Sage Research Methods Data and Research Literacy: How-to Guide

Authors: **Please complete the white fields below.** Direct any questions to your editorial contact.

Title. ▪ *Maximum of 20 words.* ▪ *All principal words capitalized.*		*How to Analyse Data and Evidence: The Third Stage Of The Social Research Toolbox, QGAP*
Authors. ▪ *The order of authorship in the publication will follow the order below.* ▪ *Please add additional rows for co-authors if necessary.*		Please specify the total number of authors: **1**
1	Name	Dr Katy Wheeler
	Author email	Katy.wheeler@essex.ac.uk
	Affiliation, country	University of Essex, Department of Sociology & Criminology
	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		[Maximum of 200 words] *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☐ *I have not published with Sage before* ☐
2	Name	
	Author email	[Insert contact email]
	Affiliation, country	[Insert institution and country]
	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		Dr Katy Wheeler is a Senior Lecturer in Sociology at The University of Essex. Her research interests are in the fields of sustainability and consumption, and she has published widely within leading academic journals and two authored monographs, *Fair-Trade and the Citizen-Consumer: Shopping for Justice?* (Palgrave: 2012) and *Recycling and Consumption Work: Social and moral economies* (Palgrave: 2015). She has also co-authored the textbook *How to do Qualitative Interviewing* (SAGE, 2022 – with Morgan Brett). She has taught social science methodology courses for many years (both qualitative and quantitative), from introductory undergraduate to advanced postgraduate. *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☒ *I have not published with Sage before* ☐
3	Name	
	Author email	[Insert contact email]
	Affiliation, country	[Insert institution and country]

	SAGE Author ID	[office use only]
Author biography. *Bios will not be copy-edited; please ensure they are correct.*		[Maximum of 200 words] *I have published with Sage and want to use my existing bio* ☐ *I have published with Sage and want to update my bio* ☐ *I have not published with Sage before* ☐
For which **student level** is this guide most suitable?		**Introductory Undergraduate**
Methodology categorization		**Other / Not Applicable**
Discipline		Guides should be **cross-disciplinary**. If the content of your guide is specific to a discipline(s), please let your editorial contact know.
Series Editor. *For office use only:*		
	Series Editor Name	
	SAGE Author ID	[office use only]

Sage Research Methods: Data and Research Literacy is an educational resource which will equip undergraduate and postgraduate students and researchers with the tools to **understand and critically evaluate research methods and methodologies**, **manage and interpret data**, and **conduct robust social research with integrity** and confidence.

Guides will be **authoritative and accessible** resources which **combine research principles with research practise**, incorporating practical and ethical considerations, to help prepare students and researchers for working with data, evaluating research, and conducting their own research.

When writing your guide, we recommend using **real-world research examples** to keep the reader engaged. You may choose to use one consistent example throughout the guide, or multiple examples.

Each how-to guide is limited to **4000 words**, with a 10% leeway. For topics which require more than 4000 words there may be the option to write multiple guides; please raise this with your editorial contact if required. Guides may include direction to further resources through which the reader can explore each topic in more depth.

You can view two how-to guides from previously published collections here:

- From *Diversifying and Decolonizing Research*

- From *Doing Research Online*

Please ensure you have read the **manuscript guidelines** before you begin writing your guide.

SRM requires the disclosure of any **AI-generated content** in your work. Please read our full policy here and notify your editorial contact if you have any questions, or if you need to disclose use of generative AI in your work.

References should conform to American Psychological Association (APA) style, 7th edition, and should contain the digital object identifier (DOI) where available. Sage will not accept guides that are incorrectly referenced; please ensure accuracy before submission. For help on reference styling see https://apastyle.apa.org/style-grammar-guidelines.

Abstract

The abstract should be a concise summary of your how-to guide. What aspect of the research process, working with data, or specific methodological and practical challenges will your guide address? It should be succinct and enticing, and should incorporate key words and concepts discussed in the body of the text. Please do not cite references within the abstract.

[Insert here: Maximum of 250 words]

This guide offers an overview of some of the key tasks required for analysing qualitative and quantitative research data and evidence. As part of the social research toolbox, or QGAP, series, it considers what should happen once you have gathered data and evidence in pursuit of your research question. The guide begins with the important task of preparing your data for analysis which involves formatting it so that it is free from errors and that all identifiable features (which could compromise anonymity) are removed. It then considers the key features of social science data analysis – which are argued to be categorisation and comparison. Theory also plays an important role in this process as comparisons within our data are used to test, refine or develop theories. We finally discuss the use of software packages in the analysis stage and some of the possibilities they offer for managing, categorising and visualising relationships within your data.

Learning Outcomes

Learning outcomes must explain what the reader will learn from reading your guide. How will the reader be able to apply what they have learned to their own research practice?

Consider what the **most important aspects of this topic** are. Bear in mind the guide is limited to 4000 words. **The content and structure of your guide should explicitly correspond with these learning outcomes**.

See the links below for guidance on writing effective learning outcomes:

- *Writing learning outcomes*

- *Blooms Taxonomy Action Verbs*

Insert 3–5 learning outcomes, **beginning with an action verb**, completing this statement:

Having read this guide, readers should be able to . . .

- Prepare their data for analysis by removing errors and identifying information.
- Understand the central role of categorization and comparison in social science data analysis and explain how these processes are used to test, refine, and develop theories.
- Recognise how software tools can assist the analysis stage.

Introduction

Build on the abstract to further describe what methodological issues will be discussed in this guide; what the student reader will gain from reading the guide; how the guide will be structured; which real-life research examples will be drawn upon, etc. You may wish to begin with a brief positionality statement.

This guide forms part of the social research toolbox or, QGAP, series, which offers a simple way of conceptualising the research process. You can learn about QGAP by watching this short animation [hyperlink to be added once available], and instructors can download some teaching PowerPoints that support the series here [hyperlink to be added]. Within the toolbox, there are four stages; 1) Questions, 2) Gathering data and evidence, 3) Analysing the data and evidence and 4) Presenting answers to your research questions. The four stages occur in all social research projects (whether qualitative, quantitative or mixed-method), and though they have been presented in a linear way for teaching purposes, the reality is that each stage is shaped by the others, and it is usual to move between stages at different points of the research journey, depending upon the type of research you are doing.

In this guide, the focus is on the 'Analysis' stage of the social research toolbox. Once you have gathered your data /evidence (G) to address your research question (Q), you then need to analyse this (A) – e.g. systematically explore the data and evidence to reveal patterns, trends and meaningful insights – before presenting your findings (P). As with the other parts of the toolbox, the analysis stage is closely tied to the other stages and often analysis does happen at the data gathering stage (in terms of the design of the research collection tool) and it can be implied by the form of the research question (e.g. whether a quantitative or qualitative question or one based on primary or secondary data). Through analysis the aim is to develop answers to your research question therefore the form of that research question and the types of data it has generated will guide our analytical strategy. But the key tasks of analysis (regardless of data type) are to categorise your data into a form that enables you to compare different observations, spot relationships between them and generate insights that are linked to, or generative of, theory. Before considering the purpose of the analysis stage, the guide begins with the practical step of preparing and formatting your data and evidence for analysis – including the need to deal with data entry errors and for anonymisation. You may decide to use computer software to store your data and to explore the relationships within it, so the guide closes with a discussion of how software can enhance analysis through its embedded tools and functions.

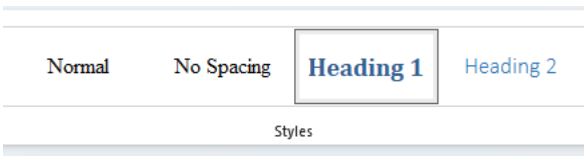
Main Body of Text

Write the body of your guide below. The text should be between **2,000 - 4,000 words.**

We encourage the use of **headings** and sub-headings to **structure your guide into sections.**

We recommend using **800 words or fewer for each section**.

For section headings please use Word Style 'Heading 1'. For any sub-headings within sections use Word Style 'Heading 2'. To use Word styles in Microsoft Word, select the text you want to format, click the "Home" tab and then use the "Styles" pane.



Every section must be followed by a **Section Summary.**

Each Section Summary should consist of 3-5 bullet points, written out as full sentences, which summarize the key information in the section.

Preparing your data for analysis

The data gathering stage can produce lots of different types of data (spreadsheets, audio recordings, photos, notes and documents) which will need to be organized and formatted before you will be able to systematically analyse them. Though this may seem like a practical task unrelated to analysis, decisions made in this phase – such as what is relevant and where to store it – will follow you throughout the analysis stage. You might decide to store things in separate files and folders or use a software program to place all files in one overarching database (more on this later in the guide). If you had to write a data management plan when you submitted your project for ethical approval, you will be obliged to follow these procedures (to learn more about data management plans and best practice see Corti et al., 2020). In this section, we focus on two important tasks when preparing data for analysis – cleaning to remove errors and anonymisation.

Cleaning the data

Data cleaning procedures depend on the type of data you have, and the instrument used to collect it. If you have an audio recording of an interview, transcription is required to convert it into a textual form. How you do this will depend on the sensitivity of the data you have collected, the quality of the recording you have and the analytical purpose your transcription needs to fulfil. Some data might be for your ears only so using a service for transcription might compromise the confidentiality you promised – another reminder to consider this stage of the toolbox when negotiating consent to gather data. On recording quality, I remember doing an interview next to a coffee machine in a busy café and the resulting recording was very difficult to hear. This was in the days before automated transcription, but I don't think these services could have produced an accurate transcript for this interview and my own transcription had lots of [inaudible] parts, which affected the quality of that data. Recording equipment has advanced over time, as has software to process audio recordings (such as Otter.ai and online conferencing tools which can produce a reasonable transcript), but the validity and reliability of these transcripts must be carefully checked (see Brinkmann & Kvale, 2015 for more on validity and reliability of transcription). I would recommend listening back to the original audio recording alongside your transcript (whether produced by automated transcription or a human transcriber) to check for accuracy. This performs the dual function of cleaning your data for errors, as well as re-immersing you into that interview to aid the familiarization needed for qualitative analysis. There are different styles of transcription and choices that need to be made about what is transcribed which usually relate to how you intend to use the data. Verbatim transcription records all the 'errs' and 'umms' as well as sentence stops, whereas intelligent verbatim removes these features so reads more like a written than an oral format (see Morgan Brett & Wheeler, 2022 for a discussion of transcription styles).

Alternatively, you may have survey data which has either been collected by hand or has been entered into a computer-assisted platform by the participant themselves or a survey interviewer. If your survey data was collected by hand, it will need to be manually entered into a digital format, which can lead to errors due to poor handwriting or data entry mistakes. You may have funds to employ two people to enter the data separately and then compare, but if not, checks can be done to ensure the data is as free from error as possible. Even if data was entered using a computer platform, it may still contain errors or skipped questions. Sometimes errors can be spotted because answers across several questions are not internally coherent – for example when I recently analysed a survey which asked separately for age and level of education, I found one participant had entered their age as 12 and also stated they had a degree-level qualification. I looked at the data from their survey responses to help me to determine which answer seemed most likely. You may make the decision to delete both answers to create missing data excluded from analysis. If you have lots of missing data, perhaps because a participant has left the survey after only a few questions, a decision needs to be made about whether this indicates withdrawal. The consequence of withdrawal will directly relate to the terms upon which you negotiated consent with the participant. If consent for data contributed was stated to be *up to the point of* withdrawal, what was contributed can be used but if withdrawal was possible *at any point* with no details on data already contributed, then you may need to delete that entry completely.

Cleaning data to remove errors is important for all types of data, and I recommend recording the changes made in a project journal to keep track of these decisions. Your changes should be about increasing the validity of the data so that you can trust your findings as you move through the analysis process.

Anonymising the data

Another key consideration when preparing your data for analysis is the removal of identifying information. Most participants are offered anonymity in social research but when data is first gathered – e.g. raw data – it will still contain information that could reveal identities. Identities can be revealed through direct identifiers – such as names, address, postcode, telephone numbers and pictures – and indirect identifiers – information like workplace or occupation which if revealed *in combination* with other information like socio-demographic characteristics might identify a person (Corti et al., 2020). It is common practice to remove all direct identifiers (such as using pseudonyms instead of real names) and store these separately from the rest of your data (perhaps in a secure folder or locked filing cabinet). But do also think about other identifying information. We often want to keep the context, so it is better to replace rather than remove entirely. As an example, suppose a Sociology Professor from Essex University was interviewed - rather than provide those details of status, workplace and subject, they could be described as a 'UK Higher Education academic' to retain the context but widen the potential pool of participants. With quantitative data, aggregate precise values of variables to prevent identification (e.g. use year of birth rather than date of birth, report data in categories rather than as discrete values). If you have postcode or geospatial data, you might want to transform that data into an alternative variable such as deprivation index position, population density or a larger area coordinate. If removing the precise data will affect your research question, it might be that you need to secure access provisions for its use, so that analysis is carefully controlled and results produced checked to ensure confidentiality is maintained.

Section Summary

- *It is important to organise the various the types of data gathered before you start analysing it. This initial organisation serves the dual purpose of familiarising yourself with your data and evidence.*
- *You should 'clean' your data to minimize errors – such as data entry errors created by transcription processes or manual entry validation of survey data.*
- *Remove all direct identifiers and modify indirect identifiers (as necessary) from raw data to protect participant's confidentiality whilst maintaining useful context.*
- *All changes and choices made at the stage of data preparation should be documented in your research journal to track key decisions.*

What is the purpose of your analysis?

Once your data and evidence are organized, and in a form ready for analysis, your task is then to thoroughly familiarize yourself with their contents to identify elements that will be of interest to your research objectives. The word 'analysis' has its origin in Ancient Greek language where it meant 'breaking up' or 'taking apart'. Analysis will always involve breaking down data and evidence into its constituent parts to understand it better and to systematically identify patterns and trends. This is then paired with 'synthesis' (also a Greek word) which means bringing those parts back together to make connections and draw out meaningful insights and conclusions that are relevant to your research question. Most analysis (regardless of theoretical and methodological approach) therefore involves the processes of categorisation and comparison. How these processes are realized vary depending on the type of data you have and the goal of your analysis – whether it is to generate new theories and insights or to test existing theories, or to add to an evidence base to inform policy decisions. This section is broken into two main parts – the first describes how researchers categorise and compare qualitative and quantitative data, whilst the second, considers the role of theory within the stage of analysis. Here is not the place to give detailed step-by-step guidance but I do point you in the direction of some sources to learn more.

Categorisation and Comparison

Comparison forms the core of social science analysis – whether that is comparison within a case or across groups of cases, comparison is a necessary precondition to understand the different elements within your data. To systematically compare these different elements, you need to transform (or break up) your data into categories or chunks, or more scientifically 'codes', which will enable you to view your data according to key features of interest. The process of coding your data does differ between qualitative and quantitative projects but essentially the purpose of this coding remains the same.

In qualitative research, you will have texts (interview transcripts, fieldnotes, newspaper articles, videos, photos or other documents) which need to be sorted into meaningful units of analysis. There are different approaches to analysing qualitative data, from thematic analysis (Braun & Clarke, 2006, 2022) to qualitative content analysis (Hsieh & Shannon, 2005; Schreier, 2012), to grounded theory (Charmaz, 2014; Glaser & Strauss, 2017), to discourse analysis (Fairclough, 2009; Keller, 2013; Taylor, 2013), to visual analysis (Pauwels, 2019; Rose, 2001) and narrative analysis (Holstein & Gubrium, 2012; Riessman, 2008) to name a few (and suggest references so you can learn more). Whilst there are differences between these approaches and your choice of approach will be shaped by your research question, they all share the requirement of categorising data and then comparing these categories to reach an interpretive understanding. This involves carefully reading and immersing yourself in the data and either/both developing a coding frame from that immersion (bottom-up) or/and applying a set of criteria (top-down) to better understand what the text *says* (about social worlds beyond the data) and/or what the text *does* (use of language/visuals, construction of talk or narrative, symbolism, key messages within the data). This will usually involve chunking the data and labelling it according to key analytical concepts – these are often called 'codes' (but terminology varies across approaches) - and as analysis develops comparing how these concepts are applied and used between your units of analysis (e.g. transcripts, participants, texts, visual material) alongside your interpretive reflections and notes – these are often called 'memos'.

In quantitative research, decisions about categorising data are usually made at the data entry stage (or much earlier at the questionnaire design stage) and involve converting raw data into numerical categories for statistical analysis. In developing your data collection tool or gathering secondary evidence, you will have thought carefully about the concepts that need to be measured. In order to perform basic through to advanced statistical analysis, you need to create a dataset that comprises a series of variables – a variable is 'an image, perception or concept that is capable of measurement' (Kumar, 2014, p. 386). There are different ways of coding your data into variables depending on the unit of measurement. There are four main types of scale; 1) nominal variables categorise data without any order (e.g. different categories like ethnic group or occupation); 2) ordinal variables categorise ordered data with no consistent intervals between these categories (e.g. level of education, Likert scales); 3) interval variables categorise data that have consistent intervals but no true zero point (e.g. IQ scores, temperature); 4) ratio variables categorise data with consistent intervals and a true zero point (e.g. age, income). Different types of variable demand different statistical methods of analysis. For example, ordinal data, such as Likert scales, can be analysed using non-parametric tests (like chi-square or Spearman's R correlations) or ordinal regression, while interval and ratio data often allow for more advanced techniques like linear regression (see Field, 2024 for a good overview of what forms of analysis are appropriate for your data). A codebook needs to be developed to record how variables have been measured and to provide a key for what the numbers represent. Once data has been categorised, the different categories can then be compared for descriptive analysis within variables (frequencies of categories, measures of central tendency and dispersion) and more complex bivariate and multivariate analysis across variables (crosstabs, correlations, t-tests, regression) (see Kumar, 2014; Williams et al., 2021 for a good introductory overview). The purpose of this analysis is usually to identify differences, relationships, patterns and trends within our data which can be inferred to the general population from which our sample was drawn (see MacInnes, 2022 for an introduction to inferential statistics). This leads us to the role of theory in analysis, which for quantitative approaches is often about testing the probability that our hypotheses (*a priori* theories) about relationships between variables are false (the hull hypothesis)– which can only be achieved by comparing categories.

In mixed methods research, data analysis involves applying standard qualitative and quantitative techniques of categorisation and comparison, but the key difference lies in the integration of these findings. For a convergent design, such as in my study of Fairtrade consumption, qualitative and quantitative data were analysed separately and then compared to draw complementary insights. I analysed a national attitude survey to explore how individuals rated the effectiveness of Fairtrade consumption compared to other individual actions, while qualitative interviews and focus groups provided deeper context to these perceptions. In this case, the integration and comparison of results revealed that Fairtrade consumption is often perceived as a valuable action but not a substitute for organized political efforts to address poverty (Wheeler, 2012). In addition to convergent designs, sequential designs use one type of data to inform the collection or analysis of the other – such as using the findings from a focus group to craft the questions in a survey – with integration of analysis techniques occurring progressively. Embedded designs involve nesting one type of data within the other – such as selecting a sub-sample of participants from a quantitative survey to interview so as to better contextualise the patterns and trends observed – enabling close comparison between different method-types and data (see Creswell, 2013 for a discussion of mixed-method designs).

The role of theory in analysis

All analysis is driven, at least in part, by theoretical assumptions drawn from the existing evidence-base. No researcher gathers data without at least some idea about what social phenomena is interesting and what mechanisms might explain why things happen in the way they do. A literature review is precisely about learning how your research question fits within an existing body of evidence and how your research might contribute. When it comes to data analysis, the insights gained through exploring existing evidence will help you to spot links within your data, to connect your findings to those of others and to generalise (often in a limited way) from your data to broader populations or processes. This moves us to the synthesising stage of the analysis process which is about bringing our analysis together to make connections with theoretical frameworks and draw out relevant conclusions. Theory is not an add-on but integral to all social science research. As we touched on in the first guide in the series, research questions come from different epistemological and ontological orientations which shape research design, data gathering and analysis (Wheeler, 2025). Research questions from the interpretivist tradition are more likely to be exploratory in their nature and so generate will theory through data analysis (sometimes referred to as an inductive approach) (Braun & Clarke, 2013; Ritchie et al., 2014). Whereas research questions from a positivist tradition are likely to start with a theoretical framework from which hypotheses about relationships between measurable phenomena are developed so that data is gathered and analysed to test these theories (sometimes referred to as a deductive approach) (Kumar, 2014; Williams et al., 2021). Though it is common to hear that theory either plays a deductive or inductive role, it is rarely as simple as this in practice (Clark et al., 2021). What is important is that researchers pay attention to the theories and assumptions which have shaped how data has been gathered and in so doing acknowledge the synergies between theory and research findings. Theory is critical for understanding and drawing conclusions from data.

Section Summary

- *The key tasks associated with data analysis are categorisation and comparison. These processes break down data into smaller parts so the researcher can identify patterns, trends and relationships relevant to the research question.*
- *Categorisation and comparison are essential for both qualitative and quantitative projects but how these processes are achieved does differ depending on data type.*
- *Theory is integral to all stages of the social research toolbox and helps us to connect research findings to broader frameworks and existing knowledge.*

Using computer software to aid analysis

In this section, we consider some of the advantages and limitations of using software to assist the analysis of your data. Most researchers will use computers to store research data and write up their analysis – this may seem obvious, but it was not always the case and norms around the safe handling of participant data have shifted as technology has evolved. I am old enough (just) to remember a time when interviews were stored on tapes and reports written on typewriters. Some of your data may be in a paper format, such as hand-filled questionnaires or signed consent forms (and we discussed above the need to digitise survey responses and mitigate against data entry errors). Increasingly data is digital, and it is produced in all sorts of spaces, not always for research purposes but because of how company infrastructures operate online – and there is a growing literature on using this sort of data for social science research (Castellani & Rajaram, 2021). Technology is not a deterministic force, but it does shape what is possible, influencing how data analysis is performed.

Before the availability of packages like SPSS, STATA, R and Excel, advanced statistical analysis was less possible, or it required mathematical knowledge that took a great deal of time to perform. Nowadays, social researchers with quantitative data will use statistics software and computers as standard. Undergraduate training across the social sciences will require students to be familiar with at least one of these programmes. The same is not yet true for qualitative software or CAQDAS (Computer Assisted Qualitative Data Analysis Software) packages, like MAXQDA, NVivo and ATLAS.ti. There is both a skills gap here and a reluctance amongst some qualitative researchers to embrace the opportunities of these packages – which are many and evolving (Silver, 2023; Silver & Woolf, 2015).

Packages for both types of data enable data management, categorisation and comparison features, as well as visualisation tools. Being able to store all observations and different data types/files is one of their key benefits. You can also keep a record of decisions made about coding data and the procedures undertaken (in project logs, memos or by keeping the syntax of commands) – and your project file could be the place where you store your project journal which records key decisions throughout the research journey. There are also opportunities to explore relationships within your data and to use visual tools to display those relationships, which is very useful when it comes to presenting your findings at the last stage of the social research toolbox.

The drawbacks of software relate to the learning required to effectively use the tools within these programmes. It is often a steep learning curve and from experience teaching both qualitative and quantitative software to undergraduates, many students feel they would rather not invest this time. Though timetabled sessions introduce students to software, the real learning happens through repeated practical application and trial and error. The cost of software packages can also be a barrier as institutions may not subscribe to preferred packages – though freeware like R and RQDA is available, and standard operating packages like Word and Excel can performs some analysis functions. From a critical point of view, software does embed a range of functions and promotes a particular way of 'doing analysis' – just because something can be done does not necessarily mean that it ought to be. Whilst statistical analysis is often a process of 'following the rules', qualitative analysis is usually more flexible and iterative which is one of the reasons some qualitative researchers are reluctant to use it. Qualitative analysis can be, and often is, done without specialist analysis software. Though Silver & Woolf's (2020) five-level-QDA approach offers tips on how to harness CAQDAS for purposeful qualitative analysis.

In sum, software offers lots of opportunities for assisting and enhancing data analysis and can be used to manage processes of categorisation and comparison. But it should be used strategically in the pursuit of research objectives rather than researchers being led by its capabilities. This is especially pertinent in an era of Generative Artificial Intelligence which has been embedded into some CAQDAS packages and can be used to help assist with the writing of code for statistical analysis. These tools cannot do the analysis for you, but they can be harnessed to enhance your processes.

Section Summary

- *Computer software has transformed the possibilities for data analysis with tools for managing, categorising and comparing both qualitative and quantitative data.*
- *Whilst software use is common for quantitative data analysis, its use in qualitative analysis projects varies.*
- *Though software can enhance the data analysis, these tools should be used strategically rather than allowing software functions to dictate the research process.*

Conclusion

Includes a **summary of the key lessons** discussed within each section of your guide.

What can readers learn from this guide and apply when conducting their own research and evaluating the research of others?

In this guide, we have considered the third stage of the social research toolbox and some of the key tasks to be executed when 'Analysing the data and evidence' (A). Whilst it can be tempting to think about analysis as something that happens once data has been gathered, as with the other stages of the toolbox, it is informed by the earlier stages (research question formulation (Q) and research design choices (G)) and shaped by the intended audience for the presentation of findings (P). The guide began by considering the importance of preparing your data and evidence for analysis. We discussed removing data entry errors and direct and indirect identifiers to ensure confidentiality. The mechanics of how to do data analysis will depend on the type of data you have and the approach to analysis you are taking, but the key tasks in all forms of analysis will be to categorise your data and evidence and to make comparisons. These two features form the core of data analysis processes and will be applied in tandem with theoretical frameworks to draw out key findings and conclusions that are pertinent to your research question. The guide concluded by discussing how computer software can assist the analysis of both qualitative and quantitative data, highlighting some of the opportunities it offers for data management, categorisation and comparison. It was pointed out that software can take time to learn, and its use should be strategic, so you are research objective-led rather than allowing software to dictate the research process. The next guide turns to the final (maybe) stage of your research toolbox – presenting your findings - and it considers how we communicate our answers to our research question to different audiences.

Multiple Choice Quiz Questions

Multiple Choice Quiz Questions should:

- Test readers' understanding of your guide.

- Focus on relevant aspects of data and research literacy.

- Not require any information that is not included in this guide.

Multiple Choice Quiz Questions should not:

- Include 'all of the above' or 'none of the above' options, or implausible responses.

- Require information not included in the guide.

Example:

1. What is critical reflexivity?

a. An understanding of how a researcher relates to and actively engages with the complex contexts and dynamics within which the research is embedded. ***[CORRECT]***

b. An understanding of how over-researched populations can experience research fatigue when directly engaged by researchers.

c. An understanding of anonymity and confidentiality in research.

Guidance for writing MCQs can be accessed using these links:

- *Tips for writing effective multiple-choice questions*

- *The process of writing a multiple-choice question*

[Insert three to five multiple choice quiz questions below. **Each MCQ must have three possible answers (A, B, or C), with one correct answer.** Please indicate the correct answer by writing [CORRECT] after the relevant answer.]

1. Which of the following steps are required when preparing data for analysis?
   a) Choosing a research topic
   b) Correcting data errors and removing confidential information [CORRECT]
   c) Writing the final report

2. Which of the following is a key advantage of using software for qualitative data analysis?
   a) It completely automates coding, thus eliminating the need for researcher involvement.
   b) It simplifies the data gathering stage of the research process.
   c) It allows researchers to store, categorize, and visualize data, aiding in systematic comparison [CORRECT]

3. In social science data analysis, what is the primary purpose of categorization?
   a) To group data into manageable units to aid the identification of patterns and relationships [CORRECT]
   b) To eliminate any duplicate information that might affect results
   c) To protect participant anonymity by removing identifiable information

Further Reading

Please ensure that the recommended readings, web resources, and cited references in the guide are inclusive, and represent a diversity of people. Given our global readership, we aim for content that allows individuals with a broad range of perspectives to see themselves reflected in our published resources.

[Insert list of up to six further readings here]

- Corti, L., Van den Eynden, V., Bishop, L., Woollard, M., Haaker, M., & Summers, S. (2020). *Managing and sharing research data: A guide to good practice* (2nd edition). SAGE.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th edition). Sage.
- Kumar, R. (2014) *Research Methodology: a step-by-step guide for beginners* (4th Edition), SAGE (good introduction to quantitative approaches)
- Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE Publications Ltd. https://doi.org/10.4135/9781473906907
- Thomas, G. (2017). *How to do your research project: A guide for students*. SAGE Publications.

Web Resources

[Insert links to up to six relevant web resources here]

- CAQDAS Chats with Christina, Podcast series, available at https://open.spotify.com/show/28usVeqag9q7irrrAgJTBh?si=7e44fcf1362c441d (accessed 28/07/2024)
- Getting Help with R - The R Project for Statistical Computing, available at https://www.r-project.org/help.html (accessed 3/10/2024)
- UCLA Advanced Research Computing Statistical Methods and Data Analytics, ‘Annotated Output’, available at https://stats.oarc.ucla.edu/other/annotatedoutput/ (accessed 3/10/2024)
- Qualitative Data Analysis Services, Blog posts, available at https://www.qdas.co.uk/blog (accessed 29/07/2024)

References

[Insert bibliography of references cited in text here]

References should conform to American Psychological Association (APA) style, 7th edition, and should contain the digital object identifier (DOI) where available. Sage will not accept guides that are incorrectly referenced. Please ensure accuracy before submission. For help on reference styling see https://apastyle.apa.org/style-grammar-guidelines.

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
- Braun, V., & Clarke, V. (2013). *Successful Qualitative Research: A practical guide for beginners*. Sage Publications.
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.
- Brinkmann, S., & Kvale, S. (2015). *InterViews: Learning the Craft of Qualitative Research Interviewing* (Third Edition). Sage.
- Castellani, B. C., & Rajaram, R. (2021). *Big Data Mining and Complexity*. SAGE Publications Ltd. https://doi.org/10.4135/9781529682755
- Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.
- Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). *Bryman's social research methods*. / Tom Clark, Liam Foster, Luke Sloan, Alan Bryman ; editorial advisor: Elena Vacchelli. Oxford, United Kingdom: (Sixth edition). Oxford University Press.
- Corti, L., Van den Eynden, V., Bishop, L., Woollard, M., Haaker, M., & Summers, S. (2020). *Managing and sharing research data: A guide to good practice* (2nd edition). SAGE.
- Creswell, J. (2013). *Research Design: Qualitative, quantitative and mixed methods approaches*. Sage.
- Fairclough, N. (2009). *Discourse and social change* (Reprinted). Polity Press.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th edition). Sage.
- Glaser, B. G., & Strauss, A. L. (2017). *The discovery of grounded theory: Strategies for qualitative research*. Routledge.
- Holstein, J., & Gubrium, J. (2012). *Varieties of Narrative Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781506335117
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687
- Keller, R. (2013). *Doing Discourse Research: An Introduction for Social Scientists*. SAGE Publications Ltd. https://doi.org/10.4135/9781473957640
- Kumar, R. (2014). *Research methodology: A step-by-step guide for beginners* (Fourth edition). Sage.
- MacInnes, J. (2022). *Statistical Inference and Probability*. SAGE Publications Ltd. https://doi.org/10.4135/9781529682748
- Morgan Brett, B., & Wheeler, K. (2022). *How to do Qualitative Interviewing*. Sage.

- Pauwels, L. (with Mannay, D.). (2019). *The SAGE Handbook of Visual Research Methods* (2nd ed). SAGE Publications.
- Riessman, C. K. (2008). *Narrative methods for the human sciences*. Sage Publ.
- Ritchie, J., Lewis, J., McNaughton Nicholls, C., & Ormston, R. (Eds.). (2014). *Qualitative research practice: A guide for social science students and researchers* (2. ed). Sage.
- Rose, G. (2001). *Visual methodologies: An introduction to the interpretation of visual materials*. Sage.
- Schreier, M. (2012). *Qualitative Content Analysis in Practice*. Sage.
- Silver, C. (2023, May 5). What's a foot in the Qualitative AI space? *Qualitative Data Analysis Services*. https://www.qdas.co.uk/post/what-s-a-foot-in-the-qualitative-ai-space, (accessed 23/07/2024)
- Silver, C., & Woolf, N. (2020). Five-Level QDA Method. In P. Atkinson, A. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE Research Methods Foundations*. SAGE Publications Ltd. https://doi.org/10.4135/9781526421036818833
- Silver, C., & Woolf, N. H. (2015). From guided-instruction to facilitation of learning: The development of Five-level QDA as a CAQDAS pedagogy that explicates the practices of expert users. *International Journal of Social Research Methodology*, 18(5), 527–543. https://doi.org/10.1080/13645579.2015.1062626
- Taylor, S. (2013). *What is discourse analysis?* Bloomsbury Academic.
- Wheeler, K. (2012). *Fair Trade and the Citizen-Consumer: Shopping for Justice*? Palgrave Macmillan. https://doi.org/10.1057/9781137283672
- Wheeler, K. (2025) 'How to develop research questions: the first stage of the social research toolbox, QGAP'
- Williams, M., Wiggins, R., & Vogt, P. R. (2021). *Beginning Quantitative Research*. SAGE Publications Ltd. https://doi.org/10.4135/9781529682809