate Change?
www.defra.gov.uk/environment/climatechange/about/index.html, Accessed 1st May 2016

At the end of your Independent Investigation, it should be absolutely clear to the reader which ideas are from someone else and which ideas are your own. Good use of referencing is essential to authenticate your own work, and also essential to avoid accusations of plagiarism.

Section 1 – Introduction

The first section of your Independent Investigation should introduce the reader to what your research will aim to do and give a focus to your whole research process. Your planning stage will have given you a clear idea of a broad research area as well as some more detailed research questions which you will aim to answer.

One of the first things you can do in your introduction is state the broad **Aims of the Study** – what at things you hope to achieve and, very briefly, how you hope to achieve them. Before stating the more detailed research questions or hypotheses, it may be a good idea to provide the geographical context to your study through a discussion of what you have read in your background reading. Formally this can be called a **Literature Review**. Citing from a wide range of sources, a Literature Review should aim to:

- Show what is already known about your topic by geographers
- Show any models or theories in relation to your topic
- Show where gaps in the existing geographical knowledge occur (and how your study aims to fill them)
- Suggest why it is important to study this particular topic

In order to avoid plagiarism, it is important to remember to reference any published work which you are using in this section.

If well written and comprehensive, the Literature Review should naturally lead the reader to consider the **Research Questions and Hypotheses**, and no further explanation of their inclusion should be needed.

The introduction should also include a **Justification for the Investigation**. This should explain to the reader why it is important that studies such as yours are done and what value they might hold in the wider geographical world.

The **Location** of your study should also be explored. This can be done through a written description as well as visually through a series of linked maps, showing the location of the research at increasingly more detailed scales. Use of mapping websites can be useful here as can GIS packages; ask your teacher about what is available.

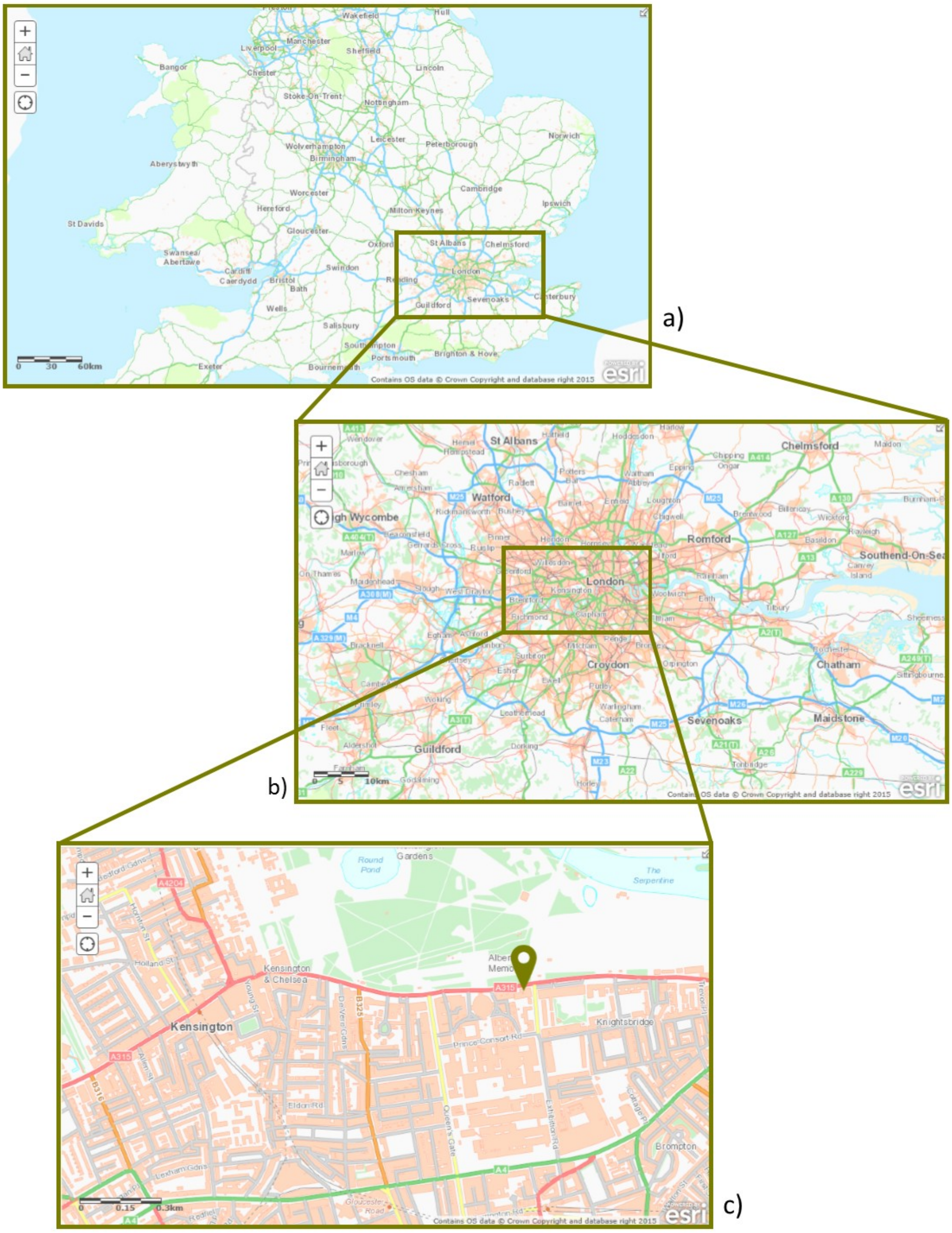


Figure 1: The location of the study site in the context of the UK (a), London (b) and Kensington (c).

Map Source: ArcGIS Online

The introduction can also include a **Study Framework** showing the different stages you are going to go through in order to meet your aims and be able to answer your research questions. This framework may be presented graphically in the form of a flow diagram, so the reader can clearly see how your thought processes will develop through the research procedure. It may read a little like a list but should not appear as one; a simple discussion of what you are planning to do at each stage of the Independent Investigation is sufficient.

Common Pitfalls:

- **Discussing your study location as if the examiner knows it intimately.** Your teacher may know where you plan to carry out your Independent Investigation but few others will be able to locate a site by name unless you show them its location on a map.
- **Justifying your study by claiming it will change all geographical thinking.** As much as you might value your study and fully believe in it as a piece of 'original' research, it is unlikely to change the course of geographical history. It is far better to speak modestly of possible outcomes and describe how it may inform others to research the topic in more depth.
- **Repeating a research question that you have already found the answer to in your Literature Review.** If you have already read about it, there is little value to your research. Instead, why not see if someone else's theory holds true in your own location?
- **Using research questions that are either too vague or too closed.** Make sure your research questions and hypotheses are actually answerable by providing them with the right wording and level of detail.
- **Discussing possible conclusions in your Study Framework.** If you have not yet carried out the investigation then it is not possible to draw conclusions, even if you do have a strong idea about what the outcome might be.

1a – A Guide to Hypotheses

Royal Geographical Society with IBG
Advancing geography and geographical learning

Hypotheses are simple statements that a researcher uses in the early stages of a piece of research to state how their data may create a conclusion. They should not be viewed as a prediction but instead a proposed explanation based on the limited evidence the researcher already has or based on the geographical theories about which the researcher has already read.

They are used to give focus to the research and are particularly useful in statistical tests. In these, it is often the case that a **Null Hypothesis** is used. This is a statement which goes against the general trends or theory that is already known. In empirical research, such as that carried out by most geographers in the field, it is ultimately impossible to prove truth exists. As such it would be very wrong for the geographer to claim to have found a positive statement (an **alternative hypothesis**) to be true. In fact, all they have done is simply proven that a false statement is not true. Geographers who consider this line of thinking challenging should think about a criminal court in which a person on trial has to be assumed innocent until proven guilty. It is the prosecutor's job to prove guilt not to prove innocence. In the same way, one's data has to be considered false until a known falsehood has been proven as such.

The null hypothesis therefore serves as a means of allowing geographers to draw conclusions when data, by its nature, cannot provide absolute truths.

For example, geographical theory suggests that the bedload of a river should decrease in size with distance from the source of the river. Therefore, a sensible positive or alternative hypothesis (shown as the symbol H_1) would be

$\boldsymbol{H_1}$: "The size of the bedload in River X is directly correlated to the distance in which it is found from the source."

However, a good researcher should not look to prove true a statement they suspect to be true to begin with. Instead they should prove a false and opposite statement to in fact be false first. Therefore, a null hypothesis (shown by the symbol H_0) should also be quoted: giving the geographical researcher scope to conclude that a false statement is indeed false.

$\boldsymbol{H_0}$: "The size of the bedload in River X is in no way correlated to the distance in which it is found from the source."

In the process of the investigation the researcher is tasked with finding the evidence to reject the null hypothesis, usually through the use of statistical tests such as **Chi-Squared** or **Spearman's Rank Correlation Coefficient**.

For example:

"The results show a negative correlation between the size of the bedload and the distance from the source. This allows me to reject my null hypothesis which states that there would be no such correlation and therefore accept my alternative hypothesis."

Section 2 –
Data collection

Royal Geographical Society with IBG
Advancing geography and geographical learning

Data is any kind of information that, once analysed, can help you to answer your research questions and either prove or disprove your hypotheses. The **Data Collection** section of your Independent Investigation is the point where you describe and explain the methods you plan to use, or those you have used, in order to get hold of your data.

Earlier planning stages of your investigation should have given you a clear idea of what data you actually need to collect and by having discussions with your teacher you will be able to gain advice about where and when you will be able to collect this data. You may even be able to borrow some field equipment from your school to help you collect the data. It is important to remember that there is no definitive set of rules about how you should collect your data – each investigation is individual and so the circumstances of one method may work well for one person and their investigation but may not be suitable for another at a different time or location.

Think inventively about how you might gain data. If the data you require involves, for example, measuring the frequency of something then it may be enough to simply count and record the phenomenon passively. If, however, you are looking to see how people feel about a certain idea, you could see whether they respond positively or negatively to different statements or ask them to come up with three words to describe the idea: inventive data collection techniques that really target the type of data needed for the investigation will certainly create more interest for the reader and marker of your study. In order to produce a good range of data presentation techniques you may need to ensure that your data collection methods produce data that can then be used numerically, graphically and cartographically.

It is also important to make sure that the data collection method you choose allows you to do something meaningful with the data it provides. Good ideas for studies can sometimes produce meaningless results, not because the research questions were flawed, but because the manner in which the data was collected only allowed for certain conclusions to be made.

Research Question:

2. What have been the primary reasons for EU migration to Lincolnshire?

Data Collection method for this Research Question:

Ask farm workers as part of an interview what their primary reason for moving to Lincolnshire was.

If you ask one hundred farm workers this question you may end up with one hundred different answers and no way of bringing their responses together to conclude anything meaningful.

Instead before you interview the farm workers, maybe come up with six to eight possible answers for the respondents to choose from, as well as an 'Other' category where they can create their own response if none of your answers 'fit' their ideas. This way you can turn their responses into numerical data, showing which answer gained the highest number of positive replies.

Research Question:

2. The highest levels of pollution will be found in areas that have the median levels of socio-economic status.

Data Collection method for this Research Question:

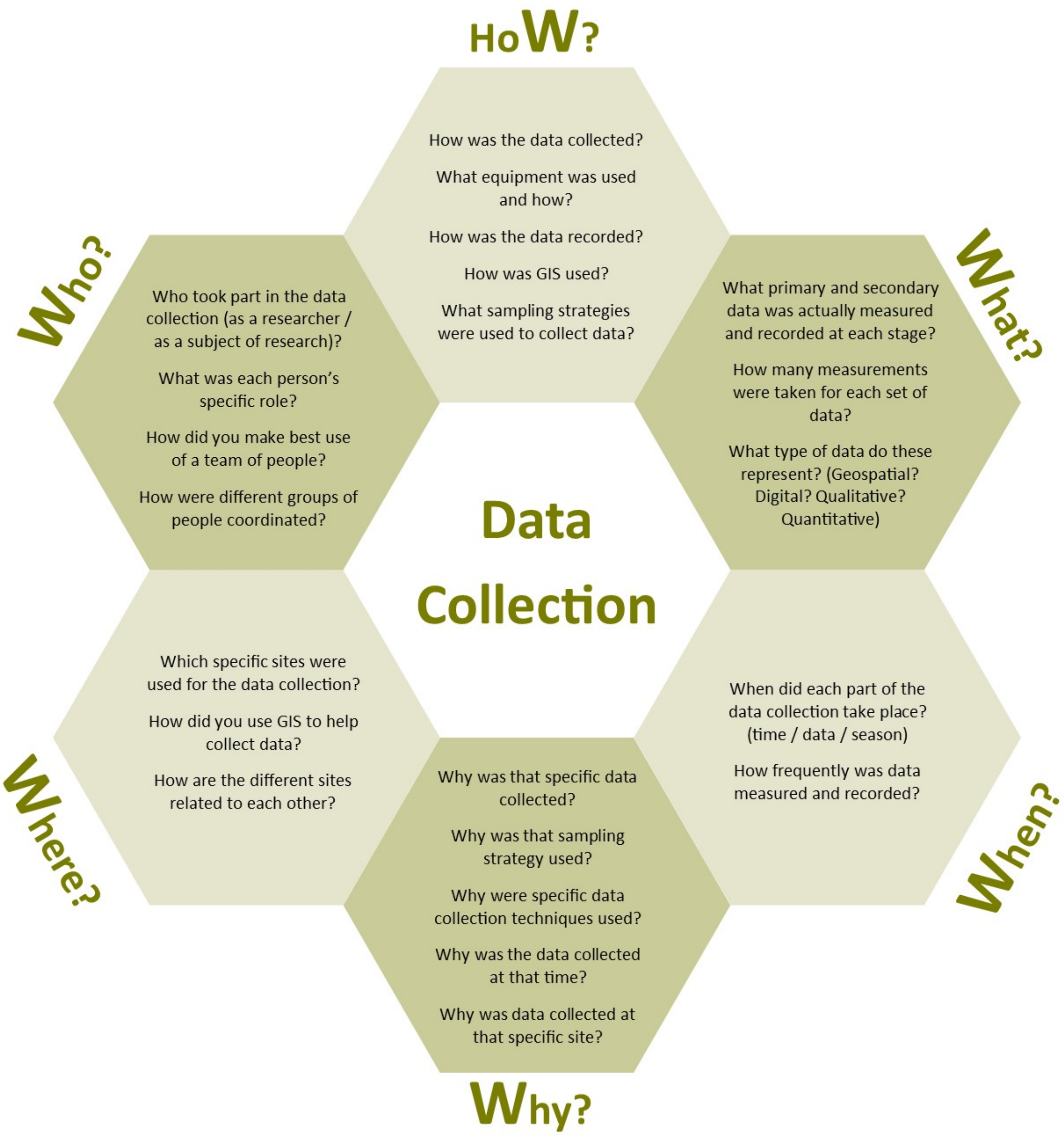
Sample the river water in the River Chelmer in Chelmsford and use it to test the acidity and level of phosphorous.

This data collection technique will only tell you about the water quality at that particular point on that particular day and very few conclusions can meaningfully be drawn from just this data.

A variety of samples should be taken at each site, at different times of the day and at slightly different points. This way you can find the mean, median or mode of the samples during the Data Analysis stage. Doing so will allow more accurate conclusions to be drawn and any conclusion made will stand up to greater scrutiny as they are based on more than one piece of data.

The **Data Collection** section of your investigation can either be written before you go into the field (in the future tense) or after you return (in the past tense). The latter of these may be slightly easier to do if you have never visited the data collection site prior to when you collect your data. For example, this may be the case if you are visiting the site on a school organised field trip and are leaving some of the precise details of the data collection methodology to be decided once you have seen the site.

It may be useful to use the **'6W' questions** as a framework in the write-up of your Data Collection section. A key indication of a well-written Data Collection section is if someone who does not know anything about your topic or the locality of the data collection site is able to go out and recreate exactly the methods you used. Think of your data collection methods as an instruction manual that may enable another student to repeat your fieldwork at a later date. Writing in a '6W' framework may go some way to ensuring that you have all the relevant details covered. Not all of the questions listed below may be answerable due to the particular type of research you are carrying out but most of these questions are likely to require answering by most researchers. You may also wish to use your research questions as a way of structuring your Data Collection write-up: citing the data needed to answer each research question and then describing and explaining the method you have used to collect that data. You should certainly **refer back to your research questions** when you think about why you chose to carry out the data collection in a particular way: it is important that the reader of your study can see that you are keeping the research questions at the centre of everything you do.



Using **GIS** may be a useful way of both recording data in the field and ensuring it is easier to process once back in the classroom. If any part of your data collection has a geospatial element (i.e. the data is recorded in different places) then location information should be recorded. This spatial data, once entered into a spreadsheet will allow you to map any results you produce.

Depending on the type of data you are collecting you may have to carefully consider how you might manage any **ethical or social issues** that may arise. Names and addresses, for example, may have to be kept anonymous and you may have to behave sensitively to personal opinions on social challenges and what information people might be prepared or willing to share with you (remember they are not obliged to answer your questions). In a natural environment, think about the measures you will take to minimise your own negative impact on the area and whether the act of you collecting data actually causes harm to the environment you are in. Whilst it is very difficult to plan for every eventuality, describing in your write-up how you best tried to manage difficult circumstances will show how well you planned your methodology.

No set of data collection methods is flawless and an acknowledgement of this in your Data Collection write-up puts you in a good position with the reader of your study. Briefly highlighting the **limitations** of your methods is very good practice at this stage.

Common Pitfalls:

- **Putting off writing up your Data Collection section once you return from the field site.** You will be surprised how easily key details that you thought you would remember are forgotten once you return to the classroom. Avoid this by making notes in the field on everything you consider important. Then write up your Data Collection, even just in skeleton form, as soon as you can once you get back.
- **Badly organised or illegible field notes.** If you cannot read the raw data once you are back in the classroom, it has little value for your study. If environmental conditions were against you on the day of your data collection and your notes look weather beaten, smarten them up at the first opportunity.
- **A lack of detail in the write-up of your methods.** The person reading your study was most likely not with you when you collected the data, so do not assume they will automatically know what you are talking about when you skip over certain details.
- **Using vague justifications for your chosen methodologies.** It is not acceptable to say that the reason you chose a particular data collection method was because ‘it was easier’. At least describe why this method made things easier, or better still, really think about how that method improves either the quality or quantity of the data.
- **Using a method that closes the scope of the study.** If you only collect data at ‘location A’ then you can only comment on what is happening at ‘location A’. However, if you collect data at a number of points you may be able to show how something changes over a distance or how other factors may be involved, placing you in a position to make higher level conclusions.

Other information

See the relevant specification for further guidance on lone and group working during Data Collection

---

2a – A Guide to
Different Types of Data

Royal Geographical Society with IBG
Advancing geography and geographical learning

Data is a collection of facts and numerical statistics that can be drawn together and analysed in order to answer a research question or start to draw wider conclusions.

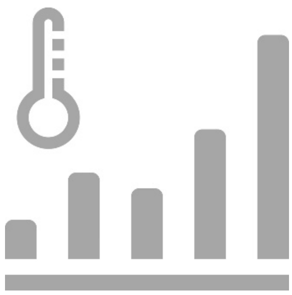
Researchers think carefully about the type of data they wish to collect before they start planning their data collection methodologies. Broadly speaking, there are two types of data:

- **Qualitative Data.** This is data that does not hold a numerical value.
- **Quantitative Data.** This is data that can be measured using numbers.

For example, the average distance a person travels to work is quantitative data, while the place they travel from is qualitative data.

Quantitative data can be further divided into **discrete** data and **continuous** data. Discrete data is that which can only be measured using a specific numerical value (such as a percentage). Continuous data can take any value and in theory can carry on to an infinite number.

For example, the percentage of days in a year where the maximum daily temperature in the UK reaches over 20°C is discrete data because there is a limit to the number of values that can be chosen as the answer (in this case because we are considering percentages there are 101 possible choices – the values zero to one hundred).



However, if a researcher wanted to measure the average maximum daily temperature for the UK over a year, they would be looking at continuous data because, in theory at least, there is no end point to the possible highest temperatures recorded.

If the research demands that relationships between different variables be examined, it may not always be easy to find these relationships using qualitative data. However, in some cases, qualitative data can be translated into quantitative data. For example, in a survey of three hundred people, rather than record the exact address from which each person travels to work (qualitative data) one may be able to find the number of people who travel into work from a particular postcode (quantitative data).

Data can also be described as:

- **Primary Data** is that which is collected by the researcher first-hand, usually at a field research site.
- **Secondary Data** is data from other sources (usually previous studies) which the researcher uses.

Most researchers will use a combination of both primary and secondary data during their data collection process.

2b – A Guide to
Data Collection Techniques

Royal Geographical Society with IBG

Advancing geography and geographical learning

The way in which a researcher goes about collecting their data is very much dictated by the focus of their research and their specific research questions that require answering. The data collection methods they employ may also be designed with the particular research site in mind as well as the time and resources they have available to them.

There is no such thing as a definitive way of collecting a particular piece of data: each piece of research is individual and while certain techniques (such as those suggested in this series of guides) are well used by geographers, they should also be adapted to suit the specific conditions and needs of the investigation.

Quantitative Measuring

Many geographical research projects will involve some form of numerical measurement of a phenomenon such as width, depth, velocity, temperature or mass. This is particularly true of studies into elements of physical geography, where measurements in the field at different geographical locations allow the researcher to compare one area with another.

These measurements may be taken at a number of selected single points or along a transect - a line on a map that links two points and along which a researcher takes regular recordings. A researcher may also wish to take measurements at the same point a number of times over the course of a specified time period if they want to see how a phenomenon changes over the space of a day, a week or even a year.

Many school and university departments will have a range of fieldwork equipment which a researcher can utilise for their study. This equipment is designed to make quantitative measuring easier in the field. Some common fieldwork equipment and its uses are listed below.

Image: Flickr CC User Aidan

Anemometer: Used to measure wind speed.

Image: Flickr CC User Tre Briercliffe

pH meter: Used to measure acidity or alkalinity of a water sample.


Image: RGS-IBG

Callipers: Used to measure the dimensions of small objects such as stones and pebbles.


Image: RGS-IBG

Clinometer: Used to measure the angle of a slope.


Image: Alex Segre

Light meter: Used to measure the amount of light or cloud cover.


Image: Flickr CC User Jonny Kriedler

Compass: Used to find out a direction or bearing.


Image: Rebecca Stokes

Flow meter: Used to measure the velocity of moving water.


Image: RGS-IBG

Quadrat: Used to measure species abundance in a set space.


Image: Flickr CC User Tim Ellis

Stop watch: Used to measure periods of time.


Image: Anne Vaughan

Sweep net: Used to collect invertebrates from a water course or a tree canopy.


Image: Alex Segre

Rain gauge: Used to measure precipitation levels.


Image: RGS-IBG

Tape measure / Meter stick / Trundle wheel: Used to measure distance, height, width or depth.

Image: Flickr CC User Toshiyuki IMAI

Thermometer: Used to measure temperature.

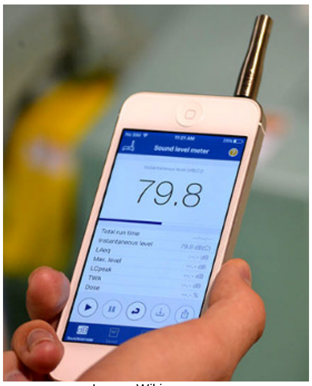


Image: Wikicommons

Smartphone Decibel Counter: Used to measure noise levels.

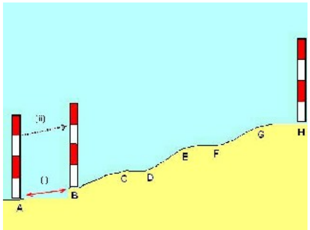


Image: RGS-IBG

Ranging Pole: Used to mark out features that occur in straight lines.

Sometimes the data itself cannot be collected directly in the field and a sample is taken which can be analysed in more detail back in a laboratory or classroom.

Handheld devices such as smartphones or tablets can be of enormous use to the researcher in the field. Quantitative measurements can be recorded directly into spreadsheet which allows the researcher the ability to manipulate and analyse the data more easily at a later date. Their use can also allow the researcher to easily download the data into a GIS package, enabling them to map their results effectively. Further guidance is available at www.rgs.org/dataskills.

The way a researcher goes about collecting their data is very much dictated by the focus of their research and the specific research questions that require answering. The data collection methods they employ may also be designed with the particular research site in mind, as well as the time and resources they have available to them.

There is no such thing as a definitive way of collecting a particular piece of data: each piece of research is individual and while certain techniques (such as those suggested below) are well used by geographers, they should also be adapted to suit the specific conditions and needs of the investigation.

Surveys

Surveys are personal studies that the researcher undertakes in the field. Using their own judgement and opinion the researcher observes the geographical phenomenon in question and records information or formulates a score that summarises what they are seeing. There are many different types of survey, and individual surveys can be designed with specific research needs in mind. Where appropriate, consider using apps like Survey123 for ArcGIS, Collector for ArcGIS and Fieldwork GB. The following three types of survey are commonly used by geographical researchers.

An **Environmental Impact Assessment (EIA)** measures the current or future possible impact a development may have or be having on a landscape. At a municipal level, their undertaking is now a statutory requirement before any major new development can be built, but in a more informal manner they can be used by geographers to compare different management options for a particular need (such as flooding or unemployment).

Commonly, EIAs are a matrix with two parts: firstly, a range of possible causes of impacts (such as traffic) against the different aspects of the environment which could be affected by any impacts (such as local people).

	Works Traffic	Location of Raw Materials	Disposal of Waste	Building Processes	Proposed Site	Comments
Plants	0	0	−2	−2	−3	
Animals	−1	0	−1	0	−2	
Noise	−3	−1	0	−2	−1	
Air Quality	−3	−3	−1	−1	0	
Water Quality	0	0	−3	−2	− 2	
People	−1	+3	−1	0	+3	

In each part of the matrix a score is given to show the perceived intensity of the impact felt, normally on a -3 (high negative impact) to 0 (no impact) to +3 (high positive impact). It is a good idea to also leave space for a comment to be made: the researcher may observe something in the field which explains a particularly high or low score and want to make note for it for further investigation.

The outside perspective of the researcher to the new development can be both a benefit and a problem to the undertaking of an EIA. As an outsider, the researcher can give a relatively impartial opinion of how they view the situation. However, they may also not have the full understanding of the phenomenon in question and so may not give a true representation of the real impacts a development may have.

A **Bipolar Analysis** is similar to an EIA but only takes into account a set number of criteria against which the researcher gives a score. This means that the researcher can use it to compare one location with another, or the same location over a set time frame.
For each criteria, paired, opposing words are used to create a scale on which the researcher places their opinion. The criteria used should also relate to the type of location that the researcher is investigating. For example, if the researcher is looking into the quality of different residential areas they may wish to comment on the upkeep of the pavements, while a woodland in a National Park will require different criteria. In the former, the two paired words may be 'smooth' (showing a positive score) and 'cracked' (showing a negative score).

		+3	+2	+1	0	-1	-2	-3	
Pavements	Smooth			✓					Cracked
Air Quality	Fresh		✓						Foul
Gardens	Well maintained				✓				Poorly maintained
Buildings	Well maintained			✓					Poorly maintained
Litter	Spotless					✓			Major
Street Furniture	Well maintained					✓			Poorly maintained
Noise	Quiet				✓				Loud
								Total Score	**+2**

Once the bipolar analysis has been completed a total value can be calculated by adding together all the scores generated, thus allowing the researcher to more easily compare one location with another. This is especially true if the researcher were to then plot the total scores onto a GIS map.

One challenge that comes with bipolar analysis is that the scores given are highly subjective to the researcher's own viewpoint. They have to ensure that their grading is consistent across all the field locations so that true comparisons can be made. It is possible that a researcher may grade an area with the lowest score possible only to move onto a new area and discover that it is dramatically worse than the last.

A **Land Use Survey** is a way of recording what geographical phenomena appear on the ground which cannot be understood from a map alone. Most often researchers use a Goad map to carry out a land use survey. A Goad map is a detailed street map that shows individual buildings and their plots, and is usually needed for land registry and insurance purposes. They can be purchased through online sources.

The job of the researcher is to note how different buildings are used (for example, municipal, residential, parkland or commercial) or how more natural landscapes might be categorised (for example, woodland, heath, grassland or moorland). This information might then be used to draw conclusions as to whether a particular dominant land use in an area corresponds to a particularly high or low value in another factor such as noise or air pollution.

Researchers may decide to survey a small area of a map, noting down the land use at a particular grid reference, for example, or as part of a team where they could map the land use of a much larger area. The use of GIS is especially good in conducting land use surveys as, with the design of a simple key, one can annotate an online map with the correct land use whilst in the field.

Interviews

An interview is a way of questioning someone connected to the chosen research topic. It involves more flexibility and a greater number of open questions than one might design for a questionnaire used with general members of the public.

Usually the process for an interview involves contacting a known person beforehand and asking permission to come and talk to them about a particular topic. These interviewees are usually connected to the issue under investigation – possibly someone in authority, involved in a local community group or someone directly affected by a possible change.

In some cases, the researcher may supply a sample of the questions they intend to ask so that the interviewee can prepare a resonance or find out the answer to particular questions. It is often the case that a set of interview questions is designed specifically for a particular person and rarely are they repeated to another participant in the exact same form.

As with a questionnaire, the wording of a set of interview questions should be carefully planned. Ambiguous questions can take the interview in an unintended direction and leading questions can sway the opinions of the interviewee unintentionally. Testing the interview on someone who does not know the aims of your study can be a good way of removing questions that could cause problems when you use it for real with the interviewee.

Generally, interview questions tend to be more open-ended and unlike questionnaires, can be expanded upon as the interview progresses. For example, the interviewee may bring up an issue that the researcher had not heard of or thought about and further, more spontaneous questions to the interviewee may result in a greater insight into an area of the study that previously had not been thought possible. A flexible approach to the questions to be asked is best and a good interviewer will change their interview plan as needed as the interview is happening rather than unnecessarily stick rigidly to their pre-written set of questions.

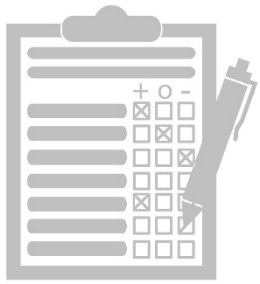
Before the interview you should check that the interviewee is happy for you to use their words as quotations, or alternatively, if they wish to remain anonymous. Equally, check the amount of time they can give you for an interview: some, less important questions may have to be omitted if you run out of time.

You may decide that it is best to record the interview to allow yourself to concentrate on what the interviewee is saying: it can be difficult to engage with someone if you have your head down whilst making notes whilst they are speaking. A digital voice recorder can be a good idea for this purpose, though do check with your participants that they are happy to be recorded in this way.

Focus Group Interviews can be an excellent way for researchers to allow people to discuss an issue without them feeling like they are taking part in a more formal process. A focus group is a group interview where the attendees are selected by the researcher and given a set of topics or questions to discuss. The researchers themselves do not play an active part in the interview itself and instead take on the role of observer, allowing the interviewees to discuss their opinions on an issue openly. Focus groups can work well when the researcher is looking into the possible impact of a new development, such as the building of a new supermarket, that may divide opinion amongst local residents. As with normal interviews, it is important to record the interview and check that the participants are happy to be quoted or if they wish to remain anonymous in the write-up.

Recruiting people to take part in a focus group interview can be difficult as they often last longer than a normal interview and all participants have to agree to be in the same place at the same time. Some researchers may decide to pay for people's travelling expenses or provide light refreshments for the participants if the interview takes place over an extended period of time. Those that have most time to take part in focus group interviews may belong to certain demographic groups (for example students and those who have retired) and not others (for example, those in full-time work) which may mean that your resulting interview transcript only reflects a certain section of opinion rather than a broader one experienced by a more diverse group.

Questionnaires



A questionnaire is a set of questions that a researcher may ask of a population set (usually members of the public) to find out facts about them and their opinions. Questionnaires are usually designed to be quick and easy to answer, with a focussed set of questions used to determine exactly what the researcher intends to find out. A set of ten questions is usually the limit to which people will respond.

The design and wording used within questionnaires are essential to their success. An ambiguous question could result in an answer that is not useful to the overall study and a respondent may not feel comfortable answering questions that they do not actually understand. For example a question such "Where are you from?" could mean a country of ethnic origin, a home address or the place they have travelled from that day. Equally, a researcher may ask a question to which the respondent simply does not know the answer, such as "How far in miles have you travelled today?"

Questionnaires should be a mixture of open and closed questions. Open questions allow the reader to choose their own response, while a closed question will have a limited number of answers; often just a yes or no response. An example of a closed and an open question is shown below.

Open question: "Use three words to best describe how you feel about the new pier."

Closed question: "Do you own or have regular use of a car?"

A questionnaire should start with fairly closed questions (such as whether they see themselves as a local or a visitor to the area in question). After that the questions may start to be more open, where the respondent may start to give their opinions on the particular issue the researcher is investigating. However, it is not necessarily true that in order for a respondent to give an opinion, the question has to be an open one – getting a respondent to choose from a pre-chosen list of adjectives can be a good way of managing responses and making them easier to analyse once back in the classroom. If it is important to ascertain the rough age category or the sex of a respondent, then these boxes can be completed by the interviewer themselves, after the respondent has gone.

The questions should be sensitive to what respondents may feel comfortable discussing. For example, a respondent may not wish to divulge their age, so it may be necessary to ask them to choose from a set of age categories. Equally, few people are happy to give details of their home address to someone they do not know, but when given the option of just providing their postcode, more respondents are likely to answer positively.

Once you have designed your questionnaire, it is a good idea to test it on someone who does not understand what you are trying to find out. Any glitches can then be ironed out before it goes live to members of the general public. You may also wish to carry out a pilot questionnaire in the field to test the suitability of your questions and the kind of data you might get back. You can then redraft the questionnaire and have a stronger and better prepared analysis as a result. The data from this pilot questionnaire should not be included in the final data set.

The process of recruiting people to take part in a questionnaire is worth some careful planning. Stopping people in the street can allow you to find more respondents than, for example, a door-to-door drop of a photocopied questionnaire that requires a written response. However, there are issues with on-the-street questionnaires that, due to the times when a researcher undertakes this process, a certain demographic of a population may be over represented compared to others. This in itself may, or may not, have an influence on the nature of the conclusions the researcher is able to draw. There are several ways of collecting questionnaire responses using a link to an online survey (e.g. Google Forms or SurveyMonkey), which may be appropriate and should certainly be considered.

As part of the recruitment process you should also consider the sampling strategy you employ. The separate guide in this series gives further details about the types of sampling open to use but the researcher may wish to consider a random selection (such as questioning every tenth person who walks past the interviewer) or a systematic sample (such as ensuring there are equal number of respondents from different age groups). Equally, the researcher should consider how many completed questionnaires are needed in order for any conclusions drawn from them to be deemed valid. In most research a minimum of thirty responses is sought initially. In some cases, the researcher will carry out a rudimentary data analysis at this stage before going back into the field to collect a further fifteen or twenty responses. If the trends in the data remain the same after the second round of responses, then it is likely that the researcher has enough data to work with.

If you are undertaking a face-to-face survey, consider whether you will read the questions to the respondent or have the respondent read them themselves. Not every member of the public may have a grasp of technical geographical language, nor indeed have English as their first language and some questions may have to be worded differently so that any person is able to answer them. If the researcher is giving the respondent a list of options, these may need to be printed separately so they can consider them in their own time. Few people are likely to remember a list of eight or nine options when they are read out to them.

It may go without saying but it is not uncommon for market research companies and other commercial enterprises to be carrying out surveys in the same space as you and as a result recruitment can be challenging as respondents may feel they are being bothered. Introducing yourself and the purpose of your research briefly as well as being polite, approachable, smiley and grateful can go a long way to getting a higher number of respondents. It is also worth remembering that the respondent has entered into the questionnaire willingly and if they decide part way through the questions that they would rather not continue, or that they wish to change an answer, they should of course have the ability to do so.

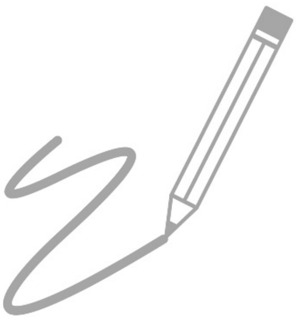
Observations, Photographs and Field Sketches

Not all forms of data collection have to involve other people or the direct collection of numerical statistics. Spending time in the area under study and making notes on any **observations** you see can be a useful way of adding value to your Independent Investigation. Observing the study area should always be done alongside any more formal quantitative data collection and may provide explanations for anomalies or unusual patterns within the data once analysed.

For example, pure numerical data on the flow rate of a river course may show that the river slows its velocity considerably at a certain point. This may be unhelpfully recorded as an anomaly in the investigation write-up. However, the use of observation notes would show that this decrease in velocity actually happened just after a natural or human-made obstruction in the riverbed which had impeded the flow of the river.

Observations can also be more open, especially when the study area might be a place that is not well known to the researcher. Investigating rebranding in a town centre might be enhanced, for example, by sitting for an hour in a town square over a number of days to see how different types of people use the space and how it transforms itself to different purposes over the week as well as how it may be misused.

Taking **photographs** or recording short **videos** of an area can also be good ways of obtaining qualitative data. These observations can be used, albeit subjectively, on the part of the researcher to show how one might feel about a particular issue, new development or natural area. **Mental maps** can be another good way of recording this information. These maps are drawn as the researcher tours an area, and record how the researcher feels about a central issue in different locations. For example, in a rebranding investigation, the researcher may record how safe or unsafe they feel in certain locations around a town, either through symbols, or by emphasising a certain feature they see and drawing it in the form of a map. The researcher can equally ask local members of the public to also create their own mental maps for comparisons with the researcher's.



Annotated Field Sketches are simplified drawings of a fieldwork site, produced in the field. The term 'sketch' can be a little misleading as it suggests that a high level of artistic ability is needed to produce a field sketch. In fact, a field sketch is a line drawing that only highlights the key features of the view that the researcher finds important geographically. They are neat, black and white drawings that do not usually require sophisticated shading or any form of 'artistic detail'.

Selecting the view to be drawn may be the most important stage of a field sketch – some geographers find a cardboard frame held up to the view useful for this exercise. Dividing the sketch into the foreground, middle-ground and background, and drawing each section separately, can help researchers who are nervous of drawing a complicated scene, as well as focus the sketcher on each part of the scene for details.

Once drawn, the field sketch can then be annotated with notes that describe and explain the most important features. Processes, and their effects, can be highlighted and relationships between people and the environment can be made. Features that are not easily identified from the field sketch, such as slope angle, rock type or percentage vegetation cover can also be noted on the drawing, allowing the viewer to create a stronger understanding of the scene. If a series of field sketches are to be used, it is a good idea to note next to each one the date, grid reference and direction of view so that comparisons can be made between them.

Field sketches and photographs can also easily be used alongside GIS. Both can be situated on a map at their respective locations and used, alongside other data, to show how a landscape may change over a geospatial area.

2g – A Guide to
Sampling Techniques

Royal Geographical Society with IBG

Advancing geography and geographical learning

A **Sample** is a selection of data chosen from all of that possibly available. Sampling is needed in almost all forms of data collection as in most research processes it is simply not possible to gain data from every available source. For example, if one wished to conduct some interviews within a town it would not be possible to interview absolutely every resident of that town. Instead, a selection of the population would be used to try to get a representation of the town's answers. The method you use to select this sample is known as your **Sampling Technique**.

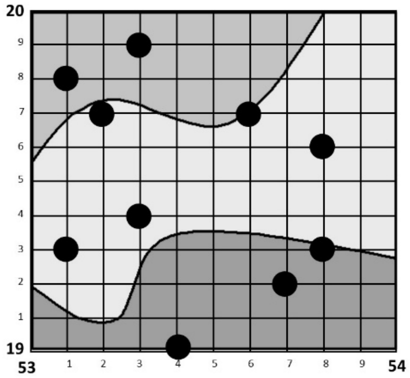
Sampling not only makes conducting your data collection possible, it can also make more efficient use of the time you have to collect your data. It may not even be necessary to collect all points of data for your research, and in many cases your overall population size (or your **Sampling Frame** – the pool of data from which you are drawing a sample) may be an unknown quantity; as a researcher you would therefore not know when to stop collecting data.

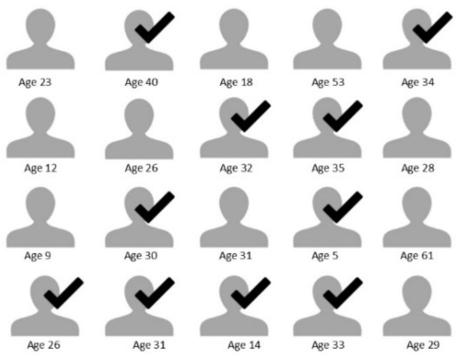
Researchers frequently put great effort into deciding on the size of their data sample. The larger the **sample size**, the more representative it is likely to be of your overall sampling frame, and as a result, the more justifiable your conclusions will be. However, the size of any sample is also dependent on the time and resources you have available and how manageable the data collection, and data analysis will be as a result.

While there are many different sampling techniques, there are three common methods used frequently by researchers: **Random**, **Systematic** and **Stratified Sampling**. In the following explanations, two examples will be used. In the first, Researcher A is trying to find a sample of ten data collection locations on a map that contains three different geographical zones. In the second, Researcher B is trying to select a sample of ten interviewees from a population of twenty people of different ages.

Random Sampling

Random sampling is where sources of data are chosen in a completely haphazard way. Once the size of the sample has been decided (maybe as a percentage of the overall sampling frame), researchers use random number generators, which can be found online, to give completely random sets of numbers. These can then be used to create grid references for data collection sites on a map or tell researchers which house numbers to survey within a street.





Alternatively, if the researcher is looking for a random transect line on a map, a random number generator can give the grid references for the start and end points of that line on a map.

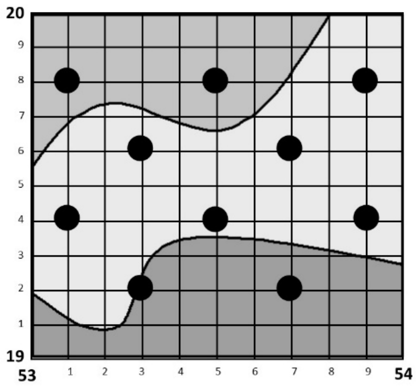
For surveys of natural terrain, where the researcher plans to use a quadrat, a common method for choosing random survey sites is to stand in the centre of the area and throw the quadrat with one's eyes closed. Surveying where the quadrat lands, and then repeating the method from that spot can create a random selection of sites.

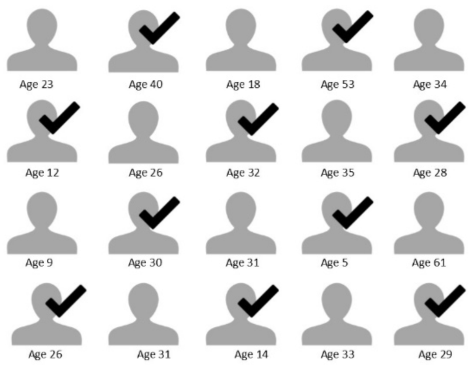
Rolling dice, choosing unseen playing cards, and picking bingo numbers out of a bag can be other ways of making random number selections.

Though randomly generated numbers take a human choice element out of the sampling process and so reduce the chance of human bias in the results, random sampling in general is not always suitable for small sampling frames as there are limited choices to be had.

Systematic Sampling

Systematic sampling is where sources of data are chosen in a completely non-random way. Here the size of the sample may not necessarily be decided before the sampling begins as the chosen system itself may create the sample size on its own. The interval size between sampling points (distance on a map, or every n^{th} person in a survey) is chosen by the researcher and stuck to without compromise.

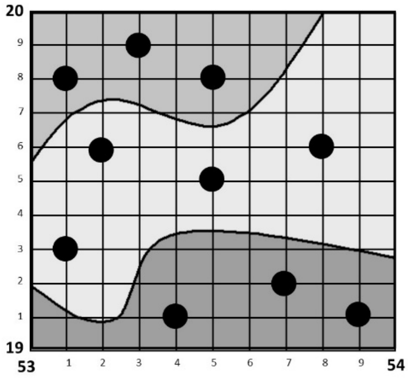


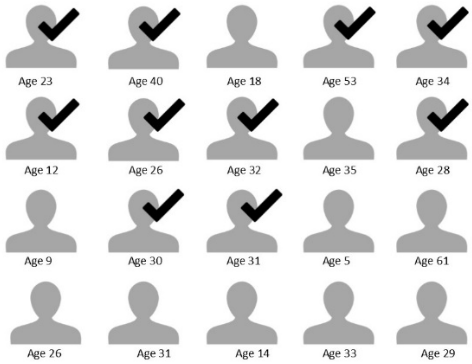


The benefits of systematic sampling are that the researcher is largely removed from the selection process and therefore bias can be avoided. However, in order for the sample to be truly representative of the study area, the researcher must also ensure that the sampling frame itself does not inadvertently create bias. For example, if the sampling frame for a survey were to be taken from a pre-selected list of people, such as the electoral role, it would automatically exclude people who were not eligible to vote, such as those aged under eighteen and some prisoners.

Stratified Sampling

Stratified sampling involves splitting the sample frame into smaller groups or **Strata** and using these strata to 'weight' the sample chosen accordingly to represent the original sampling frame. So if it is known that thirty percent of the sample frame came from a particular location were of a particular age group or belonged to a particular religion, thirty percent of the sample would also represent these strata.





Researcher A has placed more location points in the middle area as it represents a larger percentage of the overall grid square than the other two.

Researcher B has worked out the percentage of people that fall into different age categories in the original sampling frame. They have then used these percentages to determine how many of each age group should appear in the actual sample.

Many researchers believe that stratified sampling represents the most unbiased of the three techniques mentioned here as the sample becomes a true representation of the original sampling frame. However, an extensive knowledge of the sampling frame may be needed before a researcher can choose the strata and a pilot study may be needed in order to decide on the 'weighting' for each strata that is used.

It may be possible, depending on the exact nature of your research, to combine different sampling techniques together. For example, you may choose to draw a random transect line across a data collection site on a map, but then use a systematic or stratified sampling technique to choose the exact points of data collection thereafter.

2h – A Guide to Recording Data in the Field

Royal Geographical Society with IBG

Advancing geography and geographical learning

When geographical researchers plan their data collection, they also plan how the data is recorded in the field. A range of techniques can be used; a simple pencil and clipboard, smartphone or tablet, or recording straight into a GIS package.

The method you use to record the data should suit the environment you are investigating as well as the nature of the data you are trying to record. For example, data from a questionnaire will need to be recorded quickly and efficiently so the respondent is not kept waiting unnecessarily. Equally, depending on the environment you are working in, it may not be appropriate to be carrying electronic equipment which is not waterproof. Additionally, some devices may need a Wi-Fi or mobile data connection to upload data to a web-based site; in remote locations these connections are less likely to be reliable. If you are using a smartphone, tablet or other device, make sure it is fully charged and you have a portable charger.

Before you go to the field site, you need to be prepared for being able to record data in different weather conditions. High winds and strong rain can make recording data more of a challenge and some well-prepared recording equipment for these conditions can ease the role of the researcher. Traditionally, this may involve numerous spare paper data recording sheets and plenty of plastic bags to keep everything dry. However, smartphones can provide you with an excellent means to collate many different types of data in one place without having to juggle lots of different pieces of data recording equipment. Written notes, numerical data, photographs, videos and sound recordings can all be logged on most smartphones and provide an easy, pocket-sized way of keeping all the data in one place. It may be worth trialling the methods you plan on using first before you head to your data recording site. This will allow you the chance to test what data recording methods work and do not work before you are out in the field and are stuck with a technique that appears to have only looked good in the classroom. In some cases, such as those where there are inherent risks associated with the data collection, this level of preparation is essential.

Using GIS in the field to record data is becoming increasingly common. Many smartphones and PDAs have built in GPS systems that allow the user to record the exact grid location of where the data is being collected. This means that in the field, or back in the classroom at a later date, you can use GIS to locate fieldwork data and present it cartographically within the Independent Investigation. You could also consider using survey apps like Survey123 for ArcGIS, Collector for ArcGIS and Fieldwork GB.

Some more technical devices such as decibel meters (also available as an app) and flow meters can be connected straight to a laptop or tablet. This allows data to instantly be downloaded into useable forms such as Excel spreadsheets, from which you can create effective data presentation. Some pieces of data collection equipment may also have data storage devices within their hardware. This enables researchers to have devices run remotely, without the need for first-hand operation. Once the researcher returns to the device, all the data it has recorded in the time period they have been away can be downloaded and used without decoding.

Older data transfer techniques using SMS messaging may still be useful in areas where there is a 2G mobile signal, but no reliable data connection.

As with any part of your data collection methodology, the recording of your data is equally open to evaluation and a good researcher remembers to make a note of justifications and limitations of the data recording methods while in the field, just as they would for the data collection methods themselves.

2i – A Guide to
Avoiding Biased Data

Biased data is that which cannot necessarily be trusted as a true – or as true as is practicable – representation of the geographical phenomenon being researched. The integrity of the data in question has in some way been compromised as a result of the way in which it has been collected or by the inherent nature of the data itself.

While it is difficult to ever be sure that all traces of bias have been removed from an investigation, researchers can go a long way to reduce its amount and effect by carefully planning their data collection methods so that the way in which the data is brought together is as fair as possible.

Firstly, the **nature of the sampling frame** may in itself be biased. The chosen population set for a questionnaire, for example, may exclude people from a certain age group, or socio-economic background, making the data collected non-representative of those groups from the beginning. If this is the intention of the researcher, then of course this should be made explicit to the reader of the Independent Investigation, but otherwise the researcher should try to include all those in the sampling frame from which they may expect to make a conclusion.

Equally, the **size of the sample** used may not be high enough to really show the true diversity of the data available, nor allow the researcher to really comment with any certainty about any conclusions they try to make. In general, the larger the sample size, the more certain the researcher can be that the conclusions they make are an accurate representation of the geographical phenomenon. However, the researcher has to also consider the time and resources they have available and what quantity of data is a reasonable amount for them to work with.

The **time frame in which the data is recorded** may have an effect on the nature of the data collected**.** The time of year, the day of the week, or the time of day can all have an impact on the data. For example, a weekday may create a very different type of shopper in a town centre than one might find at the weekend. Equally, road traffic surveys may show a far busier state in the morning rush hours and during the afternoon school run than at other times of the day and the researcher may have to vary the timing of the surveys to make sure that a fairer representation of the issue is achieved. Physical geography studies into river systems and woodland, for example, will rarely be able to capture the nature of the landscape throughout the whole year, and providing the researcher is clear that their study is only representative of the landscape at that particular time of year, the data they collect at that time can be used to draw some (albeit limited) conclusions.

The exact **location** of the data collection may have a biased impact on the nature of the data. This may be especially true in physical geography studies when the point of data collection is highly dependent on the surrounding area. For example, a bend in a river or an especially steep river cliff may have an impact on the velocity of the river's flow at that point, but have little impact at another point. Taking a variety of readings and averaging them may be one way a researcher deals with this, or acknowledging to the reader, before the conclusion, that observations made in the field may affect the nature of the data. Similarly, how busy a town centre may appear is highly dependent on which part of town the researcher observes and selecting a variety of locations may be a better way to judge the level of activity in the area.

The **role of the researcher** in the data collection may unintentionally have an impact on the nature of the data itself. Data collected from methods such as surveys and observations may be very much based on subjectivity and the personal background of the researcher (where they live themselves, their own socio-economic background, or their prior knowledge about a place) may have an influence on how they score an area or how fair they are in comparing it to other places and locations. It is difficult to avoid this form of subconscious bias in these forms of data collection but with acknowledgement of the difficulties of being a researcher with their own sense of place, they can go some way towards being able to make more reliable conclusions from their findings. Equally, the researcher can ensure that if value scoring is to be undertaken at different locations (such as in surveys), the same collector of this data is responsible for this task in each location. While it may be more practical to have many people collecting the data at the same time, their different judgements of what constitutes a particular score in a bipolar analysis for example, may unfairly tip the locational results towards the standards of one data collector more than another.

Field sketches and photographs are also subjective as they rely on the researcher selecting **certain viewpoints** over others. For example, in an investigation into the amount of litter in a town centre, the photographer may be more likely to record a single scene that highlights the most severe form of fly tipping, rather than a scene that is actually more characteristic of the whole area, and less 'entertaining' to the reader of the study. Taking a photograph at each data collection location point, and maybe facing the same compass direction each time, may be one way of reducing the bias in this form of data collection

In a similar way, when collecting quantitative data using **field equipment**, researchers should be sure to use the same equipment each time they take a measurement and in the same manner too, reducing the chance that different pieces of field equipment, by having different margins of error in their design, will bias the final results. Ideally, the same researcher should use the same piece of field equipment each time too, reducing the chance of individualised user error.

Questionnaires and interviews can lead to biased results due to the **nature of the questions asked**, and those that are not asked. If members of the public are only asked to comment on certain ideas, then the conclusions that can be drawn from the questionnaires may present a picture where only those ideas are seen as important by members of the public. A broad range of questions, including many with scope for the interviewees and respondents to answer freely, can reduce the chance of bias in the results from these data collection methods.

Regardless of the precision of the data collection methods and the meticulous way a researcher may try to deal with data bias, it is important to remember that complete elimination of some form of unfairness or subjectivity in the data is highly unlikely. Acknowledging the limitations of the data collection methods used and the level of bias they are likely to create are more sensible ways to deal with the bias in the Independent Investigation.

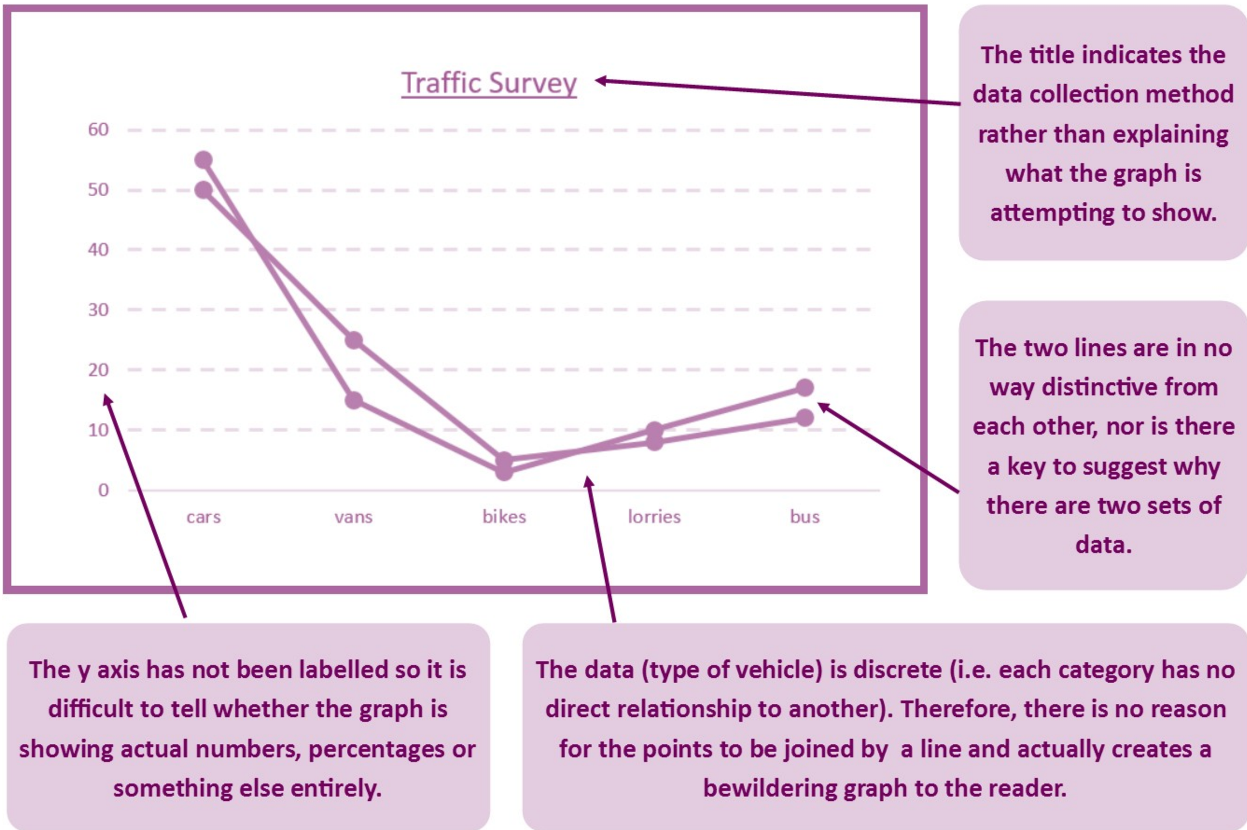
Section 3 –
Data Presentation

Once the researcher returns from their data collection in the field they are ready to address their **research questions** and hypotheses through a presentation of relevant data. Presenting data involves the use of a variety of different graphical techniques to visually show the reader the relationship between different data sets, to emphasise the nature of a particular aspect of the data or to geographically 'place' data appropriately on a map. The data presentation section is commonly written up at the same time, and alongside, the performance of **data analysis** (Section 4 of this guide). It is highly likely that it will make sense to analyse some data before it is presented graphically and vice versa.

Data presentation is not just used to make your Independent Investigation look more aesthetically pleasing – though good data presentation will also make the reading of the results more interesting to the reader. Instead, the primary reason for extracting the relevant data from your results and presenting it is to demonstrate to the reader and marker of your study that you can select the data most appropriate for answering your research questions and graphically work with the data to allow it to highlight its own inherent correlations and relationships. While a comprehensive data table that stretches for many pages may technically do the same thing, leaving the reader to try to 'find' the relevant data amongst a jumble of numbers is a sign of poor research practice. It is a good idea instead to structure your data presentation with your research questions in mind. Each one can be addressed in turn, with the appropriate data extracted and presented.

Readers and markers of your study will expect to see the use of a **variety** of data presentation techniques which are being used appropriately – not just for the sake of adding different ways of looking at the data. Putting every piece of data you have into a pie chart, histogram or scatter graph is not only boring for the reader and demonstrates a lack of imagination in the researcher and an inappropriate use of techniques, but also shows a lack of understanding of the benefits of using one data presentation technique over another. Bar charts for example may not be appropriate for showing a particular type of data. Simply typing the data into a spreadsheet package such as Microsoft Excel and selecting one of their generic data presentation techniques may illustrate that the researcher has not thought about the unique nature of the data and the research questions they are trying to answer.

A poor example of data presentation



It is important to ensure that the technique used to present the data is **appropriate** for the data in question. For example, if the researcher is trying to show the spatial relationship between a series of sites and correlating numerical data, they may wish to explore the use of GIS in their presentation. Equally if two sets of data are collected in such a way that the researcher is trying to show a correlation between them, it may be advisable to draw this as a scatter graph, so that a line of best fit may be drawn, allowing the researcher to comment on the strength of the correlation. If the data collected is able to be converted into a percentage of a whole, a bar chart may not be the best way to show the data. Instead, a pie chart or a percentage bar may more easily show the relative size of each of the categories compared to each other. It is important to consider the type of data you are dealing with: continuous and discrete data lend themselves more easily to some data presentation techniques than others.

More **sophisticated techniques** are likely to place your Independent Investigation in a higher marks level and the use of **original data presentation**, designed by you, is also encouraged. However, it is also important to remember that sophisticated techniques do not necessarily have to be complicated: the key to a good data presentation technique is that it is easily readable by the person marking your report. The use of GIS in your data presentation is also strongly encouraged if any type of spatial data is needed to answer the research questions.

Good data presentation can easily be let down by the inaccurate **execution** of the technique itself. Not giving a scale, or not labelling an axis can mean that the reader is not able to actually understand what the data is showing. Allowing someone not familiar with the data in question to ‘read’ the data presentation critically may help the researcher to avoid these type of errors before the report is submitted.

Common Pitfalls:

- **Presenting data that has nothing to do with a research question.** If the data does not have a role in the answering of the main aims of the study then it should be ignored, even if a lot of time was spent collecting that particular set of data.
- **Presenting only some of data.** If you think you will want to draw conclusions from it or refer to the data in some supportive way, it must be presented.
- **Using inappropriate and generic data presentation techniques.** Selecting a data presentation technique simply from a drop down menu in a computer-based spreadsheet shows a lack of imagination and there is a danger that a poor selection will result in demonstrating a lack of understanding of the complexity of the data in question.
- **Using a data presentation technique which is inappropriate for the data itself.** Think carefully about the type of data (continuous or discrete) and whether the technique you have chosen is appropriate for that type of data.
- **Using the same data presentation technique more than once.** Show some imagination and try to come up with an original data presentation technique, unique to your particular data.
- **Combining more than one data presentation together.** Remember, the ultimate aim is to make the data, patterns and relationships easy to see, not to create confusion for the reader.
- **Presenting the same piece of data more than once.** This can waste time and effort – choose the most appropriate technique only. If that technique does not show everything you want it to, it is not the right technique to use.
- **Using techniques in an inaccurate fashion.** Spend time checking that you have labelled keys and axes appropriately – marks can be easily lost without these checks.

3a – A Guide to
Column Charts and Histograms

Royal Geographical Society with IBG
Advancing geography and geographical learning

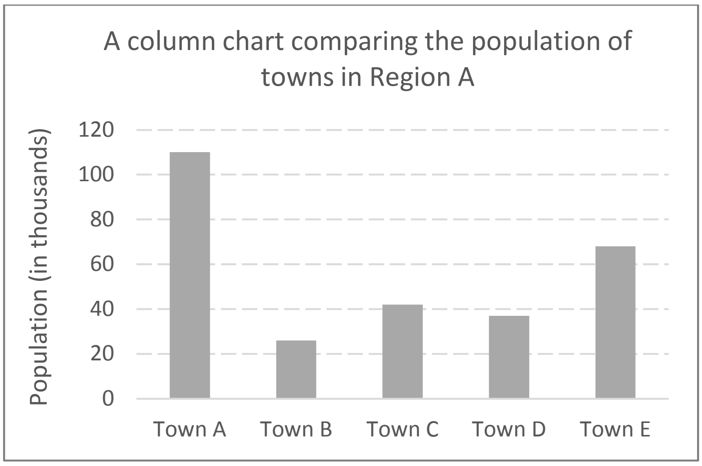
Column and Bar charts are a common form of data presentation, used to show the frequency of something with a data set. Column and bar charts show discrete data, where each variable is unrelated to the next. They are commonly confused with **Histograms** which instead work with continuous data, and show how the data is distributed across a range of values such as time or distance.

Why would we use column, bar charts and histograms?

This form of data presentation is commonly used for its ease of design and interpretation. A researcher would aim to use this method of data presentation when faced with relatively few sets of data (four to eight different variables for example) that are distinguishable from each other by their respective quantities.

Column Charts

Column charts use coloured columns standing vertically to show the frequency of data within discrete categories. To emphasise that the data is discrete, the columns should be separated by a gap along their x axis.

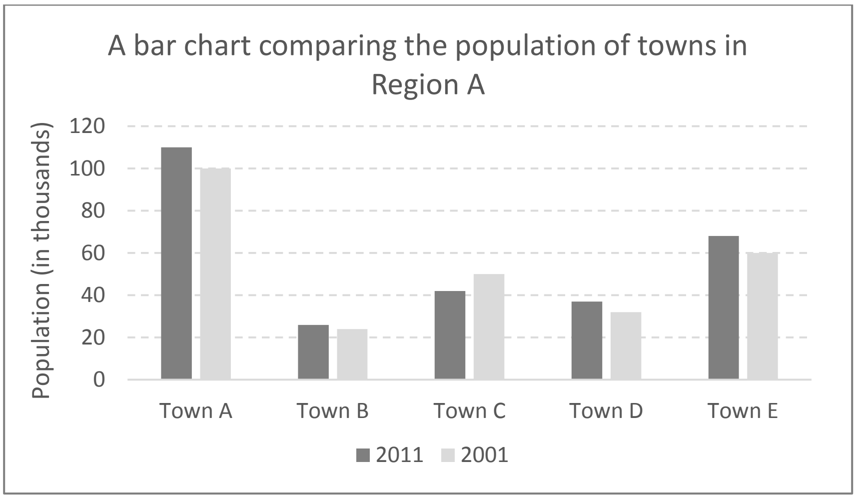


Bar Charts

Bar charts represent data in exactly the same manner as column charts but with the bars running horizontally. The bars should be separated by gaps along their y axis to show the discrete nature of the data.

Comparative Bar Charts

Placing another set of data alongside the first can allow a researcher to more easily make comparisons between them. A key may be needed to distinguish between the two sets.

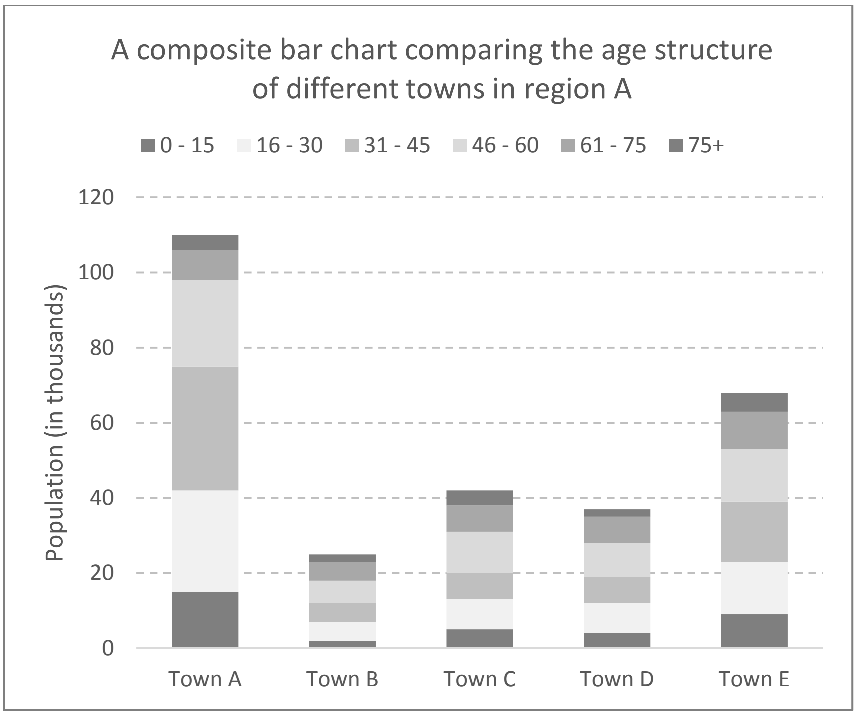


Composite Bar Charts

A composite bar or column chart can be used if the researcher wishes to show more than one type of data for each variable. These can be expressed as a percentage of the whole or as the actual figures within that total.

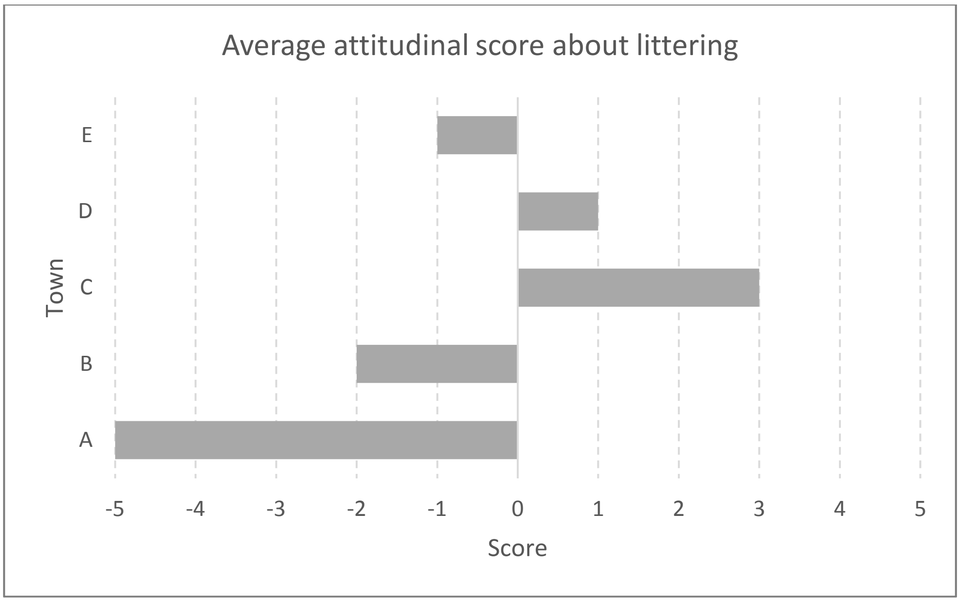
Why would we use composite bar charts?

A researcher would use a composite bar or column chart when there are a reasonable number of categories in each bar. More than six categories can make reading the chart quite difficult.



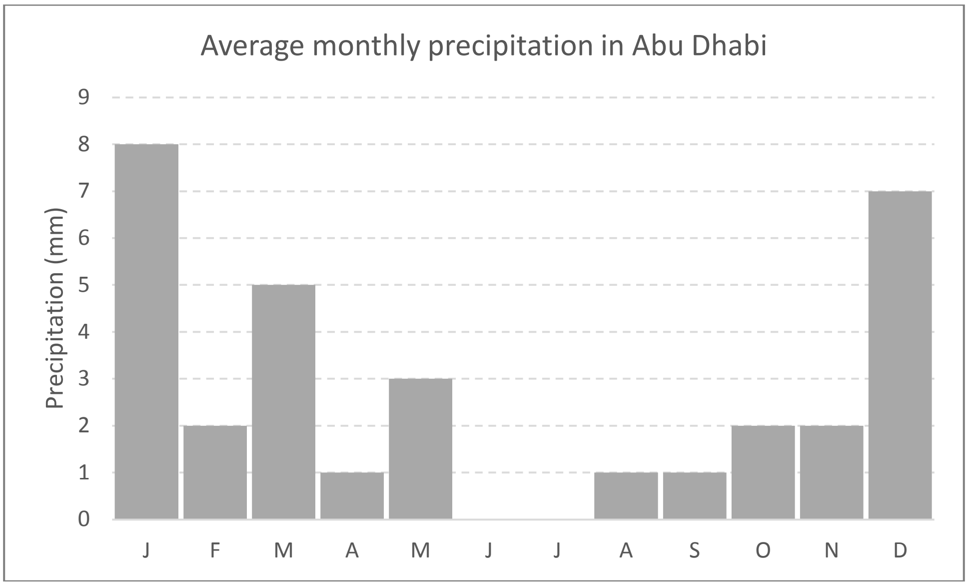
Bipolar Bar Charts

These bar charts have one y axis and an extended x axis, allowing the researcher to show two forms of data at once. The most common bipolar bar chart that geographers use is actually a histogram: the age-sex pyramid. However, if the researcher wishes to display data on opinions from a bipolar survey, they can show both the positive and negative opinions, visually on the one bar chart.



Histograms

Histograms are bar charts that are showing continuous data. This means that it is possible for some zero values to also be displayed alongside the positive values. One histogram commonly used by geographers is a rainfall graph. Unlike other types of graphs, the user should not be tempted to shade the columns differently as they are all representing the same type of data.



3b – A Guide to Pictograms

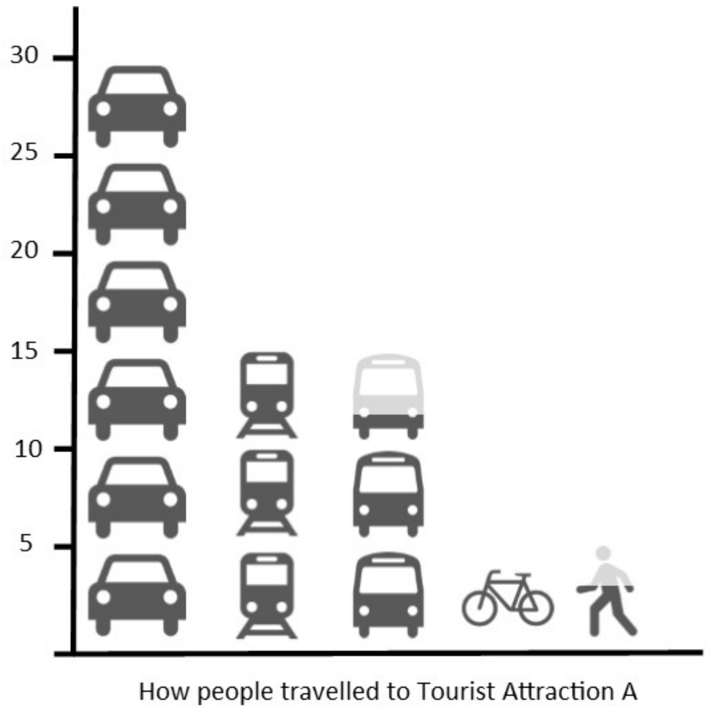
A pictogram involves the use of a symbol in place of a word or statistic.

Why would we use a pictogram?

Pictograms can be very useful when trying to interpret data. The use of pictures allows the reader to easily see the frequency of a geographical phenomenon without having to always read labels and annotations. They are best used when the aesthetic qualities of the data presentation are more important than the ability to read the data accurately.

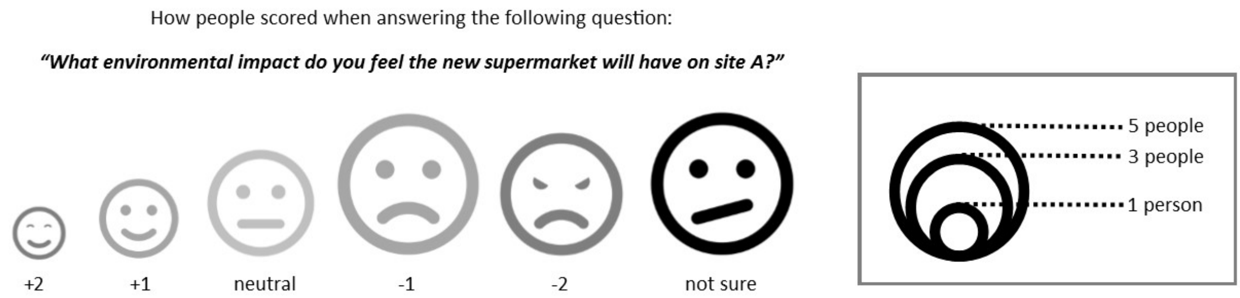
Pictogram bar charts

A normal bar chart can be made using a set of pictures to make up the required bar height. These pictures should be related to the data in question and in some cases it may not be necessary to provide a key or explanation as the pictures themselves will demonstrate the nature of the data inherently. A key may be needed if large numbers are being displayed – this may also mean that 'half' sized symbols may need to be used too.

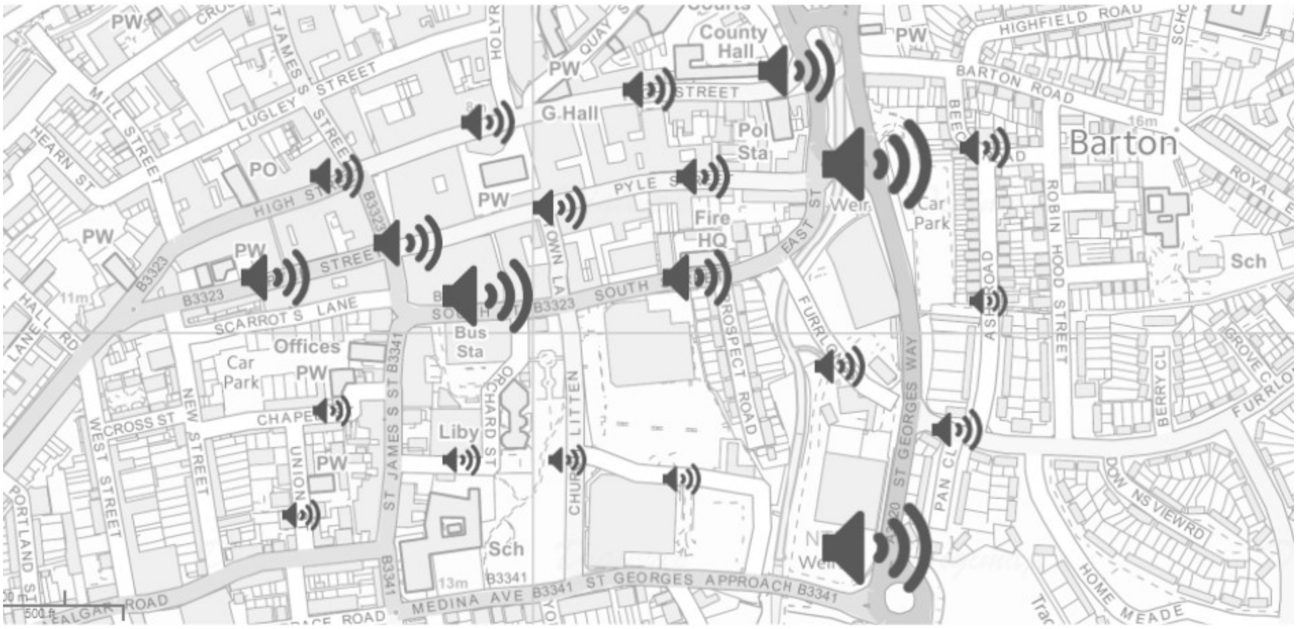


Proportional shapes and symbols

Scaling the size of the picture to represent the amount or frequency of something within a data set can be an effective way of visually representing data. The symbol should be representative of the data in question, or if the data does not lend itself to a particular symbol, a simple shape like a circle or square can be equally effective.



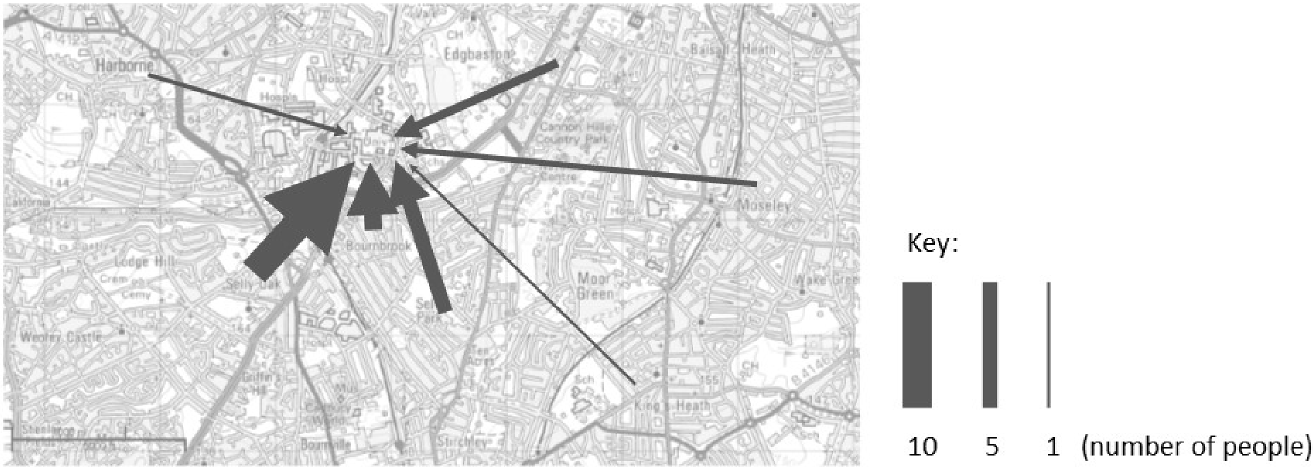
Proportional symbols can work well with GIS, where the symbols can be placed on different sites on the map to show a geospatial connection to the data. Providing a key with a scale to the size of the symbols may not be necessary if the relative values of the shapes is the main point of the data presentation. If precise data needs to be read from the proportional symbol, then it may not be the best data presentation method to use in the first place.



A map to show relative noise levels across Newport town centre

Flow lines

Flow lines (sometimes called desire lines) are arrows drawn on a map which show a connection between two places. The size of the arrow (its width) is proportional to the frequency of that connection. For example, a wider arrow may show that a large number of people travel from Town B to Town A, while a narrower arrow will show that fewer travel from Town C.

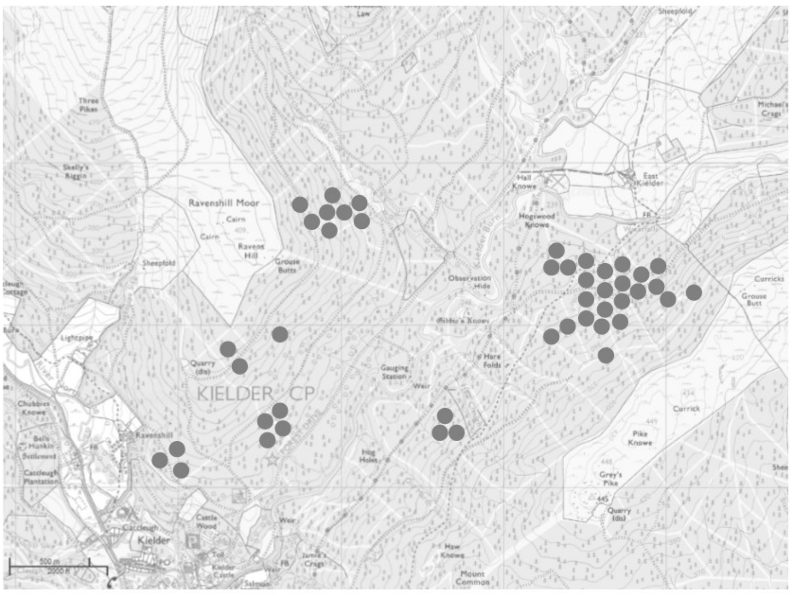


Where the questionnaire respondents had travelled from to visit the university that day

As with proportional shapes, a scale bar may not be necessary if only the relative differences between the arrows needs to be displayed. Flow lines can also be used with maps to produce effective GIS based data presentation.

Dot maps

Dot maps are frequency distribution maps where the amount of a variable is shown by the number of dots there are in a particular location. A single dot may represent a single occurrence of a phenomenon or a dot may represent a set number: the amount should be indicated in the map's key.



● = 5 occurrences of Species A observed

Dot maps work best when they are representing a large spatial scale: trying to create a dot map for a small area, like a single street in a town may simply result in a lot of dots singularly spaced and a feel for the distribution of a variable will not be easy to read. With more data samples and a wider area under observation, a pattern may be more identifiable.

3c – A Guide to Pie Charts

Royal Geographical Society with IBG

Advancing geography and geographical learning

Pie charts are commonly used by researchers to show how a whole set of data is split between different categories. They are normally used to show the relative frequency of data rather than be a presentation of raw data, as generally they are displayed without the data actually listed alongside each segment.

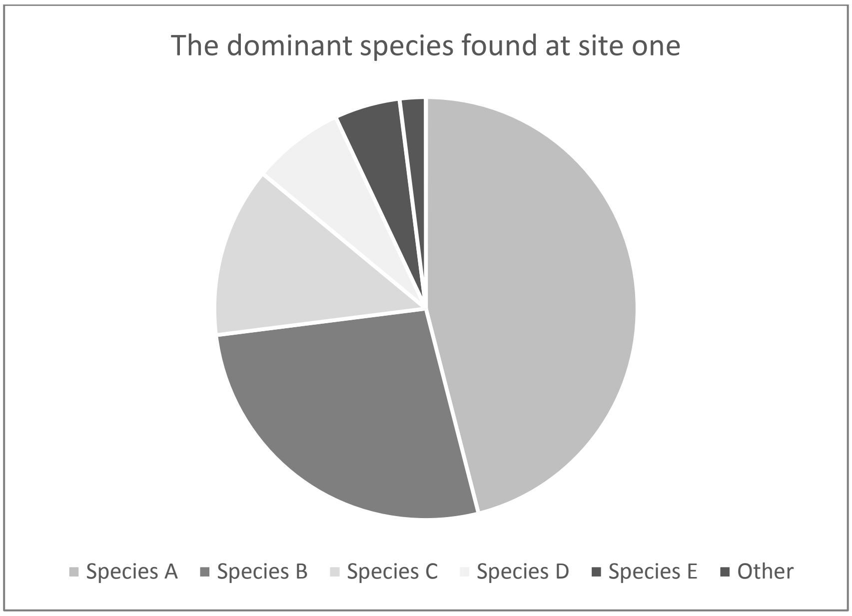
Why would we use a pie chart?

Pie charts are best deployed by researchers trying to show the amount of a phenomenon in a relatively small number of categories. Through the use of different colours, pie charts can allow the reader to quickly see the majority and minority categories without having to read exact numbers in the data set.

To calculate the size of each part of the 'pie', the following equation should be used:

$$\frac{\text{value}}{\text{total value}} \times 360 = n^\circ$$

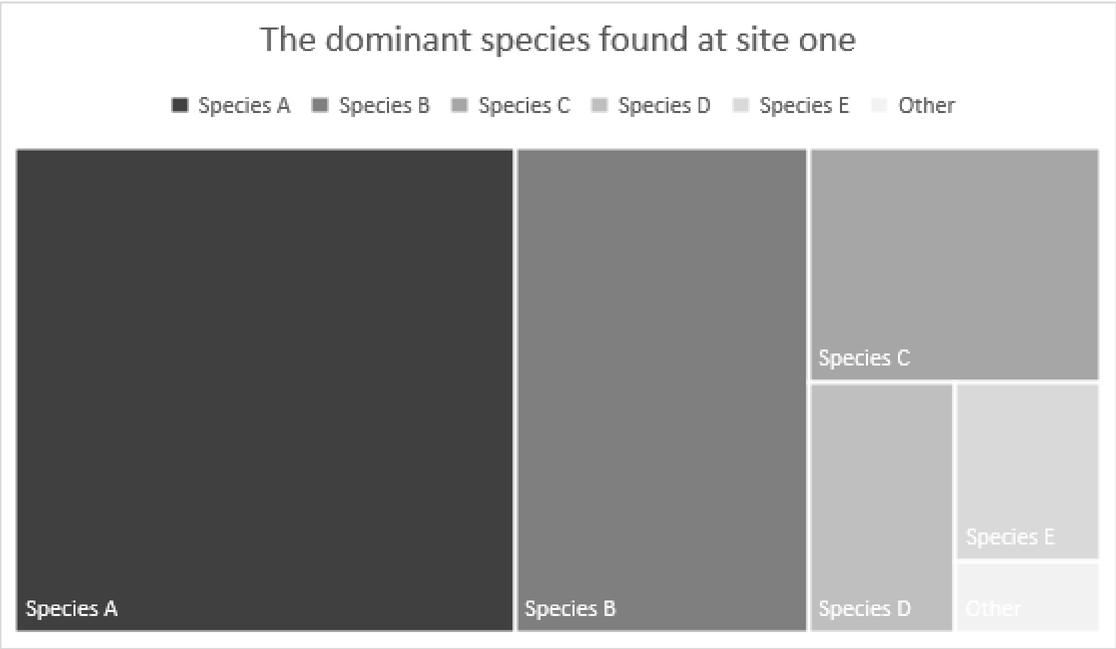
Pie charts can be difficult to read if there are too many categories with very low values – this creates a chart where many categories are compressed into one part of the pie. Depending on the nature of the data, the researcher may be able to create an 'Other' category which encompasses all these small values. However, this method should be used with caution as creating an 'Other' category also effectively hides some of the data that might actually be significant to the conclusions the researcher wishes to make.



Pie charts can also be used with GIS. If the data changes from place to place, the pie charts can be placed as an overlay on a map of all the data collection sites, allowing the reader to compare one site's data, relatively, to another.

Tree Maps

Tree maps are the same as pie charts in all ways except their shape. Instead of a circle, a rectangle or square is used and is split according to the values of the different categories of data.



Generally, tree maps are set out with the highest to lowest values reading from top left to bottom right.

3d – A Guide to Scatter and Line Graphs

Royal Geographical Society with IBG

Advancing geography and geographical learning

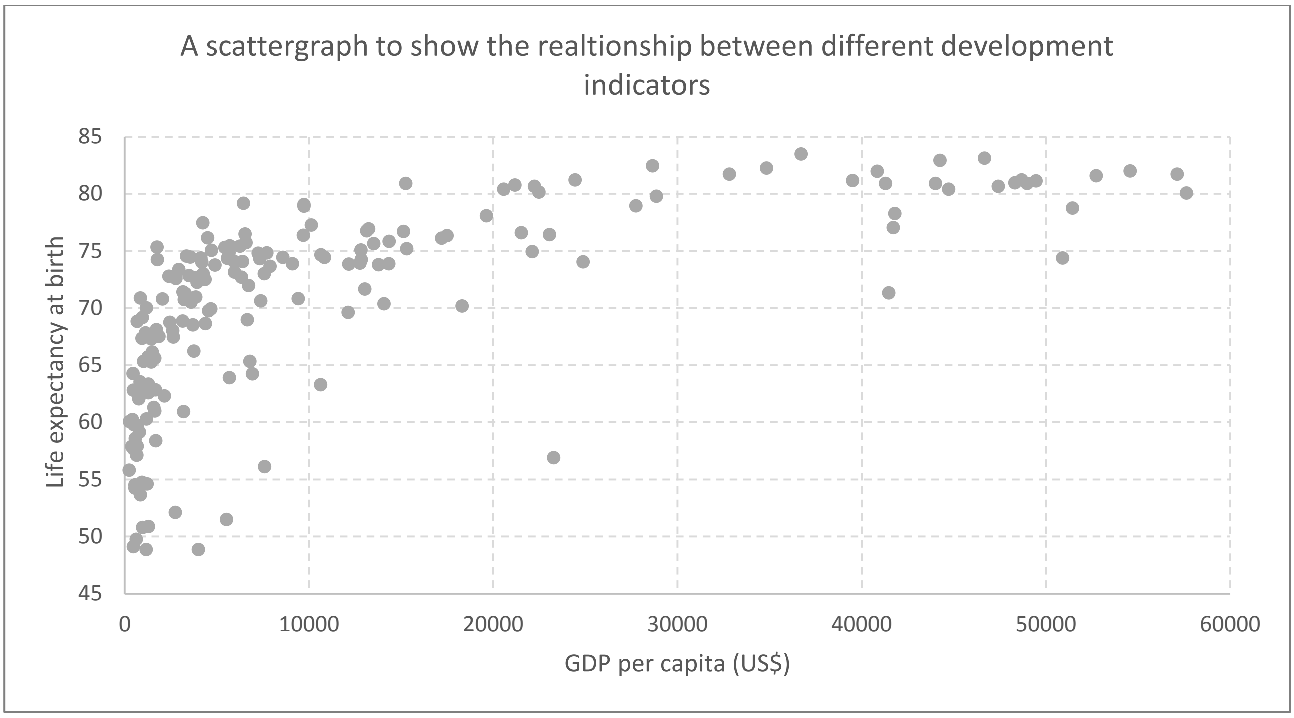
Scatter graphs and line graphs are used to show the potential correlation between two different variables.

Scatter graphs

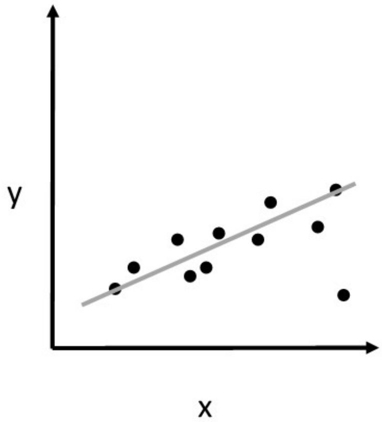
Scatter graphs can be used when the data from both variables under investigation is continuous. In order to be able to judge whether there is a correlation between the two variables, the researcher should try to include as many points as possible: fewer than thirty plots can make it difficult to draw any meaningful conclusions from the data.

Why would we use a scatter graph?

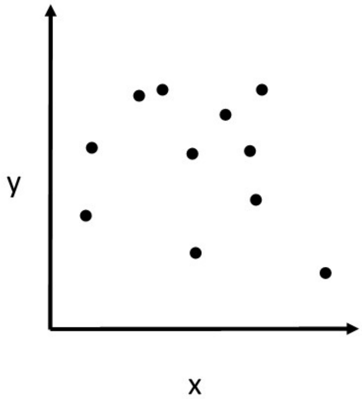
Researchers tend to use scatter graphs when they are faced with very large sets of data. Scatter graphs allow researchers to identify anomalies in the data more easily as well as the overall trend and relationship between the variables. The strength of any relationship can also be visualised.



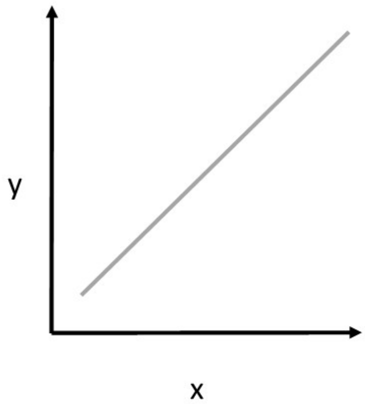
A line of best fit can also be drawn on the graph if the researcher suspects that there is a correlation. This line does not necessarily have to go through any points at all, but should fall in the middle of the plots and aim to have an equal number of points above and below the line. An outlying point may be interpreted as an anomaly. The angle and direction of the line of best fit can tell the researcher what type of correlation the data has.



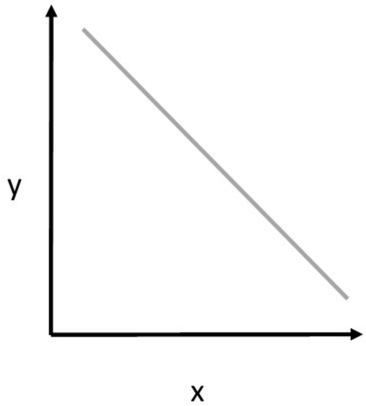
a)



b)



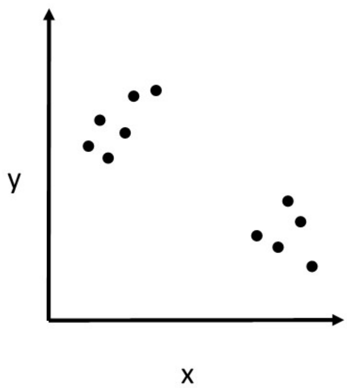
c)



d)

a) Drawing a line of best fit. The outlying point may be thought of as an anomaly.
b) No correlation. It is not possible to draw a line of best fit by eye.
c) Strong positive correlation
d) Strong negative correlation

The nature of the data used in a scatter graph may mean that a line of best fit is not possible. Instead, clusters of data may be identifiable and rings may be drawn on the graph to show how one idea is true for one scale of data and another is true at the other end of the scale.

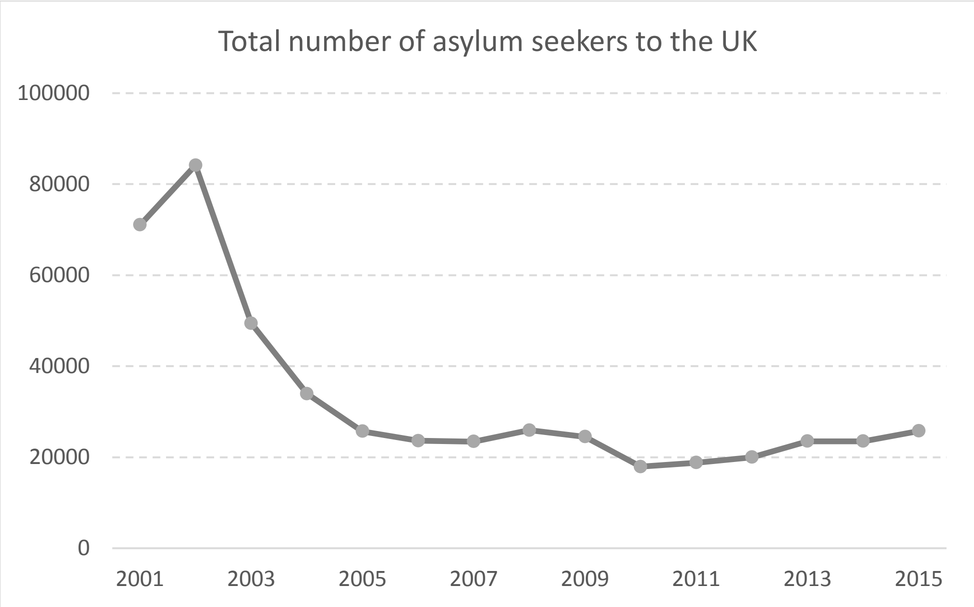


Line graphs

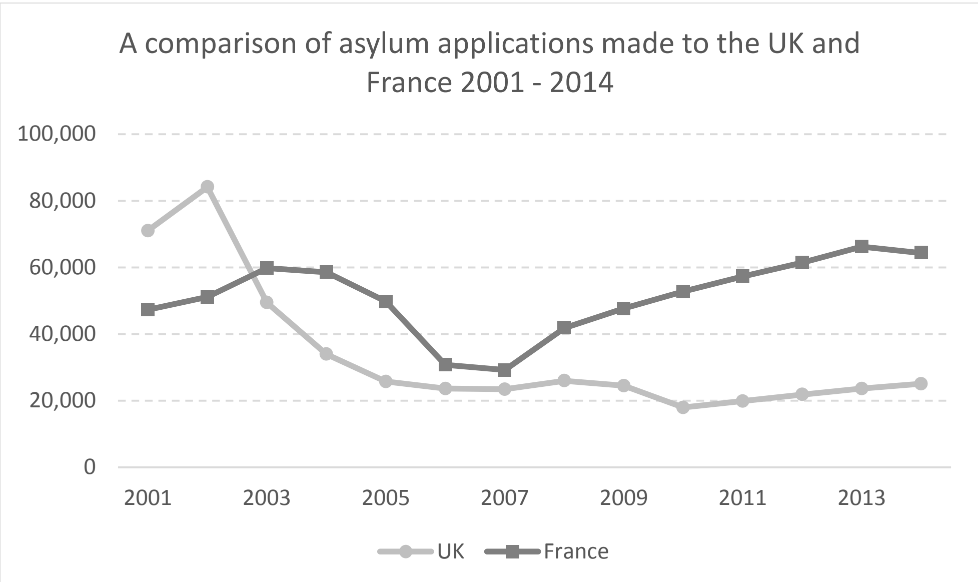
Line graphs are used to show the change in a relationship between two variables over a form of continuous data such as distance, or most often, time. Points are plotted in the same way as one would for a scatter graph but then joined together in sequence.

Why would we use a line graph?

Researchers would use a line graph to show the nature of a relationship, not whether there is a relationship or not between two sets of data. They can be used to instantly see the 'journey' made by a geographical entity over time or distance which can be useful for creating a narrative around a geographical idea.

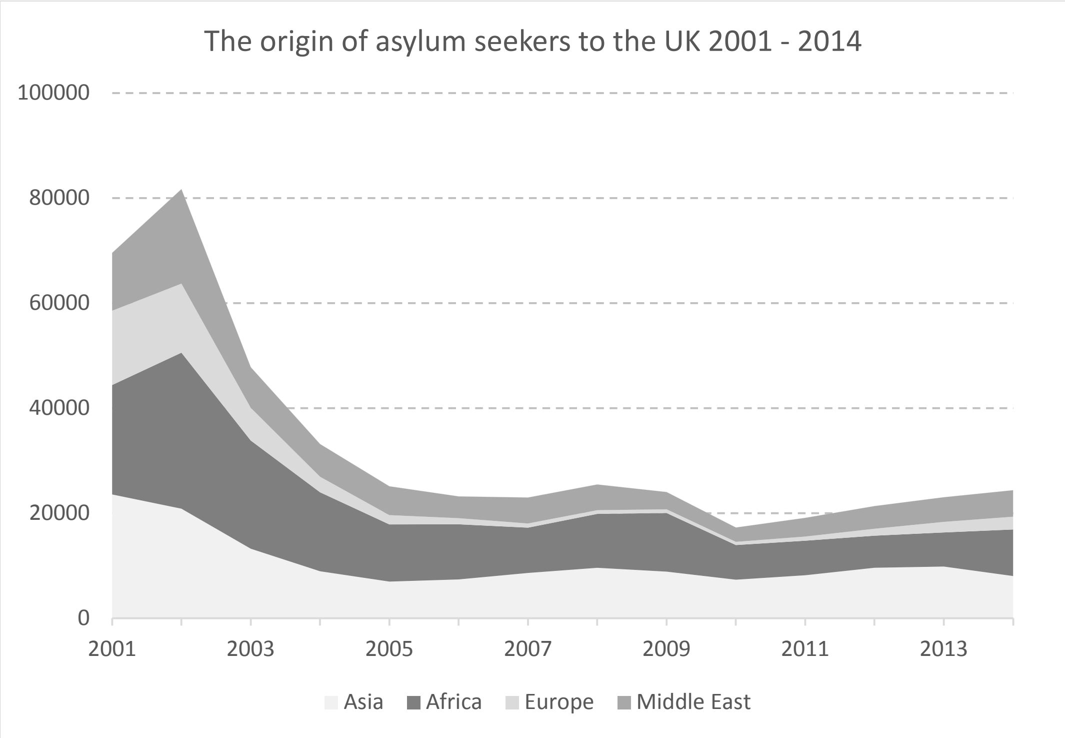


More than one line can appear on a graph, along with a key, to show different sets of data.



Compound line graphs

For some data there may be a total value, and further component values, that can be shown on the same graph. For example, the total number of asylum seeker applications a country receives may change over time but the places from which those asylum seekers come will vary in different years.



Using Logarithmic Scales

Sometimes the spread of the data is so large between the minimum and maximum values that producing a graph with a meaningful scale on its axes can be challenging. In these cases, a logarithmic scale might be helpful. A logarithmic scale is one where values along an axis are compressed to make the whole spread of data fit into one graph (known as a log graph). Usually this compression occurs by powers of ten, and instead of the axis showing units going up one at a time in a linear fashion (0, 1, 2, 3, 4 etc), the units go up in non-equal parts (1, 10, 100, 1000, 10000 etc). If both sets of data have a large spread, both sets of axes can be converted into a log scale (and the resulting graph is known as a log-log plot).

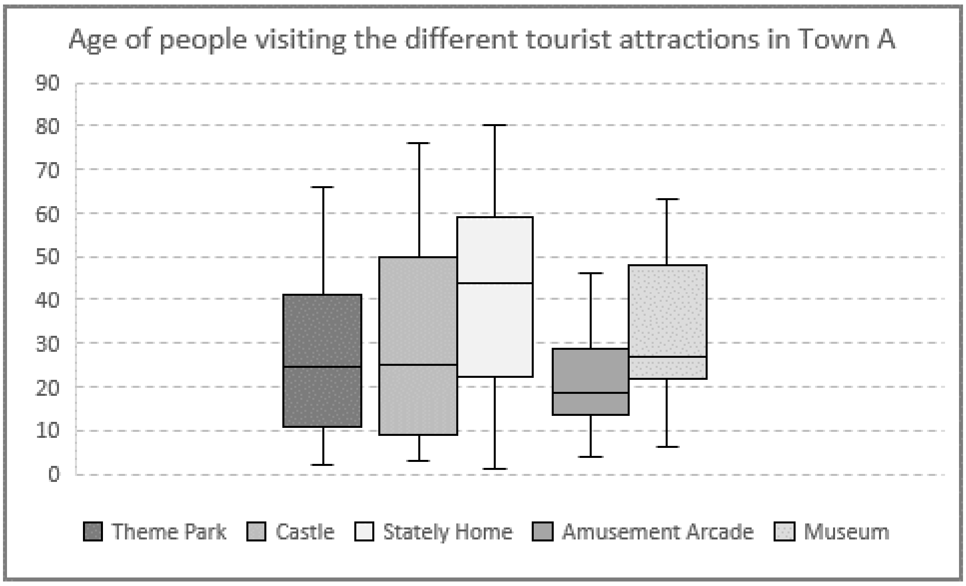
Drawing a graph that uses a logarithmic scale can be done using online downloads and packages designed to work with very large data sets. If you wish to draw a logarithmic graph by hand, log graph paper can be found online and printed off. Plotting the points on a logarithmic scale is done in exactly the same way as when using a normal scaled axis: the researcher needs to simply take a little more time as it can be easy to misplace a point using these scales.

3e – A Guide to
Box and Whisker Graphs

Box and whisker graphs are used to show the spread of a data set across a range of values. The bars on the chart appear to 'float' with their extent covering the most frequently quoted values in the data set. The top of the box is placed on the lower (or first) quartile value, while the bottom of the box extends to the upper (or third) quartile. The x axis generally shows the different categories of data (such as different locations).

Why would we use a box and whisker graph?

A researcher would use a box and whisker graph rather than any other method when they need to compare distributions of a number of variables quite directly with each other. The many elements to a box and whisker graph means that one graph can show a number of different data points.



Worked Example:

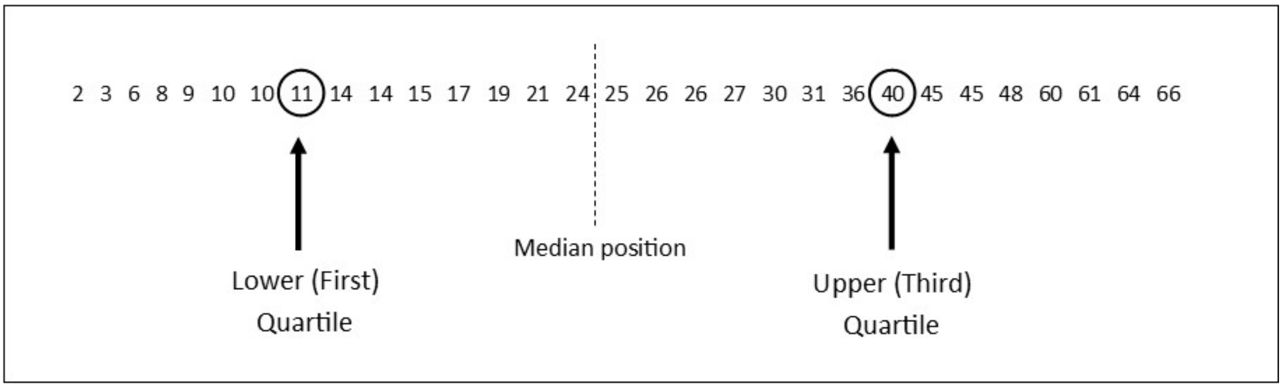
There are a number of ways of calculating the first and third quartile values. The following method represents a simple and commonly used method, but it is by no means the only method and spreadsheet packages that can calculate the lower and upper quartiles automatically may use a different method.

Firstly, the data should be placed in value order from lowest to highest.

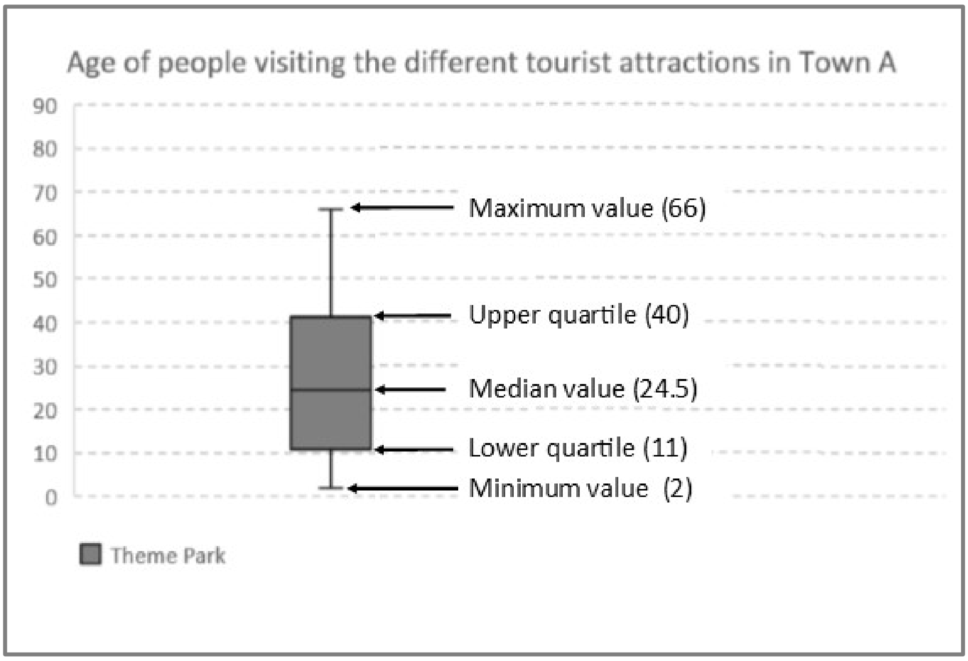
Theme Park:

2 3 6 8 9 10 10 11 14 14 15 17 19 21 24 25 26 26 27 30 31 36 40 45 45 48 60 61 64 66

By dividing the total number of values by two, you can find the median value (that that falls exactly in the centre of the range of data). If the total number of values is odd, the median will represent an actual value within the set. If the total number of values is even, (as in the above example), the median represents a position between two values.
The lower quartile sits between the median point and the minimum value, representing twenty-five percent of the total number of values. Similarly, the upper quartile is the value that sits between the median point and the maximum value, representing seventy-five percent of the total number of values.



At the top and bottom of each box on the chart, ‘whiskers’ can be added. These are lines that extend out to the minimum and maximum values within each data set. Additionally, a line can be drawn in the centre of the box to show the median value for each set.



3f – A Guide to
Kite Diagrams

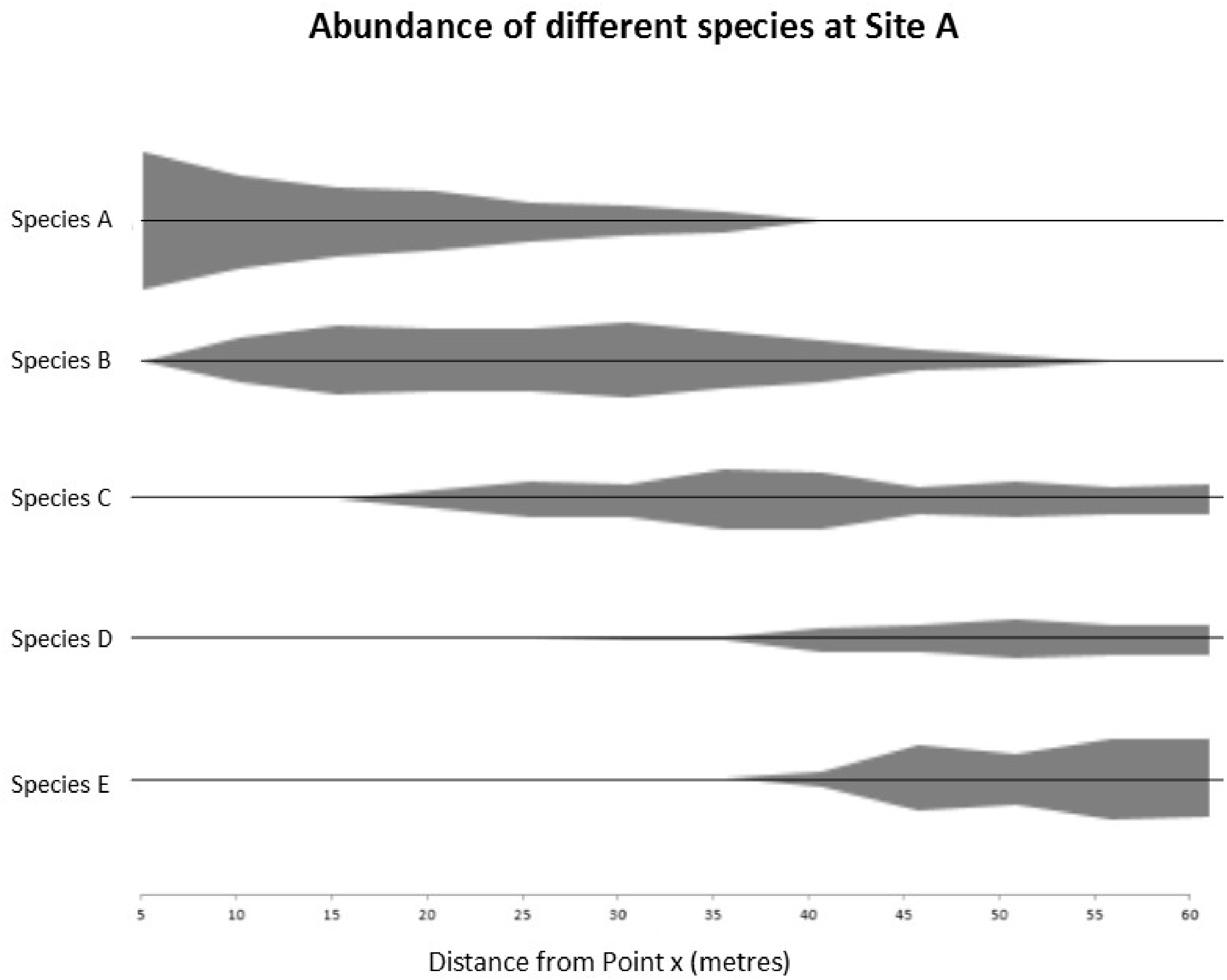
Royal Geographical Society with IBG
Advancing geography and geographical learning

Kite diagrams are used to display several observations seen at various points along a transect.

Why would we use a kite diagram?

Kite diagrams allow researchers to graphically compare the different frequencies of different observed data along the same transect. As the frequency is shown by an area on the graph, potential and hypothesised relationships between different observations can be more easily identified.

Their most common, though not exclusive, use is for showing the frequency values of different species on a surveyed transect in a certain environment.



The kite diagram is actually a more complex series of bar charts. Along a single x axis (of which there are as many as there are species to represent) the data value is split, with equal amounts on each side of the axis, creating a ‘kite’ shape. In some cases, the researcher may wish to label the

y axis with values so that the actual data can be read off the diagram. In general, though, kite diagrams are used to show relative abundance. The area of the kite shape for each observed phenomenon then represents the abundance of that particular phenomenon.

If a large range of values is being presented, it may be more efficient to calculate the values as percentages of the total and use these rather than the raw data.

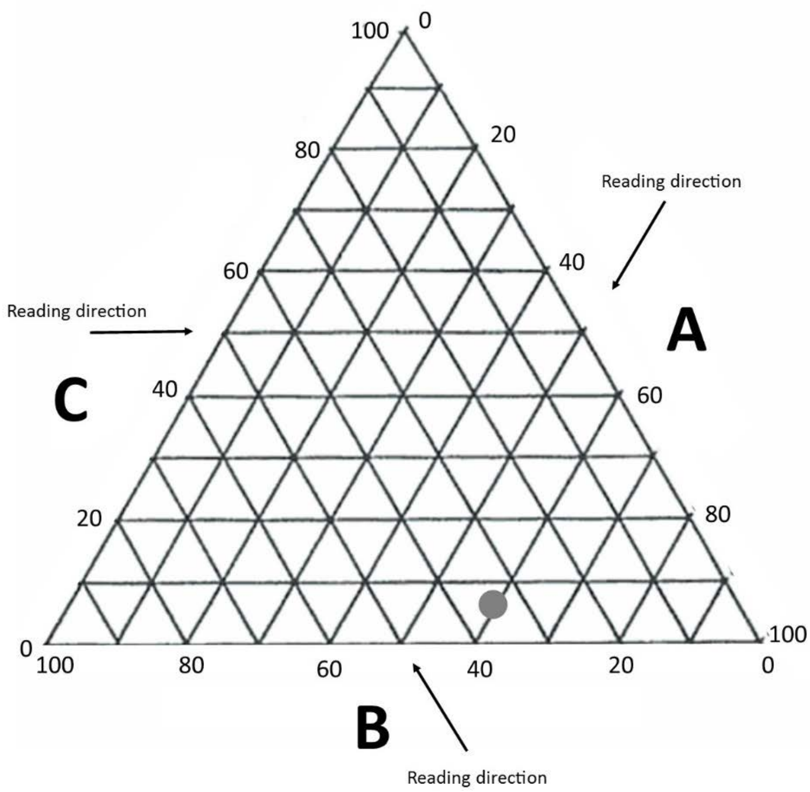
Kite diagrams can be drawn by hand or using online packages, however, few spreadsheet software packages have kite diagrams as a standard graph choice in their system.

3g – A Guide to Triangular Graphs

Triangular graphs (sometimes known as ternary graphs) offer an opportunity to display data based on three variables simultaneously. They can only be used for three variables where their total equals one hundred percent of the data.

Why would we use a triangular graph?

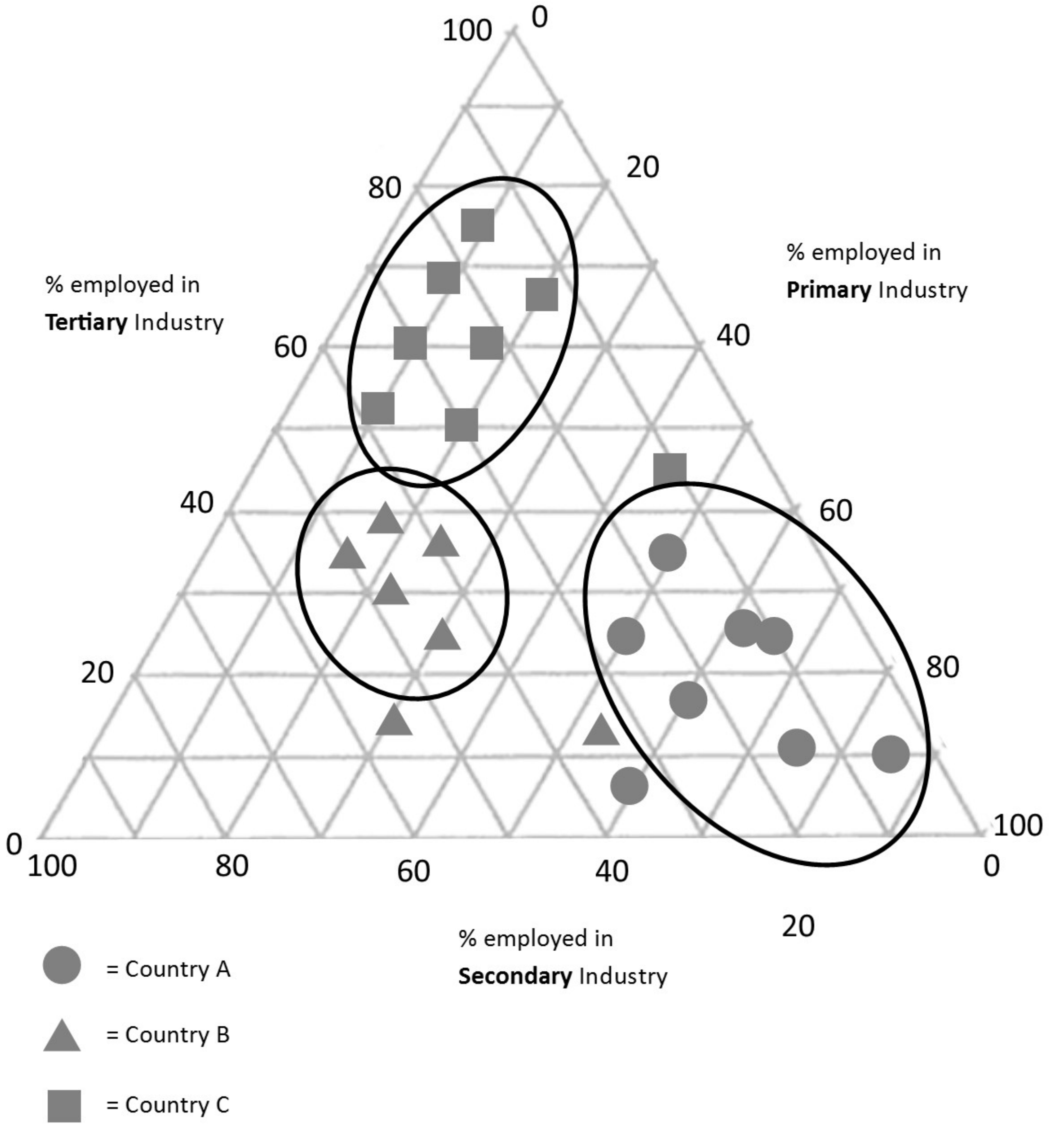
Though not all data categories neatly fall into three and only three sub-categories, for those that do, triangular graphs offer a spatial method of seeing the relative abundance and position of such data. They are easy for the researcher to read and create; they can utilise colour to show further subdivisions in the data.



In the above example, the point shows:

A = 59%   B = 34%   C = 7%

A triangular graph is an equilateral triangle with three ‘axes’. Points can be plotted by reading the correct value off each of the axes and placing a point on the triangular grid. With many points, patterns may be observed and clusters of points may indicate a relationship between a place and a corresponding phenomenon. In this case, placing cluster circles on the triangular graph can be a useful way of highlighting these to the reader.



Triangular graph templates can be downloaded from online sources and hand completed or, if downloaded into a publishing software package, completed digitally.

3h – A Guide to
Rose and Radial Graphs

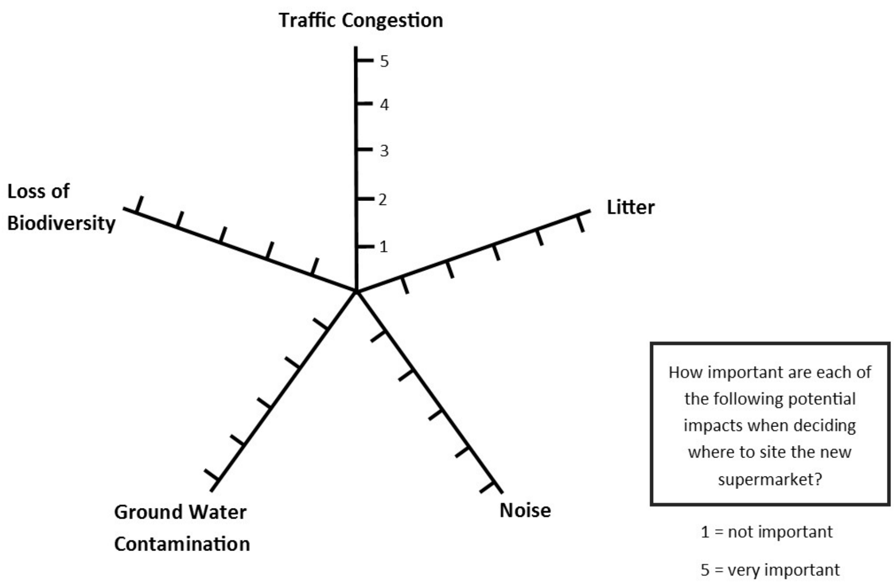
Royal Geographical Society with IBG

Advancing geography and geographical learning

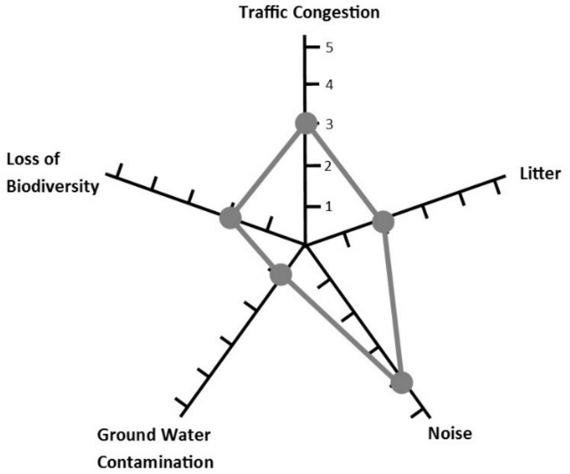
Radial Graphs are multi-axis graphs that show a number of similar ideas in one graphical presentation. They are so called because the axes, of which there are usually five or six, radiate out from a central point.

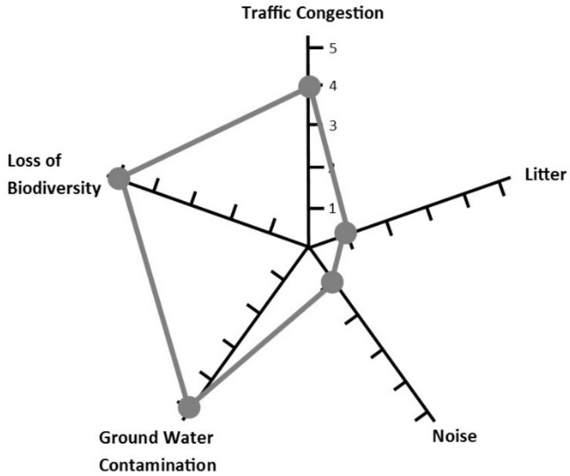
Why would we use a radial graph?

Researchers tend to use radial graphs not only to allow them to present five or more sets of data on one diagram at once, but also because it allows them to compare the relative strengths of the different variables at the same time as the overall strength of the location compared to others.

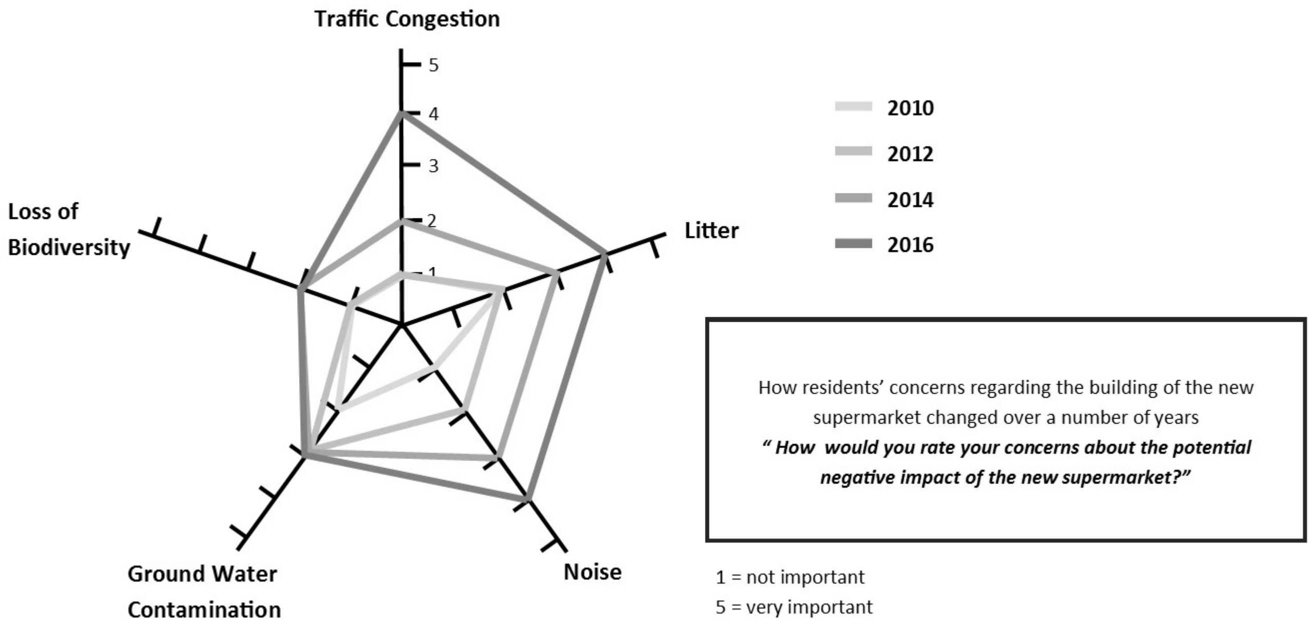


Once plotted, the points of a radial graph can be joined together to form a shape. This shape can be compared with others that have been plotted on the same axes and this comparison may allow the reader to identify patterns across a range of different sites. A shape with a clear point may indicate that one variable dominates in one field, while a more circular shape will show that all the variables may hold roughly equal weights.

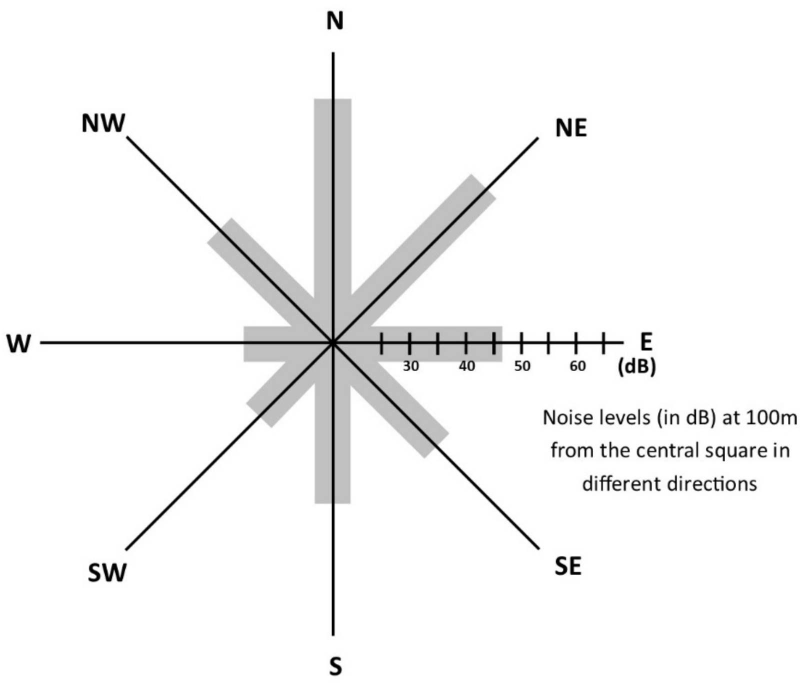




Equally the different shapes can be drawn in different colours over one another on the same set of axes, allowing a comparison to be made more easily between different sites, different groups of people, or different years.



One form of a radial graph is a **Rose Graph**. This is a circular histogram, with multiple axes radiating out from the centre. Instead of points being plotted along an axis, bars are drawn to show the quantity of that variable. A common use of a rose graph is in showing wind speed or noise levels. The 'axes' for the graph are the directions of a compass and the speed of the wind or the decibel level is then drawn in the appropriate direction.

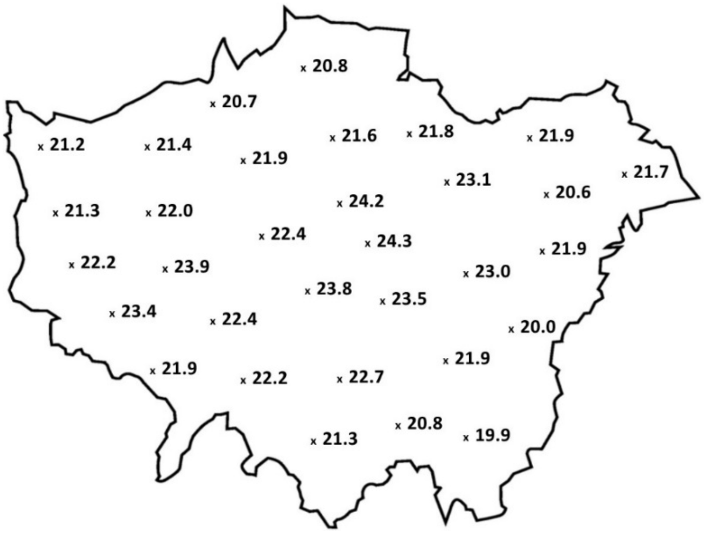


3i – A Guide to Isoline Maps

An **Isoline Map** is a way of presenting numerical data cartographically.

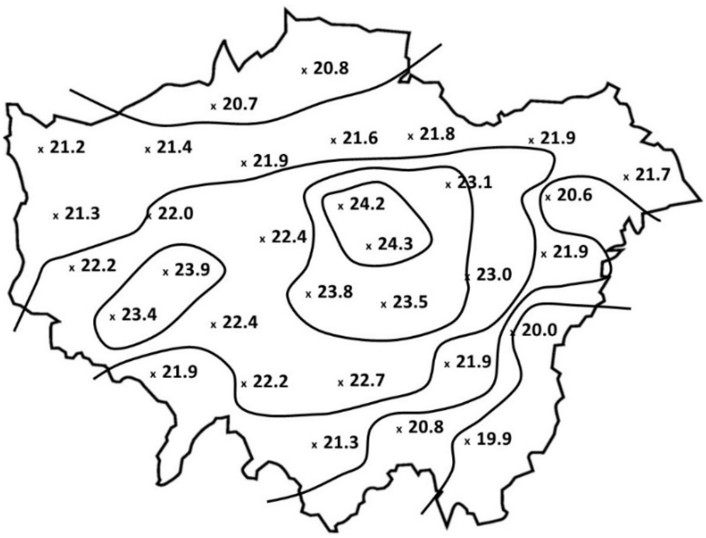
Why would we use an isoline map?

Isoline maps help the reader to recognise patterns and relationships between the geography of an area and data that might have been collected on the ground, such as air temperature.

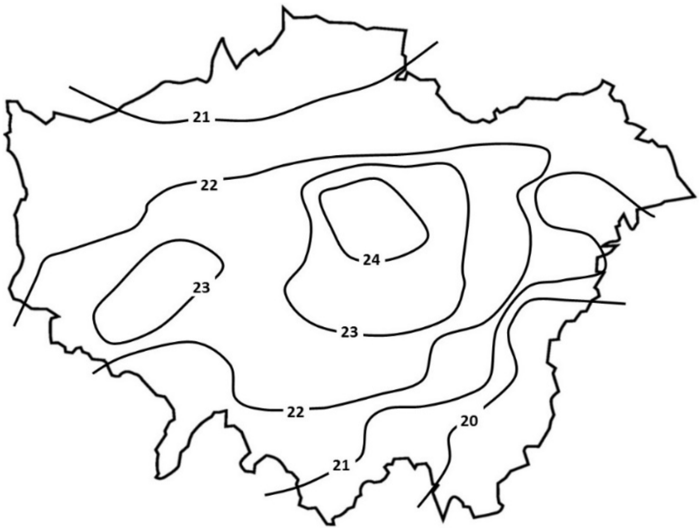


An outline map of Greater London showing the maximum temperature (in °C) recorded on a given day in the summer.

Isolines are lines drawn on a map connecting data points of the same value. They are commonly used by geographers. Contour lines, for example, show relief and connect points on the map that have the same height. Equally, isobars show bands of high and low pressure and connect points that have the same atmospheric pressure.



The same map with appropriate isolines used (at 1°C intervals).



This map can be further simplified by removing the actual data points and only showing the isolines with their labels.

Isolines should have equal intervals between them numerically. The scale used (for example, whether the value goes up in tens or hundreds) depends on the nature of the data being used and at which scale the map will tell the reader the most information. As equal intervals are used, it is unlikely that an isoline will actually pass through every point that has been plotted, instead passing either side of the point depending on whether the value of the isoline itself represents a higher or lower value than the data point.

If the areas between the isolines are shaded in a choropleth fashion, the graphic is known as an **Isopleth Map**.

3j – A Guide to
Choropleth Maps

Royal Geographical Society with IBG

Advancing geography and geographical learning

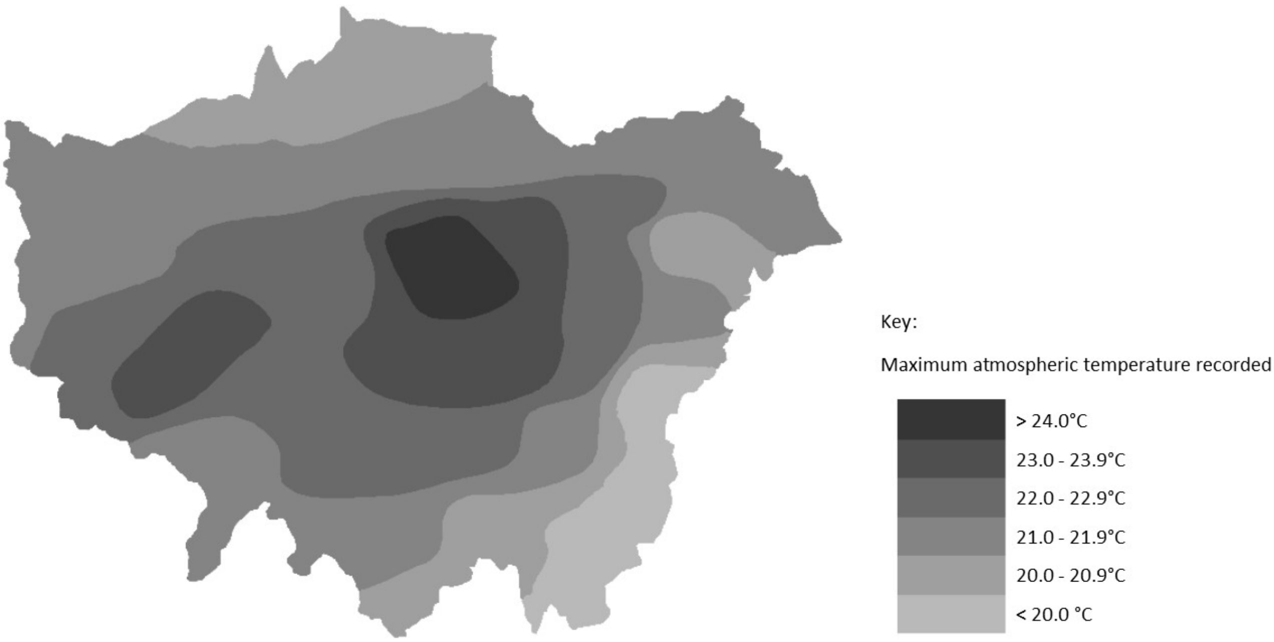
Choropleth maps are shaded maps where the intensity of the colour is indicative of the intensity of the phenomenon in question. A good choropleth map is one where the reader does not have to consult the key in order to understand the pattern or relationship that the map is conveying.

Why would we use choropleth shading?

A researcher would use choropleth mapping when data can be arranged spatially and into different categories. It allows the reader to easily see patterns in the way the data is spatially arranged.

The best type of choropleth map uses colours that are symbolic of the variable being displayed. For example, a map which shows the maximum rainfall data across a wide area may be coloured blue in varying intensity of shades. The use of different intensities of one chosen colour is what makes choropleth maps so easy to understand. The more there is of a phenomenon, or the greater the frequency of a certain observation, the darker, or bolder, the shade of that colour. The less there is of that phenomenon, the lighter or weaker that same colour will appear. If the map is going to be presented in black and white copy then a simple grey scale, such as the one below, should be used.

In some circumstances, it may be easier for the map reader if different colours are used alongside each other. In these cases, a colour spectrum (which still uses symbolic colours) is best used to allow the reader to instantly identify the pattern or relationship. For example, on a map that depicts temperatures over an area, one may use ‘cool’ colours, such as blues and greens, for the lower temperatures and ‘warmer’ colours, such as oranges and reds, for the higher temperatures.



A choropleth map showing the maximum recorded temperatures in London on a given day.

It is generally accepted that choropleth maps should be constructed from no more than seven different shades, though five is the number normally used. If the researcher uses too many different shades it can be difficult for the reader to differentiate between them.

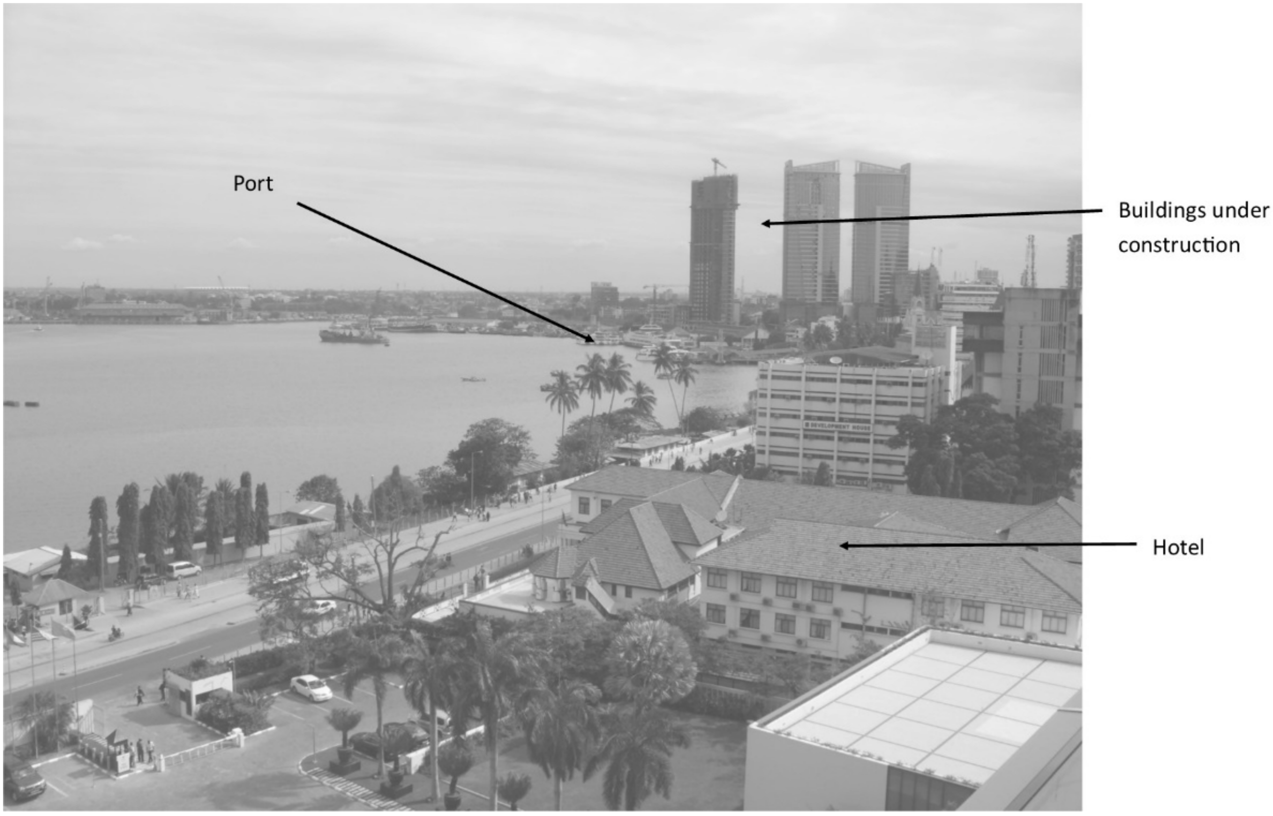
In order to draw a choropleth map, it may be necessary to first draw a map with the appropriate isolines for the data. This is where lines separate the borders of one colours from another.

Choropleth shading usually relies on the researcher creating data bands for each colour shade (i.e. a range of numbers from one value to another that colour shade represents) which should then be shown through a key. These bands do not have to be equal in size and the decision over the size of the bands may rest on the size of the spread of the data. Therefore, the nature of choropleth maps mean that it is not always easy to directly read the actual data for a particular location: the reader may only be able to identify the data range within which that location falls.

3k – A Guide to
Pictorial Data Presentation

Sketches, **Cross Sections** and **Photographs** form an integral part of most data collection and their presentation in a report write-up can add context to quantitative data. The use of annotations can draw the reader's attention to things they might otherwise not see and add additional information such as background data or historical issues associated with the scene.

A label is used simply to highlight something in a sketch or photograph, while an annotation is likely to explain something in the scene in greater depth.



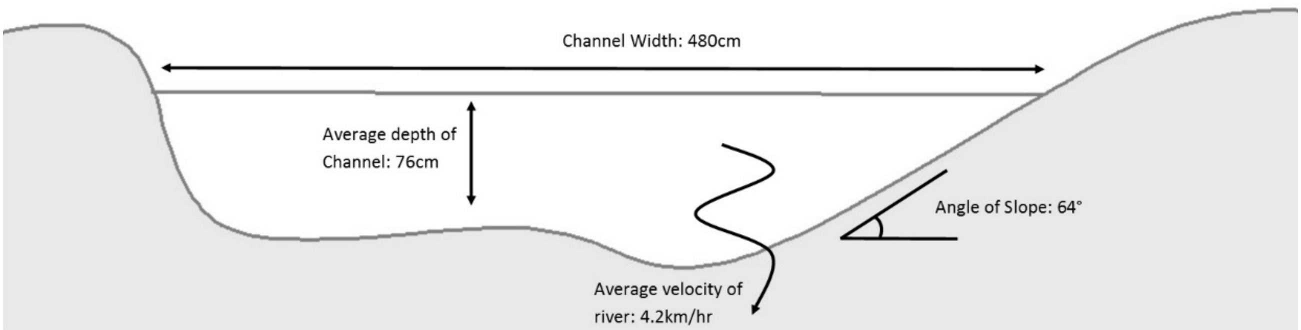
A labelled photograph of Dar Es Salaam.

If the researcher knows the direction in which the photograph was taken, or the orientation of the view of the fieldsketch, this should be included in the title of the graphic.

Title: **<u>Dar Es Salaam Port (facing South West)</u>**

The same photograph with annotations.

It is rare that a photograph, field sketch or cross section sketch will be used in a report write-up in isolation. Quantitative data that was also collected first hand may be put into the picture using arrows and appropriate annotations. This is especially true when data may be difficult to describe without a pictorial guide for the reader.



A cross section of a river annotated with data collected from one site in the field

3l – A Guide to
Presenting Qualitative Data

Royal Geographical Society with IBG
Advancing geography and geographical learning

Numerical data naturally lends itself to a wide range of different data presentation techniques and good researchers exploit a large number of these within any particular study. It is tempting to ignore the qualitative data of an investigation, such as interview transcripts and observations, until the analysis stage as many researchers feel that the 'wordy' bit of the data cannot be presented in any way other than as a copy of the literal words used.

In fact, qualitative data can, and should, be presented in a number of interesting and attractive ways.

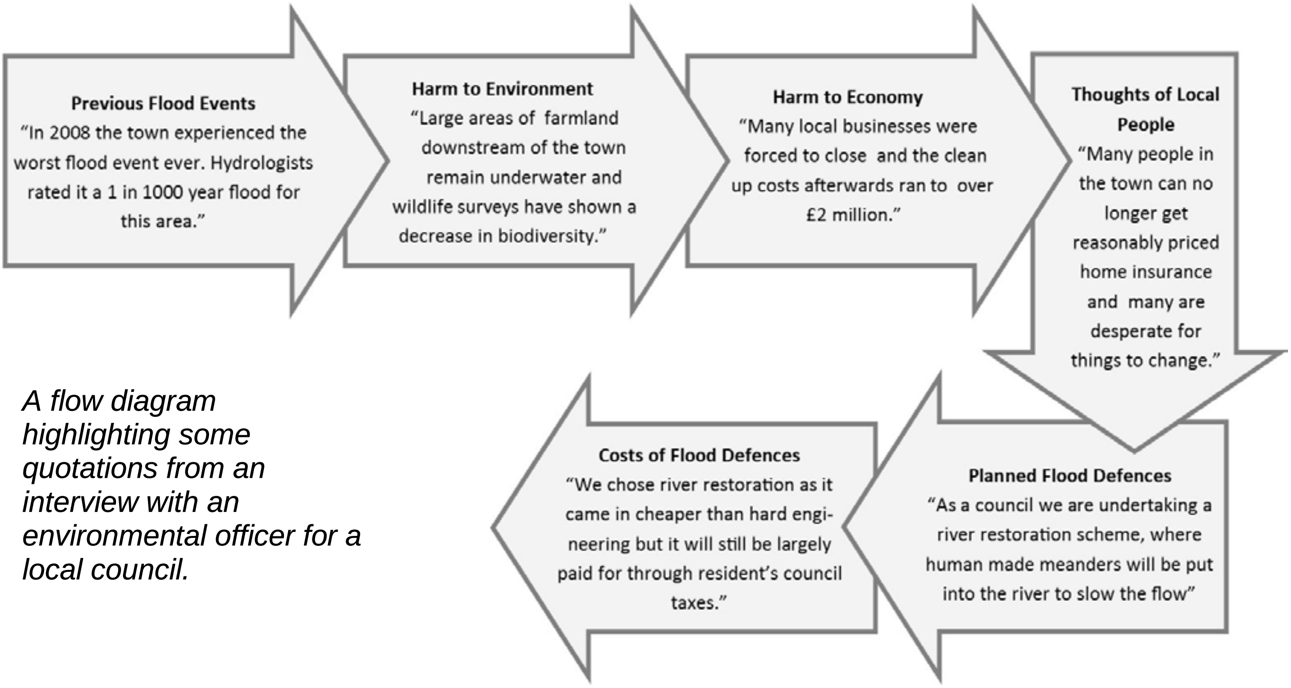
Quotation Banks

Creating a transcript of any interviews that are taken or oral histories that are recorded should provide the researcher with a number of relevant quotations, each of which can be used as a piece of data. By organising the quotations into themes or categories, a quotation bank can be created for each issue or idea and organised graphically on the page.

A quote bank about the street furniture in the centre of Town A

Flow Diagrams and Mind Maps

A good interview will be one that has a narrative running through it rather than a series of random questions. Therefore, it is possible to create a flow diagram or mind map showing how the interview has evolved, and with a series of quotations or observations rather than a whole transcript of the interview.



A flow diagram highlighting some quotations from an interview with an environmental officer for a local council.

Equally, personal observations can be mind-mapped, overlaying a true map or an interpretive representation of the space in question.

A map of some of the researcher's observations of a town.

Word Clouds

An interview, a recording of an oral history, or a questionnaire where the questions require an open answer can generate huge volumes of descriptive words and texts which can be difficult to analyse. Creating a word cloud out of adjectives can be a useful graphical way of seeing where a consensus of opinion may lie. A word cloud presents the most frequently used adjectives as the largest in the graphic and those used less frequently as far smaller.

There are a number of free online packages which can create word clouds with relative ease. Most need the researcher to simply type the data (in this case adjectives) into the programme and it will automatically generate a word cloud which can be downloaded.

A word cloud showing how local residents described the stream.

Section 4 –
Data Analysis

The **data analysis** is the stage of the investigation where the researcher manipulates the data to make it more meaningful, easier to use and in a format whereby the research questions can be answered. It is commonly written up at the same time, and alongside, the data presentation section. It is highly likely that it will make sense to analyse some data before it is presented graphically and vice versa.

For numerical data, analysis usually involves the use of statistics, in which the data is put through tried and tested methods to show whether there are any relationships between variables within it and how significant these relationships actually are. Some statistical methods are relatively straightforward, such as calculating a percentage, or the mean (average) from a set of data. Others require a more in-depth mathematical understanding though, in most cases, it may be enough to simply use the formulae given to you and understand what the figures mean at the end of the calculation. Remember to use an appropriate form of statistical analysis. Properly explaining your results can positively contribute to gaining a good mark in your investigation. It is not uncommon for geographers to feel concerned about statistical analysis. You should not feel daunted by the prospect of using statistical methods as there is plenty of guidance available, along with step-by-step instructions, to help you.

As with every part of your Independent Investigation, it is important that the statistical methods chosen are **appropriate** both for the data in hand and the research questions or hypotheses the study aims to answer. A statistical method should certainly not be chosen simply because it is easy or the only one you know: it is worth spending some time exploring the different methods available to see which best suits the hypotheses. Each statistical method is designed with different aims. Therefore, you should not enter into the data analysis under the belief that with any one set of data, you can pick and choose from a menu of statistical techniques. Instead, you should be aware that each statistical technique has one purpose. If that purpose fits the aims of the study that statistical technique has effectively ‘chosen you’. Some statistical techniques work best alongside certain data presentation techniques too.

A summary of some statistical techniques and their uses is highlighted below:

Name of technique	Use	Use with data presentation techniques
Measures of Central Tendency	To show the most common observed data	Bar chart Radial graph
Measures of Proportion	To find out what proportion of a whole the data is representing	Pie chart Triangular graph Composite bar chart Proportional shape map
Dispersion	To show the frequency of data around a central idea	Box and Whisker graph Line graph

Spearman's Rank Correlation	To show the strength and nature of a correlation between two sets of data	Scatter graph
Chi-Squared Test	To show how closely observed data matches data that is expected	Bar chart Histogram
Simpson's Diversity Index	To show the frequency of different categories as well as the abundance of those categories	Kite diagram Choropleth map Isoline map
Pearson's Product Moment	To show the strength of correlation between two variables that show a linear relationship	Scatter graph
Nearest Neighbour Analysis	To show the degree to which something is clustered or uniformly spaced	Choropleth map Isopleth map
Mann Whitney U Test	To show whether two data samples are significantly different from one another	Choropleth map Line graph Scatter graph

It is also good practice to **justify** to the reader the reasons for choosing one statistical method over another.

Non-numerical (**qualitative**) data, as well as **secondary data** can also be analysed. In the former, this may involve the conversion of 'wording' to numerical data (e.g. by counting the number of times certain negative or positive opinions are made during an interview) and the coding and indexing of transcripts can be used to make the wealth of 'word data' more manageable. Equally, secondary data can be subjected to the same statistical analyses as primary data, allowing the researcher to create a larger data set for analysis or more easily compare data from different time frames.

After reading your data analysis, the reader and marker of your Independent Investigation should be left with no surprises as to the type of **conclusions** you will be making. Therefore, the data analysis is the last stage at which the researcher can reveal any new or different connections in the data that have not already been explained; the data analysis section leads the reader naturally to the conclusion. However, conclusions can only be made if the data analysis shows that particular idea to be true: it is not good research practice to allow a hunch or personal instinct about a place and a situation to cloud one's analysis of the data. If the data does not show something to be true, no amount of data handling will make it so. Under no circumstances should the researcher be tempted to change or make up the raw data itself (such as removing outlier or anomaly data) in order to suit the conclusions that they wish to reach.

Common Pitfalls:

- **Carrying out more than one statistical test on the same set of data.** You will not gain extra marks for this – it is much better to choose the most appropriate statistical analysis method and justify your choice afterwards.
- **Only analysing some of the data.** If you intend to draw a conclusion from some data, or answer a research question from it, then it must be included in the data analysis.
- **Choosing a statistical analysis method on the ease of the mathematics required alone.** The best reason for choosing one statistical method over another is because it is most appropriate for the data you have and the nature of the conclusions you wish to draw.
- **Including pages of calculations.** It is best to put any workings in an appendix in case the reader and marker of your study wishes to check how a statistical method has been used.
- **Changing the data itself to suit a preconceived conclusion.** If the data analysis hasn't shown the conclusion you were expecting, there is likely to be a geographical reason for this, or there may be a limitation in the way you conducted your study. Both circumstances give you more interesting things to write about, compared to the situation of everything working out as you expected.
- **Making grand conclusions from under analysed data.** Unless you can prove on paper something you are saying, don't make a simplified statement.
- **Providing raw data as well as analysed data.** There is no reason to provide pages of raw data if you are going to ultimately show it in its manipulated form.

4a – A Guide to
Measures of Central Tendency

Royal Geographical Society with IBG
Advancing geography and geographical learning

Measures of Central Tendency are ways of working out the most representative or central value within a data set.

For the examples below, the following data set will be used.

8	12	3	5	23	19	14	8	6	16	10	8	13	9	27	16	2	9	5

Distances (in km) that interviewees travelled to access the High Street of Town A.

Mode

The mode for a set of data is that which occurs most frequently within the set.

When the example data set is put in value order, one can see that the mode is **8 km**.

2	3	5	5	6	8	8	8	9	9	10	12	13	14	16	16	19	23	27

It is important to remember when using the mode that it does not always represent the centre of a distribution. Equally, identifying the mode is not always possible for all data sets – continuous data for example may produce a mode that is meaningless as there may be one frequency of every value. In some data sets there may be more than one mode, making the use of this analysis impractical.

Median

The median is the middle value when the data set is placed in value order. If there are an odd number of values in the data set, the middle value is taken as the median. If there are an even number of values in the data set, the mid-point between the two middle values is taken as the median.

When the example data set is put in value order, one can see that the median is **9 km**.

2	3	5	5	6	8	8	8	9	9	10	12	13	14	16	16	19	23	27

Identifying the median is only possible when the data set can be ordered, but using the median can offset the possible problems associated with outliers and anomalous data values.

Mean

The mean is the sum of all the values in the data set divided by the number of values within the data set.

For the example data set the sum of all the values is **213**.

2 + 3 + 5 + 5 + 6 + 8 + 8 + 8 + 9 + 9 + 10 + 12 + 13 + 14 + 16 + 16 + 19 + 23 + 27 = 213

The total number of values within the data set is **19**.

213 divided by 19 = 11.2

Therefore, the mean is **11.2 km**.

One advantage of using the mean is that it can be used for almost all types of data set. However, it is also true that the mean can be greatly influenced by an anomaly within the data set.

4b – A Guide to Measuring Proportions

Royal Geographical Society with IBG

Advancing geography and geographical learning

Calculating the proportion of a data set that display a particular quality can be a good way of making comparisons between places or circumstances.

Fractions

Fractions are the expression of a value out of 1, where a value of 1 represents the whole of the value of the data set.

Fractions should always be expressed in their smallest form: for example $\frac{1}{4}$ is used rather than $\frac{3}{12}$.

In a research study, if there are forty people within a data set and ten of them are from Town A, ten are from Town B and twenty are from Town C, the following fractions are true:

$\frac{1}{4}$ are from Town A $\frac{1}{4}$ are from Town B and $\frac{1}{2}$ are from Town C.

Percentages

A percentage is a fraction expressed as a value out of 100. It can be expressed as a percentage figure or a decimal.

The method for calculating percentages will already be well known to students at this level.

Percentages can be a useful way of presenting your findings.

Ratios

A ratio is normally used to show the quantity of one variable compared to another, and though in most cases this results in only two values being expressed against each other. It is also possible to have more than two variables in a ratio.

For example, if out of twenty people in a seaside café, five are local people compared to fifteen people who are tourists, the ratio of locals to tourists is 5 to 15, or 5:15. Ratios are always expressed in their lowest common form, so in this case, **1:3**

If, however, there are five local people, six tourists from the UK and nine tourists from overseas in the cafe, the ratio is **5:6:9**, (which cannot be expressed in any lower form).

Stating a ratio on its own is not really enough as the reader will not know without explanation in which order you have considered the variables (for example, what the 9 relates to exactly in the 5:6:9 ratio).

4c – A Guide to Measures of Dispersion

Royal Geographical Society with IBG

Advancing geography and geographical learning

The range of values over which a data set is found can tell you a lot about the degree of consensus within the data. This is known as **dispersion** and can be measured using some statistical analyses.

Standard Deviation

The standard deviation (SD) of a data set is a measure of the extent to which the range of its values differ from the mean value.

A low standard deviation tells the reader that the values of the data set are close to the mean, and the range of the set is relatively low. This means that the reliability of any conclusions you draw, by using the mean as your indicator, will be higher.

A high standard deviation dictates that the values of the data set are spread out over a wider range and are relatively far from the mean. Therefore, any conclusions that are drawn using the mean as an indicator, will be more tenuous and less reliable.

The size of a single standard deviation on its own is somewhat meaningless, but when one compares it with another data set (such as that of another geographical site) the standard deviation starts to have more meaning.

Calculating the standard deviation for a data set is a little complicated and for very large data sets it is worth considering entering the data first into a spreadsheet which will allow you to carry out each stage of the calculation to all of the values at once. The following shows a step-by-step guide to identifying the standard deviation for a relatively small data set.

For this example, the following data set will be used:

8	**12**	**3**	**5**	**23**	**19**	**14**	**8**	**6**	**16**	**10**	**8**	**13**	**9**	**27**	**16**	**2**	**9**	**5**

Distances (in km) that interviewees travelled to access the High Street of Town A.

Step 1. Calculate the mean value of the data set.

2 + 3 + 5 + 5 + 6 + 8 + 8 + 8 + 9 + 9 + 10 + 12 + 13 + 14 + 16 + 16 + 19 + 23 + 27 = 213

213 / 19 = **11.2 km**

Step 2. For each value in the data set, subtract the mean.

8	12	3	5	23	19	14	8	6	16	10	8	13	9	27	16	2	9	5

minus 11.2 =

-3.2	0.8	-8.2	-6.2	11.8	7.8	2.8	-3.2	-5.2	4.8	-1.2	-3.2	1.8	-2.2	15.8	4.8	-9.2	-2.2	-6.2

Step 3. Square the resulting figures, thus removing any negative values.

-3.2	0.8	-8.2	-6.2	11.8	7.8	2.8	-3.2	-5.2	4.8	-1.2	-3.2	1.8	-2.2	15.8	4.8	-9.2	-2.2	-6.2

squared =

10.24	0.64	67.24	38.44	139.24	60.84	7.84	10.24	27.04	23.04	1.44	10.24	3.24	4.84	249.64	23.04	84.64	4.84	38.44

Step 4. Calculate the mean of the figures produced from stage 3.

10.24 + 0.64 + 67.24 + 38.44 + 139.24 + 60.84 + 7.84 + 10.24 + 27.04 + 23.04 + 1.44 + 10.24 + 3.24 + 4.84 + 249.64 + 23.04 + 84.64 + 4.84 + 38.44 = 805.16

805.16 / 19 = **42.38**

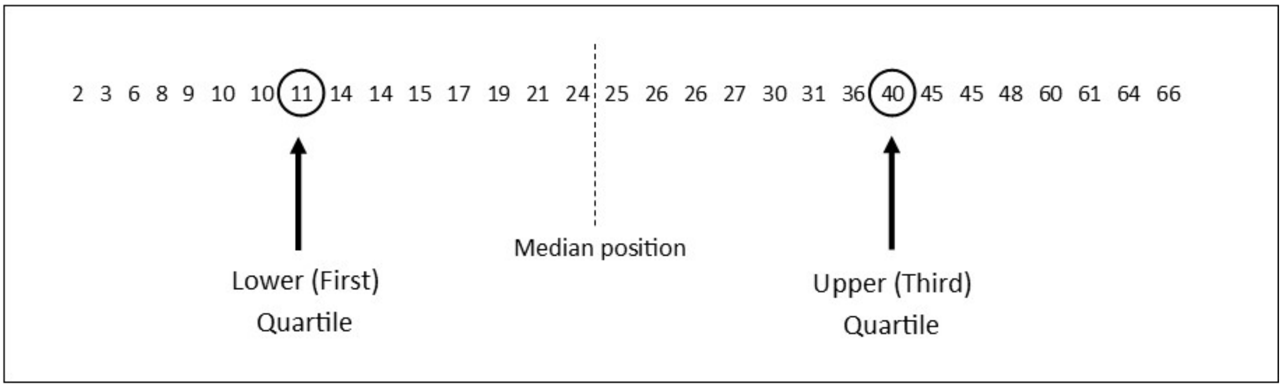
Step 5. Square root this mean. This is the standard deviation for this data set.

$\sqrt{}$ 42.38 = **6.51** $\boldsymbol{SD}$

Interquartile Range

A quartile is a term that refers to 25% of the values in a data set. The first (or lower) quartile is the value that, when the values are placed in numerical order, sits halfway between the median value and the minimum value. The second quartile ***is*** the median value and, therefore, not generally quoted as such, while the third (or upper) quartile is the value that sits halfway between the median point and the maximum value.

So, if the following set of data represents a sample of the age of visitors to a theme park on a particular day, the lower quartile is 11 years old and the upper quartile is 40 years old. The median value is 24.5 years old (the difference between the two values that sit either side of the median position).



The interquartile range is the difference between the lower and upper quartiles. In the above example, therefore, the interquartile range is 29 years (40 years old minus 11 years old).

As with standard deviation the interquartile range on its own is not wholly meaningful. The larger the interquartile range, the greater the spread of the data across the range of values. The smaller the interquartile range, the more likely the use of the median is going to be for making reliable conclusions as the spread of the values from the median is smaller.

4d – A Guide to Cost-Benefit Analysis

Royal Geographical Society with IBG

Advancing geography and geographical learning

Cost-Benefit Analysis is a useful method for identifying the overall consensus about an issue when considering many different opinions on a range of different aspects. Cost-benefit analysis makes use of **weighted scores**, whereby the relative importance of different issues is taken into consideration when scoring their negative and positive points.

Why would we use a cost-benefit analysis?

A researcher looking into future proposals or prospective modelling for a geographical idea will find it is very useful to consider the relative weaknesses and merits of different options. A cost-benefit analysis allows a researcher to consider different perspectives with a greater degree of objectivity by allowing mathematical scores to determine an outcome rather than emotions.

Worked Example:

A researcher may ask questionnaire participants to consider the relative merits of a range of different flood protection schemes, and the strength of their negative impacts in a range of different areas:

	Impact you feel there will be on… Score 1 (low negative impact) to 5 (high negative impact)				
	Visual attractiveness	Traffic	Noise	House prices	Insurance Premiums
Scheme A					
Scheme B					
Scheme C					
Scheme D					

To some sets of participants, the importance of a change in the value of their home may be far greater than a change to the visual attractiveness of their area. Therefore, more 'weight' should be added to the former score as it is more important to them. The researcher can choose the exact scale given to this weighting, or they might like to carry out a pilot study that asks public respondents to allocate a 'weight' to each of the criteria based on how important they feel those issues are. In this case, weighting the score out of three, where three is most important and one least important, allows the final score to reflect the relative importance of each of the categories.

	Impact you feel there will be on… Score 1 (low negative impact) to 5 (high negative impact)														
	Visual attractiveness			Traffic			Noise			House prices			Insurance Premiums		
	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**
Scheme A		2			2			1			3			3	
Scheme B		2			2			1			3			3	
Scheme C		2			2			1			3			3	
Scheme D		2			2			1			3			3	

Weighting a score means that the original score is multiplied by the ‘weight’ added to it. The same weighting must be used for each choice (in this case flood management scheme) so a direct comparison between them can be made. To allow the researcher to choose the flood management scheme most favoured by the participants, the total weighted score for each scheme should also be considered.

	Impact you feel there will be on… Score 1 (low negative impact) to 5 (high negative impact)															**TOTAL SCORE**
	Visual attractiveness			Traffic			Noise			House prices			Insurance Premiums			
	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	Score	Weighting	**Weighted Score**	
Scheme A	1	2	**2**	3	2	**6**	1	1	**1**	2	3	**6**	1	3	**4**	**19**
Scheme B	3	2	**6**	3	2	**6**	2	1	**2**	3	3	**9**	1	3	**4**	**27**
Scheme C	5	2	**10**	1	2	**2**	4	1	**4**	5	3	**15**	2	3	**6**	**35**
Scheme D	2	2	**4**	1	2	**2**	2	1	**2**	1	3	**4**	2	3	**6**	**18**

Therefore, according to this cost-benefit analysis, Scheme D is deemed by the participants in this survey to have the least negative impact, while Scheme C will have the most negative impact.

A Guide to
Spearman's Rank

Royal Geographical Society with IBG

Advancing geography and geographical learning

The **Spearman's Rank Correlation Coefficient** is a statistical test that examines the degree to which two data sets are correlated, if at all. While a scatter graph of the two data sets may give the researcher a hint towards whether the two have a correlation, Spearman's Rank gives the researcher a numerical value on the degree of correlation, or indeed, the degree of non-correlation. It is a relatively straight forward analysis for those researchers whom are not wholly confident in their mathematical skills.

In order to use Spearman's Rank the researcher must have paired sets of data that are in some way related (such as the geographical site where they were collected in the field). It is a good idea for the researcher to have at least ten pairs of data to use for the analysis: any fewer than this and the result will be highly insignificant and more likely to be as a result of chance than of true correlation.

In the following example, the researcher is looking at whether for River X, the channel width increases as the distance from the source increases. Theoretically, this should be true, but the Spearman's Rank analysis will tell the researcher whether it is true in this case that there is a correlation and the strength of any such correlation.

1. The researcher should arrange the paired data in a table to allow for ease of analysis. This can be done in a spreadsheet package or through hand written methods.

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150		0.40			
2	300		0.80			
3	450		1.00			
4	600		0.95			
5	750		1.20			
6	900		1.10			
7	1050		1.30			
8	1200		1.40			
9	1350		1.85			
10	1500		2.40			
11	1650		2.55			
12	1800		3.20			
13	1950		3.80			
14	2100		3.60			
15	2250		3.20			
					Total	

Nuffield Foundation

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

2. Then researcher should rank each data, starting with 1 as the smallest figure and (in this case) 15 as the largest. Where there might be two values that are equal, the researcher should average the ranks and omit the ranking values they cover. For example, a set of rankings may read: 1, 2, 3.5, 3.5, 5, 6, 7.25, 7.25, 7.25, 7.25, 11, 12, etc

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150	1	0.40	1		
2	300	2	0.80	2		
3	450	3	1.00	4		
4	600	4	0.95	3		
5	750	5	1.20	6		
6	900	6	1.10	5		
7	1050	7	1.30	7		
8	1200	8	1.40	8		
9	1350	9	1.85	9		
10	1500	10	2.40	10		
11	1650	11	2.55	11		
12	1800	12	3.20	12.5		
13	1950	13	3.80	15		
14	2100	14	3.60	14		
15	2250	15	3.20	12.5		
					Total	

3. The difference (d) between the two ranks should then be calculated by subtracting R_1 from R_2:

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150	1	0.40	1	0	
2	300	2	0.80	2	0	
3	450	3	1.00	4	-1	
4	600	4	0.95	3	1	
5	750	5	1.20	6	-1	
6	900	6	1.10	5	1	
7	1050	7	1.30	7	0	
8	1200	8	1.40	8	0	
9	1350	9	1.85	9	0	
10	1500	10	2.40	10	0	
11	1650	11	2.55	11	0	
12	1800	12	3.20	12.5	-0.5	
13	1950	13	3.80	15	-2	
14	2100	14	3.60	14	0	
15	2250	15	3.20	12.5	2.5	
					Total	

4. d should then be squared to remove any negative values. The total value of all the d^2 can also be calculated at this stage.

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150	1	0.40	1	0	0
2	300	2	0.80	2	0	0
3	450	3	1.00	4	-1	1
4	600	4	0.95	3	1	1
5	750	5	1.20	6	-1	1
6	900	6	1.10	5	1	1
7	1050	7	1.30	7	0	0
8	1200	8	1.40	8	0	0
9	1350	9	1.85	9	0	0
10	1500	10	2.40	10	0	0
11	1650	11	2.55	11	0	0
12	1800	12	3.20	12.5	-0.5	0.25
13	1950	13	3.80	15	-2	4
14	2100	14	3.60	14	0	0
15	2250	15	3.20	12.5	2.5	6.25
					Total	14.5

5. One should then apply the Spearman's Rank equation to calculate the coefficient value (R) (the value that tells the researcher the strength of the correlation).

$$R = 1 - \frac{6\Sigma d^2}{(n^3 - n)}$$

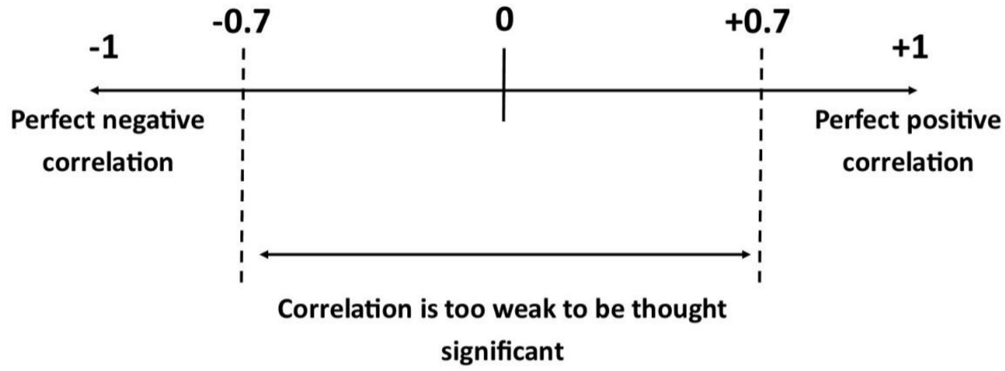
where n is the number of pairs of data collected and used (in this case 15). The sum of the d^2 values (Σd^2) in this example is 14.5.

Therefore, the equation can be calculated as follows:

$$R = 1 - \frac{87}{(3375-15)} \qquad R = 1 - \frac{87}{3360} \qquad R = 1 - 0.026$$

$$R = 1 - 0.026 \qquad R = 0.974$$

The coefficient (R) will be between a value of -1 and +1 where -1 indicates a perfect negative correlation and +1 indicates a perfect positive correlation. A value of between -0.7 to +0.7 is generally seen as being too weak to be thought of as a significant result.



Therefore, the data in this example shows a strong positive correlation between channel width and the distance from the source.

6. To check whether the result is meaningful or is just down to chance, the value for R can be compared with the critical value for n in the Spearman's Rank significance table.

Below is the significance table for some values of n, but for analysis of larger sets of data, extended significance tables can be found online.

	Significance Level		
n	0.1	0.05	0.01
4	1.000	1.000	1.000
5	0.700	0.900	1.000
6	0.657	0.771	0.943
7	0.571	0.679	0.857
8	0.548	0.643	0.810
9	0.483	0.600	0.767
10	0.442	0.564	0.733
11	0.418	0.527	0.700
12	0.399	0.504	0.671
13	0.379	0.478	0.648
14	0.367	0.459	0.622
15	0.350	0.443	0.600
16	0.338	0.427	0.582
17	0.327	0.412	0.558
18	0.317	0.400	0.543

The critical value for this example, where there are 15 pairs of data (n = 15), is 0.443. As the value of R is greater than the critical value, we can say with 95% certainty that the results we have observed have not occurred by chance. This means the results are highly significant and sound conclusions can be drawn from them.

4f – A Guide to Chi-Squared Testing

Royal Geographical Society with IBG

Advancing geography and geographical learning

The Chi-Squared test (also known as the Index of Association) looks at how closely data the researcher has collected fits with a theoretical or expected set of data. It is sometimes referred to as a ‘goodness of fit’ test as it is used to see how closely observed findings match those that are known to be true on a wider scale.

Why would we use the Chi-Squared test?

A researcher would use the Chi-Squared test when their data is showing the frequency of an occurrence of something. This test only checks to see if there is an association between two sets of data, not what the nature of the relationship might be between those sets, nor the strength of any relationship.

Worked Example:

The following example will look at the age structure of visitors to a tourist attraction in a particular town. A Chi-Squared test will be carried out to see if the observed age structure at the attraction is indicative of the age structure of the town in general.

The researcher has found secondary data which represents the annual number of visitors to the attraction in different age groups. Census data was also found to show the age structure of the town in which the attraction was situated.

Age grp	**Number of visitors**	**Population of Town**
0 – 19	280,580	41,370
20 – 39	240,500	48,830
40 – 59	121,250	44,730
60 – 79	80,160	30,500
80+	2,080	7,650
Total	724,570	173,080

1. The researcher should firstly work out the expected data (in this case the expected number of visitors). To do this, convert the known data of the town's population into percentages and use these same percentages to predict the number of visitors there will be to the attraction in each of the age categories.

Age grp	Population of Town	Population as a %
0 – 19	41,370	23.9
20 – 39	48,830	28.2
40 – 59	44,730	25.8
60 – 79	30,500	17.6
80+	7,650	4.5
Total	173,080	100

Total number of visitors = 724,570

Therefore, the number of predicted visitors aged 0 – 19 is

$$\frac{23.9 \times 724570}{100} = 173{,}172$$

This calculation should then be done for each age category for the number of visitors to create a set of predicted (or expected) values:

Age grp	Expected visitor numbers	As a %
0 – 19	173,172	23.9
20 – 39	204,329	28.2
40 – 59	186,939	25.8
60 – 79	127,524	17.6
80+	32,603	4.5
Total	724,570	100

Where no real data is available on which to base an expected set of values, the researcher should divide the total value (724,570) equally between the number of categories in order to get expected values. This would suggest that the category or variable in question would have no influence over the observed data as the values would all be the same (in this case 144,914 predicted visitors in each age category). It is then the role of the researcher, through carrying out the Chi-Squared test, to see if the categories or variable have a role to play in the geography being observed.

The expected frequencies or values do not have to be whole numbers to further work within the Chi-Squared calculations.

2. The researcher should then table the observed and expected values as follows:

	0 – 19	20 - 39	40 - 59	60 - 79	80+
Observed value or frequency (O)	280,580	240,500	121,250	80,160	2,080
Expected value or frequency (E)	173,172	204,329	186,939	127,524	32,603

3. For each category, the researcher should then calculate the value of $\frac{(O - E)^2}{E}$

For example, 0 – 19 age group

$$\frac{(280580 - 173172)^2}{173172} = \frac{107408^2}{173172} = \frac{11536478464}{173172} = 66,618$$

	0 – 19	20 - 39	40 - 59	60 - 79	80+
$\frac{(O - E)^2}{E}$	66,618	6,403	23,082	17,592	28,575

4. Add all these values together:

$\Sigma \frac{(O - E)^2}{E}$ = 142,270 This is the Chi-Squared value (shown as the symbol χ^2).

5. The researcher can then use a data table (part of which printed here or otherwise found through online searches) to work out if this value is significant.

   The 'degree of freedom' is the number of categories you have in your test minus one (in this case, we have 4 degrees of freedom).

> For the purpose of most geography research, an exact understanding of what 'degrees of freedom' means is not necessary and the reader should not worry at this stage if this term remains unexplained except for how it should be used in a significance table such as that below.

Significance Table (also known as the Critical Values of Chi-Squared Table):

Degree of freedom (df)	Significance Level				
	10%	5%	2.5%	1%	0.5%
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	**9.49**	11.14	13.23	14.86
5	9.24	11.07	12.83	15.09	20.52

A significance table tells us how reliable the result is and whether the observed data is as a result of a geographical factor, or due to complete chance. Using a 5% (sometimes written as 0.05) significance level tells us that there is only a 5% possibility of the results being as they are by chance. A 5% significance level is commonly used by geographical researchers.

If the Chi-Squared value is greater than the appropriate value in the table, the observed data is significantly different from the expected data. Therefore, there is more likely to be an association between the collected data and the variable in question (in this example, the tourist attraction). The data is said to be dependent on something, and not just a result of random numbers coinciding.

If the Chi-Squared value is less than the appropriate value in the table, then the observed data is not significantly different from the expected data. This does not mean there is no association between the observed data and the variable in question, just that there is not enough evidence to show this.

In this example, the size of the sample is huge so it is not unexpected that the Chi-Squared value is also very large, suggesting an association between the attraction and age of visitor. In many student-based studies the researcher will be working with primary rather than secondary data so it is highly likely that a smaller Chi-Squared value will be used when looking up the result in a significance table.

4g – A Guide to Simpson's Diversity Index

Royal Geographical Society with IBG
Advancing geography and geographical learning

Simpson's Diversity Index is used to calculate a measure of diversity, taking into account the number of something as well as its abundance. The index is most often used for ecological studies that measure species diversity, but the same analysis can also be applied to other principles, such as the diversity of opinion on an idea over a geographical space.

Why would we use Simpson's diversity index?

Researchers frequently need to know if the range of data they produce is showing a true level of variety or if it just appears to do so on paper. This is especially the case when researchers are dealing with very large quantities of data and the level of diversity within that data is not easy to decipher from reading a table of results.

The index measures the probability that two randomly selected individuals from a sample will be the same. The formula for calculating the value of the index (D) is

$$D = 1 - \frac{\Sigma\, n(n\text{-}1)}{N(N\text{-}1)}$$

where n is the number of individuals displaying one trait (e.g. the number of individuals of one species)
N = the total number of all individuals

Worked Example:

To best use this calculation, it is a good idea to tabulate the findings. In this example, the different species seen in quadrats at two sand dune sites are compared.

Species	Location A quadrat			Location B quadrat		
	n (number of individuals)	n-1	$n(n$-1)	n	n-1	$n(n$-1)
Sea couch grass	21	20	420	18	17	306
Sea twitch grass	1	0	0	11	10	110
Sea rocket	8	7	56	8	7	56
Prickly saltwort	1	0	0	3	2	6
Lyme grass	12	11	132	7	6	42
N (total number of individuals)	43			47		
		Σ	608		**Σ**	520

For Location A:

$D = 1 - \frac{608}{43 \times 42}$ $D = 1 - \frac{608}{1806}$ $D = 1 – 0.337$ $D = 0.663$

For Location B:

$D = 1 - \frac{520}{47 \times 46}$ $D = 1 - \frac{520}{2162}$ $D = 1 – 0.241$ $D = 0.759$

The value of D will always fall between 0 and 1, where 1 represents complete diversity and 0 represents complete uniformity.

One index value on its own holds very little value: but once the researcher is able to compare it to another, it can begin to show something. For example, in the above example, the higher value at Location B shows that from the data collected, Location B appears to be more diverse in species than Location A.

A Guide to
Pearson's Product Moment

Royal Geographical Society with IBG
Advancing geography and geographical learning

Pearson's Product Moment Correlation Coefficient measures the degree of correlation there may be between two variables. It is best used when results have already been plotted on a scatter graph and there is an indication of a linear relationship between the two factors. In order for Pearson's Product Moment to be useful, both sets of data have to be continuous and there should be no significant outliers (anomalies) in the manner in which the data is distributed.

Analysis of data using Pearson's Product Moment gives a numerical value (r) between -1 and +1 to show the strength of any correlation.

In the following worked example, the researcher is trying to find the strength of correlation between the measure of the velocity of the water in a river channel and the gradient of the river at various points.

1. Label the paired data sets x and y for ease of use.

Gradient (x)	8.5	7.1	5.2	5.5	2.0	3.5	4.3	4.7	6.0	5.5	1.1	7.4	3.2	2.4	6.4
Velocity (y)	0.51	0.11	0.24	0.33	0.40	0.35	0.27	0.24	0.15	0.18	0.48	0.08	0.31	0.38	0.16

2. Work out the following figures for each piece of data and then sum the figures. It is best to do this in a spreadsheet package so that the researcher does not have to undertake hundreds of calculations.

																Σ
x	8.5	7.1	5.2	5.5	2.0	3.5	4.3	4.7	6.0	5.5	1.1	7.4	3.2	2.4	6.4	**72.8**
y	0.51	0.11	0.24	0.33	0.40	0.35	0.27	0.24	0.15	0.18	0.48	0.08	0.31	0.38	0.16	**4.19**
x^2	72.3	50.4	27.0	30.3	4.0	12.3	18.5	22.1	36.0	30.3	1.21	54.8	10.2	5.76	41.0	**416**
y^2	0.26	0.01	0.06	0.11	0.16	0.12	0.07	0.06	0.02	0.03	0.23	0.01	0.10	0.14	0.03	**1.41**
xy	4.34	0.78	1.25	1.82	0.80	1.23	1.16	1.13	0.90	0.99	0.53	0.59	0.99	0.91	1.02	**18.4**

3. Using this summed data, the following calculations should then be worked out:

   (n = the number of pairs of data)

$ssxy = \Sigma\, xy - \dfrac{(\Sigma x \Sigma y)}{n}$ $\quad ssxy = 18.4 - \dfrac{(72.8 \times 4.19)}{15}$ $\quad ssxy = 18.4 - \dfrac{(305.032)}{15}$

$ssxy = 18.4 - 20.34$ $\quad \boxed{ssxy = -1.94}$

Nuffield Foundation

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

$$ssxx = \Sigma x^2 - \frac{(\Sigma x \Sigma x)}{n} \qquad ssxx = 416 - \frac{(72.8 \text{ x } 72.8)}{15} \qquad ssxx = 416 - \frac{(5299.84)}{15}$$

$$ssxx = 416 - 353.32 \qquad \boxed{ssxx = 62.68}$$

$$ssyy = \Sigma y^2 - \frac{(\Sigma y \Sigma y)}{n} \qquad ssyy = 1.41 - \frac{(4.19 \text{ x } 4.19)}{15} \qquad ssyy = 1.41 - \frac{(17.56)}{15}$$

$$ssyy = 1.41 - 1.17 \qquad \boxed{ssyy = 0.24}$$

4. To work out the Pearson's Product Moment correlation coefficient (r) use the above boxed results in the following equation:

$$r = \frac{ssxy}{\sqrt{ssxx \text{ x } ssyy}} \qquad r = \frac{-1.94}{\sqrt{62.68 \text{ x } 0.24}} \qquad r = \frac{-1.94}{\sqrt{15.04}} \qquad r = \frac{-1.94}{3.878} \qquad r = -0.50$$

A value of -1 indicates a perfect negative correlation while a value of +1 indicates a perfect positive correlation. A value of zero suggests no correlation at all.

Therefore, in this example, the results show a negative correlation between gradient and river velocity, which is what we might expect, but the results only show a weak correlation. The next role of the geographical researcher is to decide whether there are enough samples in the data for the result to be valid or whether there are other, site specific, reasons why the data does not match the theory.

4i – A Guide to Nearest Neighbour Analysis

Nearest Neighbour Analysis measures the spread or distribution of something over a geographical space. It provides a numerical value that describes the extent to which a set of points are clustered or uniformly spaced.

Why would we use nearest neighbour analysis?

Researchers use nearest neighbour analysis to determine whether the frequency with which something is observed spatially is comparable with other locations. It can provide researchers with a numerical value for the 'clustering' of a geographical phenomenon, allowing this value to be compared more accurately with other places.

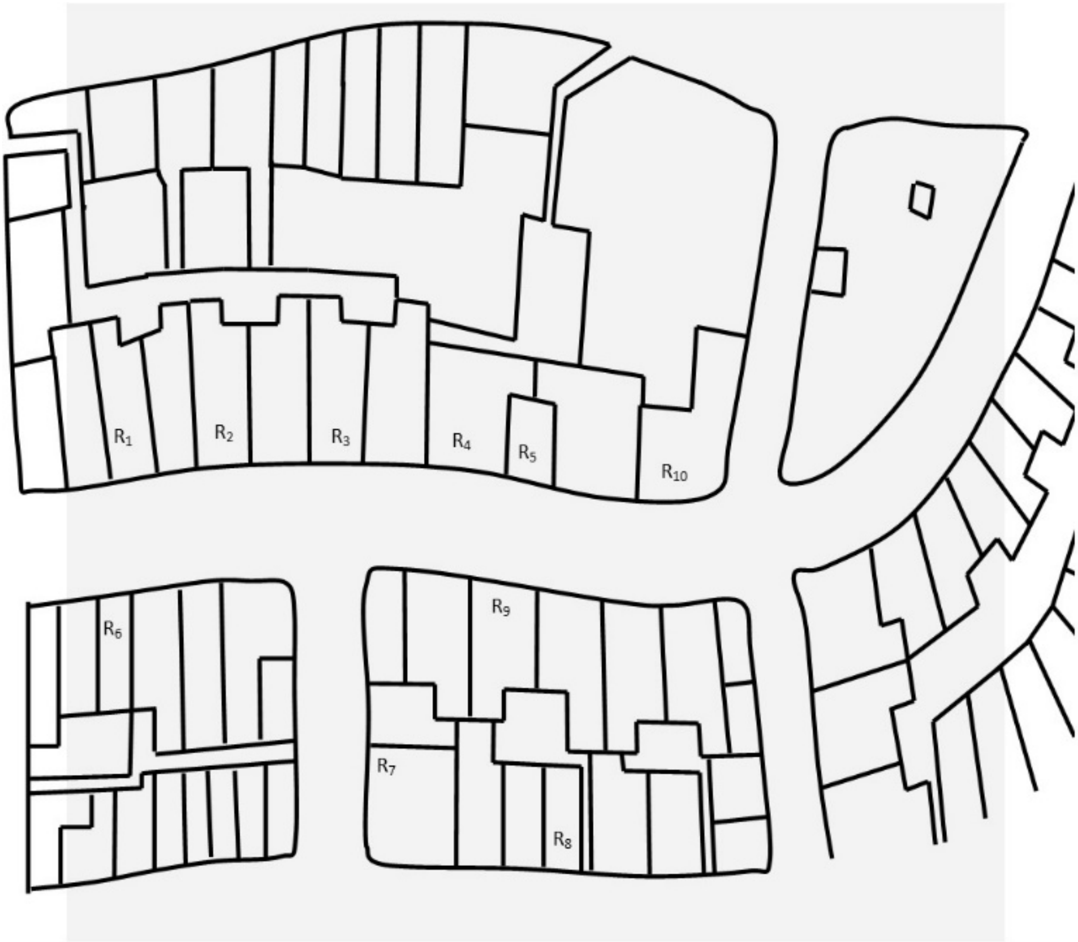
Worked Example:

In this example, researchers have mapped the land use of each building in a 200m by 200m plot in a town centre, using a Goad map. A Goad map is a detailed street map that shows individual buildings and their plots, and is usually needed for land registry and insurance purposes. They can be purchased through online sources.

The nearest neighbour analysis can be used to identify whether there are clustered land use zones within that section of the town.

The buildings have been categorised by their primary use: Retail; Financial; Municipal; Private Other; Residential (Housing), and Green Space.

1. In each category, the buildings are coded and numbered. For example, in the retail category the buildings are coded R_1, R_2, R_3 etc.



2. For each building in each category, the distance to the next nearest building in that same category is measured. Where measurements are taken between different points of the map, as in this case, the researcher must ensure there is a high degree of consistency, by always measuring from the centre of a feature or in the case of retail outlets, from the centre of the main entrance doorway. This ensures that large buildings or areas are subjected to exactly the same treatment as smaller ones. For ease in this example, only the first ten buildings of the retail category will be analysed, but to cover the full land use of the town, every category should be done in full.

Building	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
Distance to nearest neighbour (m)	8.2	4.5	6.9	9.6	9.5	12.3	23.4	18.5	5.0	12.9

3. The following calculation is used to work out the nearest neighbour value (R_n)

$$R_n = \frac{D(Obs)}{0.5\sqrt{\frac{a}{n}}}$$

where

$D(Obs)$ = the mean value of the nearest neighbour distances
a = the area sampled
n = the number of points (in this case retail buildings)

$$R_n = \frac{11.08}{0.5\sqrt{\frac{40000}{10}}} \qquad R_n = \frac{11.08}{0.5\sqrt{4000}} \qquad R_n = \frac{11.08}{0.5 \times 63.25}$$

$$R_n = \frac{11.08}{31.62} \qquad \boxed{R_n = 0.35}$$

4. Using the following scale, the researcher can then identify the extent of clustering exhibited by the category in question.

0	1	2.15
Clustered	Random	Uniform distribution

Therefore, the retail sector appears to be very clustered in the example town. However, the ten points used in the example are too few to really show an accurate picture: the researcher should aim for at least thirty to be able to say they have a significant result.

4j – A Guide to
Mann-Whitney U Test

Royal Geographical Society with IBG
Advancing geography and geographical learning

The Mann-Whitney U Test is used to analyse whether two data samples are significantly different from one another or whether any differences witnessed by the researcher are there simply due to chance.

Why would we use the Mann-Whitney U test?

Researchers who are interested in how similar two sets of data are, rather than if there is a correlation between those two sets are best using this test. It can be used when two samples are clearly independent from one another. The data should also be made up of ordinal data (data that can be placed in a clear rank order).

Worked Example:

For this worked example, the following data was gathered, showing how questionnaire participants rated the quality of service provision for two towns, Town A and Town B.
Ratings were given on a 0 to 5 scale.

Participant number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Town A	3	1	2	2	0	2	3	1	0	1	1	2	0	1	3
Town B	3	2	3	4	2	5	3	4	1	4	2	4	4	1	5

1. Place the values of both samples combined in numerical order, noting to which sample (in this case which town) each piece of data refers. If there are two of the same value, place that from Town A first in the list. It does not matter which sample you choose to be first in this case, as long as you are consistent.

A	A	A	A	A	A	A	A	B	B	A	A	A	A	B	B	B	A	A	A	B	B	B	B	B	B	B	B	B	B
0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	5	5

2. For each value for Town B, count how many Town A values come before it in the list. Add together these counts to get a U_1 value.

A	A	A	A	A	A	A	A	B	B	A	A	A	A	B	B	B	A	A	A	B	B	B	B	B	B	B	B	B	B
0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	5	5
								8	8					12	12	12				15	15	15	15	15	15	15	15	15	15

U_1 = 8 + 8 + 12 + 12 + 12 + 15 + 15 + 15 + 15 + 15 + 15 + 15 + 15 + 15 + 15

U_1 = 202

3. For each value for Town A, count how many Town B values come before it in the list. Add together these counts to get a U_2 value.

A	A	A	A	A	A	A	A	B	B	A	A	A	A	B	B	B	A	A	A	B	B	B	B	B	B	B	B	B	B
0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	5	5
0	0	0	0	0	0	0	0			2	2	2	2				5	5	5										

U_2 = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 2 + 2 + 2 + 2 + 5 + 5 + 5

U_2 = 23

4. Using a critical value table, one can tell if this result is significant or not. One compares the smaller of the U values with the critical value read from the table. The copy below can be easily found online for use with higher n_1 and n_2 values. The table below gives critical values to 5% significance.

	n_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
n_1																			
1																			
2									0	0	0	0	1	1	1	1	1	2	2
3						0	1	1	2	2	3	3	4	4	5	5	6	6	7
4					0	1	2	3	4	4	5	6	7	8	9	10	11	11	12
5				0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18
6				1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24
7				1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30
8			0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36
9			0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42
10			0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48
11			0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55
12			1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61
13			1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67
14			1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74
15			1	5	10	14	19	24	29	34	39	44	49	54	59	**64**	70	75	80
16			1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86
17			2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93
18			2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99

The size of each sample is denoted by n_1 and n_2 (in this case, the sample size for both Town A and Town B is the same). Both n_1 and n_2 are 15. This gives us a critical value of 64.

5. In order for the two samples to be thought of as significantly different from each other (and not due to chance) the smaller U value has to be equal to or less than the critical value from the table.

   In the example, the U value of 23 is less than the critical value of 64. Therefore, we can say with 95% certainty that the questionnaire respondents have rated Town A and Town B significantly differently. The next stage for the geographer is to find a reason why the respondents may have reacted this way.

4k – A Guide to
Qualitative Data Analysis

Royal Geographical Society with IBG
Advancing geography and geographical learning

It is all too easy for a researcher to fall into the trap of thinking that their qualitative data cannot be analysed in the same manner as their numerical data, or is too difficult to work with. In fact, there are a variety of **Qualitative Data Analysis Techniques** for the researcher to use, and while their methodologies may appear a little vague compared to their quantitative partners, logical stages of analysis can be applied to most forms of qualitative data.

Qualitative analysis is all about drawing out firm concepts and patterns from what at first appears to be a wealth of uncategorised behaviours and opinions. The following framework can apply to most forms of qualitative data but a good researcher will adapt it to suit the particular nature of their own data and explicitly tell the reader of their study how and why they are using a particular framework. The framework below has been applied to an interview transcript about out-of-town shopping to highlight to the reader one way of using it.

1. Transcribing

Transcribing data means making a full written record of the data. In the case of interviews, this will mean writing down word for word the responses the interviewee gave to each question and similarly for oral histories or sound files, the researcher should be prepared to note down the script used. Smartphones and tablets may be useful tools for this exercise as many devices have voice recognition apps, which allow the researcher to speak and the device to create speech into a document. Equally, there are a variety of software packages available (often with a free online trial) which can transcribe, code and index data quickly.

For data such as the research diary, observations made in the field, photographs and videos, the researcher should describe their data as impartially as they can, with detailed insight and noting separately any viewpoints drawn by the researcher themselves which are opinion based rather than factual.

Example:

> *"I used to visit the town centre quite frequently, once a day at least but when the 24-hour supermarket opened out on the Forest Road I kinda didn't need to go into town that much 'cause they sold everything: even got me work clothes there. And they were a lot cheaper too, y'know supermarket prices like. The downside was I had to drive there but when you worked it out it was alright 'cause you weren't paying for parking so I guess it was alright. Now when I do go into town it's for something special, like last week I had to get a birthday present for my niece and she wanted this particular remote control dog thing so y'know… Mum still goes in every week, but that's 'cause the buses to Forest Road don't run too frequent round her way so it's easier for her to keep going into town, plus she knows where everything is. She likes to have a coffee too in the High Street before she heads home, makes a day of it. But for me I don't have time for that 'cause of work, so for me the supermarket does the job, y'know, all in one go"*

2. Cleaning

Once the data is fully transcribed it may be necessary to 'clean' the data. This is especially true of interview transcripts which tend to be littered with asides and repeated points. Removing these from a transcript will make the data easier to work with. It is important to note that cleaning does not involve any change to the wording used, only the removal of any parts that are not to do with the nature of the interview or the removal of words that are commonly used in informal speech but are not used for more formal written reports.

Example:

> *"I used to visit the town centre quite frequently, once a day at least but when the 24-hour supermarket opened out on the Forest Road I didn't need to go into town that much 'cause they sold everything: even got me work clothes there. And they were a lot cheaper too. The downside was I had to drive there but when you worked it out it was alright 'cause you weren't paying for parking. Now when I do go into town it's for something special. Me mum still goes in every week, but that's 'cause the buses to Forest Road don't run too frequent round her way so it's easier for her to keep going into town, plus she knows where everything is. She likes to have a coffee too in the High Street before she heads home, makes a day of it. But for me I don't have time for that 'cause of work, so for me the supermarket does the job, all in one go"*

3. Identifying Character Points

A character point in a transcript is any event that clearly informs a point or brings forth a new idea. It may be a comment that identifies a variable or an issue. It is the moment when reading through a transcript that the researcher realises that they have heard something new. It takes some practice to be able to identify characters in a transcript and the researcher may find they need to read a piece a few times before they are confident they have got all the characters covered. In this way, categories for other transcripts may be created and these categories will allow the researcher to carry out further analysis at a later stage.

Example:

Character point in transcript
I used to visit the town centre quite frequently, once a day at least
when the 24-hour supermarket opened out on the Forest Road I didn't need to go into town that much 'cause they sold everything
And they were a lot cheaper too
downside was I had to drive there
you weren't paying for parking
Now when I do go into town it's for something special
mum still goes in every week
the buses to Forest Road don't run too frequent
she knows where everything is [in town]
makes a day of it [in town]
I don't have time [for enjoyment in town]
the supermarket does the job, all in one go

4. Coding

Coding is the process of attaching labels to each of the character points the researcher has identified and effectively ‘sorting’ the data. This will allow them to collate points from many different transcripts together relatively quickly and aid the identification of patterns within the data.

Example:

1. Frequency of visits to both sites
2. Quality / quantity of goods on sale at both sites
3. Economic / financial considerations
4. Advantages of town centre / disadvantages of out-of-town centre
5. Advantages of out-of-town centre / disadvantages of town centre
6. Differences in use of sites by different demographics
7. Transport issues to both sites

Section of Transcript	**Coding**
I used to visit the town centre quite frequently, once a day at least	1
when the 24-hour supermarket opened out on the Forest Road I didn’t need to go into town that much ‘cause they sold everything	1 2 5
And they were a lot cheaper too	3 5
downside was I had to drive there	4 7
you weren’t paying for parking	3 5 7
Now when I do go into town it’s for something special	1 2
mum still goes in every week	1 6
the buses to Forest Road don’t run too frequent	4 7
she knows where everything is [in town]	4
makes a day of it [in town]	4
I don’t have time [for enjoyment in town]	5
the supermarket does the job, all in one go	2 5

5. Processing

Once the data has been coded, the researcher can process it. This means that all the data from one code is brought together to see if there is consensus, to identify areas of difference and similarity and to sort theory from real experience. At this stage it is possible to see which categories are linked together across lots of different data sources. Some themes may always be discussed together and the researcher should identify any possible patterns. Key words might be identified and their frequency of use counted to identify areas of agreement: this is known as indexing. It may even be possible to convert some of your qualitative data into quantitative data and process it using numerical data analysis tools.

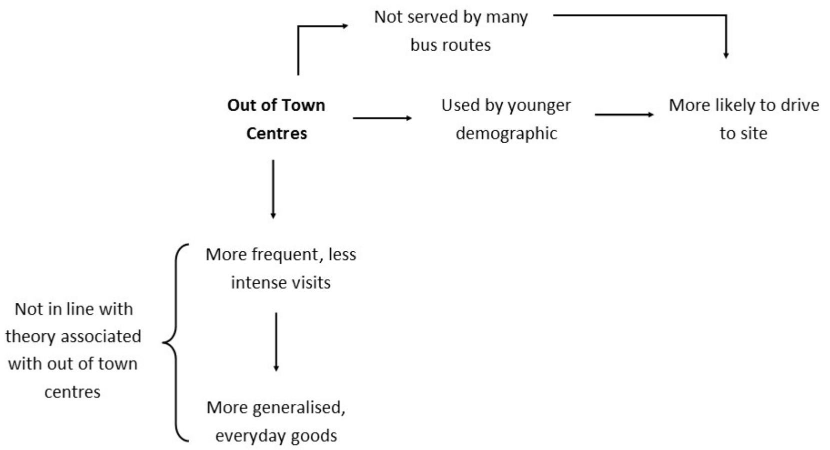
Summaries can also be drawn from the data, with each of the character points forming the basis of the summary. These summaries should include all the sources of qualitative data available.

Example:

> *"People of working age are visiting the town centre less frequently with the opening of the out of town shopping centre on Forest Road. When they do go into the town centre it tends to be for specialist items or for socialising with friends. Transport plays a key role in people deciding in which place to shop."*

Mapping the data involves drawing a mind map of how certain ideas are linked together across all the sources of data: something that can only be done once coding has occurred. Mapping can be as simple as drawing links between fact and opinion, and between descriptions and explanations within the data. More complex mapping might identify how people, places and events are linked to processes.

Example:



Drawing a matrix may also be a useful way of summarising the information one getting from the practice of coding qualitative data. Different character points are plotted against each other and the researcher uses the data to suggest whether there are degrees of conflict between the two ideas or degrees of cohesion. This works especially well when discussing different user groups.

Example:

	Local land owners	Environmental group members	Housing developers	Local renters
Environmental group members	**X X**			
Housing developers	**O**	**X X X**		
Local renters	**O**	**O O**	**O O O**	
Tourists	**O**	**O O**	**O**	**X**

X = degree of conflict **O** = degree of cohesion

6. Just as the researcher would with quantitative data, it is a good idea to have measures in place to check for the significance of any data collected. For any data source it may be possible to design a scale of credibility and a scale of neutrality, helping the researcher to see which sources of data are likely to be most reliable when drawing conclusions.

Section 5 – Conclusions

The conclusion section of the Independent Investigation is the point at which you have the chance to both **summarise** the main findings of the data analysis section and offer geographical explanations for the phenomena you are presenting. It is important at this stage to only draw on information seen in the analysis section: no new data or theories should be presented in the conclusion, as the framework for a study of this nature wants students to have already explored all of the main ideas in the previous sections. Instead the conclusion should concentrate on **answering the research questions** and, in fact, using the research questions as sub-headings within the conclusion section can be good practice for showing the reader how the study has always had these questions at its heart.

Geographical explanations for the answers to the research questions should be clearly linked to established theory where possible, though students should take care not to repeat the geographical models and theories in the detail in which they were stated in their Introduction and Literature Review section. Instead it is a good idea to link the findings of the study to the theory by showing how the investigation extends geographical knowledge or indeed confirms it within the location or setting of the study. Where the results contradict the theory the geographer is called on to look at the particular circumstances of the study and offer possible reasons why it might not match the preconceived models. There may be many reasons why the results from a study might not follow an expected pattern: from errors in the methodology to site specific factors. It is vital that you do not reject your findings simply because they do not match what you have read in a textbook – some of the best Independent Investigations are those where the results appear 'wrong' but the geographer sees the local geography as a defining factor in what made the results what they are.

The conclusion should be well sequenced and all reasoning should be logical and sound. The researcher can only conclude what their results tell them, even though it is very tempting to make cognitive 'leaps of faith' in your arguments.

Wrongly drawn conclusions:

The results show that since 2002 the seaside town in question has not experienced any coastal flooding. Groynes were built along the its coastline in 2002 and 65% of people say that the pier is now more of an eyesore than the groynes.

" The new groynes do not detract from the look of the beach." ✗

The results do not state this. When people were asked to compare the groynes with the pier, they rated the pier as being less attractive, but that is not to say the groynes are welcomed aesthetically.

" The new groynes have saved the town from flooding." ✗

The results do not state this. Though the town has not flooded there is no evidence to suggest that would have flooded if the groynes had not been built.

" The groynes have been a good investment for the town and people like them." ✗

The results do not state this. There is no indication from the statement that the groynes have given good value for money as ultimately this will always be unprovable. Nor have people been directly asked if they like the groynes or not.

Common Pitfalls:

- **Discussing data that has not been presented or analysed.** Your study should follow a logical sequence: in the conclusion there should be no new information for the reader to find.
- **Bulking out the conclusion by repeating explanations of geographical theories and models.** If you have already discussed a theory in your Literature Review simply refer back to the theory briefly.
- **Making 'leaps of faith' in your reasoning.** You cannot make a conclusion about an issue if you have not asked that question or found that answer to be true.
- **Discrediting your methodology because the results do not fit the theory.** There may be sound geographical reasons why the theory does not match up with your results. A good geographer will look at all the localised conditions to find an explanation before putting anomalous results down to flawed data collection methods.
- **Making vague links to the aims and research questions.** Remember, your research questions or hypotheses should make up the backbone of your study.
- **Claiming the investigation does more than it actually does.** It is always best to be tentative with any claims about the impact your conclusions will have on the world of geography! It is highly unlikely that your Independent Investigation will do anything more than confirm that a model in its whole, or in part, 'works' for a particular location.

Section 6 – Evaluation

Royal Geographical Society with IBG
Advancing geography and geographical learning

The **evaluation** section of the Independent Investigation is the researcher's chance to reflect purposefully on both the validity of the study and its limitations. Each stage of the investigation should be covered, from the choice of research questions to the analysis methods you have deployed. The limitations of your data collection methods are likely to form the main part of the evaluation but assessing your use of certain data presentation methods over others is just as important.

It is wrong to think that by highlighting the problems with your research, you are somehow admitting that you made mistakes and will be marked down for it. In fact, the opposite is true: most research, even at the highest academic level is imperfect. Not admitting that your study has limitations is naïve: the marker of your study will be looking to see that you recognise these problems and have tried to either resolve them or reduce their impact.

With the evaluation section appearing at the end of the study it can be tempting for students to quickly note down some possible improvements and declare the section complete, simply for wishing to finish the investigation and hand it in. However, it is worth spending some quality time on this section as it is relatively easy to gain marks in this section compared to others. Writing a **research diary** throughout the study can prove especially useful in the evaluation stage. Reading back through your field notes, your observations on your methodologies and ways that you tried to do things and failed, can provide lots of evaluative points that you might otherwise have forgotten.

Broadly, the evaluation of the different sections of the study should cover three main areas:

- The **reliability** of the study. This involves an analysis of how accurate your study is in setting out what it intended to do. You will have to look at the precision of your data collection methods, the accuracy of the way you have presented data and the significance of your results in the data analysis section.
- The **validity** of the study. This involves standing back from your study as a whole and thinking about whether it was a sound decision to study the whole of your chosen phenomena in the way you have. You might want to comment here on the level to which your study has contributed to geographical thinking.
- The **limitations** of the study. Your study was invariably limited by time, resources, location and personnel. Comment on how these made a difference to your study and how these limitations affected the possible conclusions you were able to make.

Whilst commenting on these areas, it is also worth considering the **improvements** you could make to your study. You may want to write about what you would have done differently had you had more time or different resources, as well as how you could **extend the study** and what logically you could do next if you were to continue researching. It is common practice to think about repeating the investigation at a different time of the year, or in a different location. You should also explain why you believe this might be useful and draw on the wider context of your investigation established through your reading and research.

Key words to use during an evaluation:

Don’t forget to also comment on the **strengths** of your work, as well as how you avoided pitfalls and potential problems.

The reader and marker of your study will expect to see that you have considered the potential **ethical problems** associated with carrying out your research. The evaluation is a good place to include this if you have not done so in earlier sections. This means you think about the ethical impact you, as a researcher, have had on people and the environment within your study, as well as the impact ethically that your study and the nature of your research has had more broadly.

Common Pitfalls:

- **Pretending that everything in your study is perfect.** No piece of research is without its flaws – be honest about what you would change and the inherent problems of researching your chosen topic.
- **Only focusing on the limitations of the data collection methods.** The whole of your study should come under analysis in this section, and even the wording of your research questions should be evaluated.
- **Having a fairly weak level of reflection.** Simply saying that questionnaire results were subject to human bias, that your sample size wasn’t big enough or particular circumstances limited your access to alternative sites or wider data collection is not enough for the reader to be convinced that you have really thought deeply about your evaluation.

Section 7 – Final checks

Royal Geographical Society with IBG

Advancing geography and geographical learning

Just before you submit your Independent Investigation for marking, it is worth carrying out some final checks to make sure your report is well organised. Some students still make the mistake of focusing too much of their effort on making their work look good rather than geographically sound in content.

A contents page is a useful tool for organising the study in the right order. The reader should be able to follow the framework of your thinking about the study and the whole focus of your study should be centred on your research questions or hypotheses. Each section should begin on a new page and have a clear title. Try not to cram too much on each page – give graphs and tables the space they need and check with your teacher about line spacing. Use the same font and font size throughout the main study text and be consistent in the way one page looks to the next.

Everything in your study which is not text, such as graphics, photographs, presentations of data and tables should be assigned a figure number. This is normally written just below each graphic and they should run in numerical order from the start of the study to the end. Figure numbers make it useful to refer to a graphic when it does not appear on the same page as the text in which it is referred. Use automatic figure numbering, rather than adding details manually.

Double check that every piece of geographical information you have talked about in the text of your study has a citation or reference. Then make sure that after your evaluation section, you have included a bibliography. If you decide to include an appendix, it should be the last part of the study. It should only contain things that you feel may aid the reader but are not essential to the understanding of the study – the test is that the reader could easily mark your study without it.

While checking your study for spelling, punctuation and grammatical errors (do not rely entirely on a word processing package to do this for you), it may be worth checking your use of geographical language. The marker will expect to see geographical terms used frequently and accurately throughout your study.

Always check that your submission conforms to the latest guidelines and requirements of the relevant awarding body.

Further information:

- You can join the Society as a Young Geographer and include this on your personal statement. For more information visit **www.rgs.org/joinus**
- For more information about studying geography at university visit **www.rgs.org/studygeography**
- For additional case studies, podcasts and online lectures to support your A Level studies visit **www.rgs.org/schools**