

RESEARCH METHODS FOR DATA SCIENCE

Assoc. Prof. Dr. Salha Alzahrani
Department of Computer Science

Step 3: Constructing an instrument for data collection



My Notes

~Primary
Sources

~Secondary
Sources

Major approaches to information gathering

There are two major approaches to gathering information about a situation, person, problem or phenomenon. When you undertake a research study, in most situations, you need to collect the required information; however, sometimes the information required is already available and need only be extracted. Based upon these broad approaches to information gathering, data can be categorised as:

- primary data;
- secondary data.



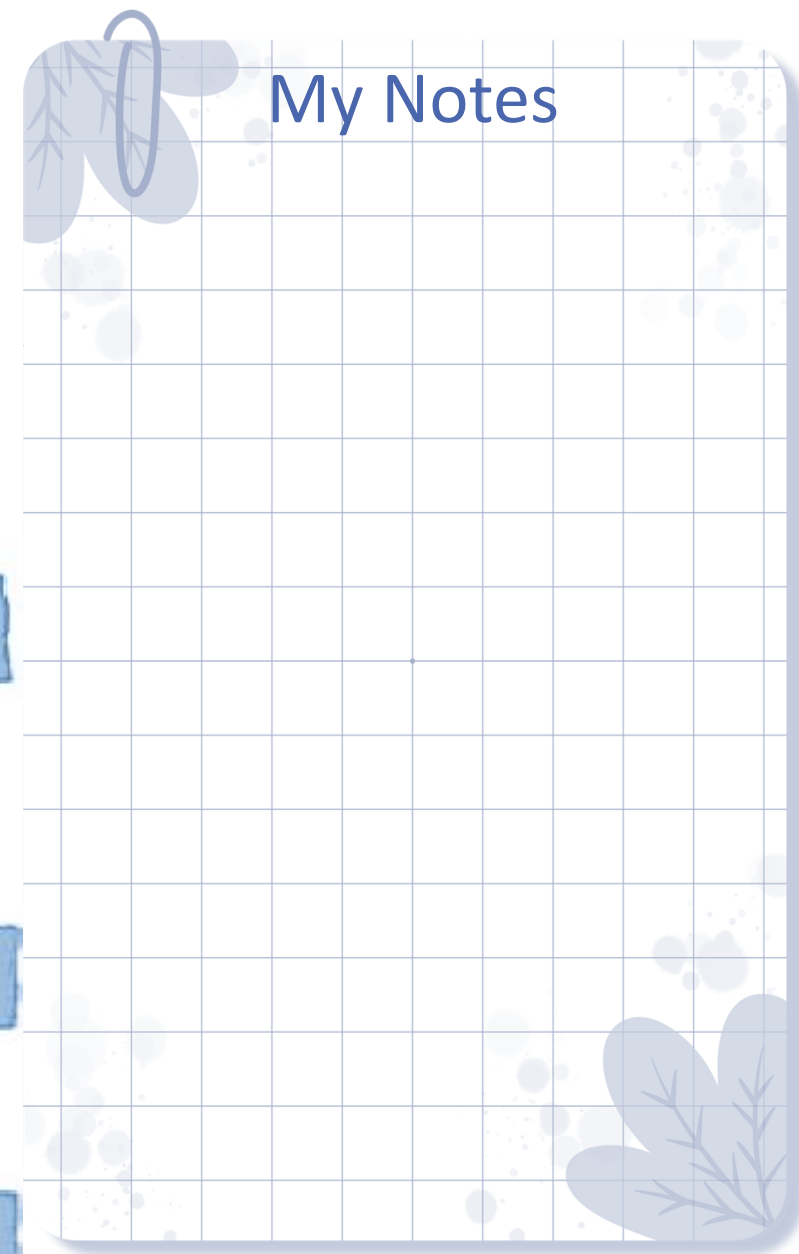
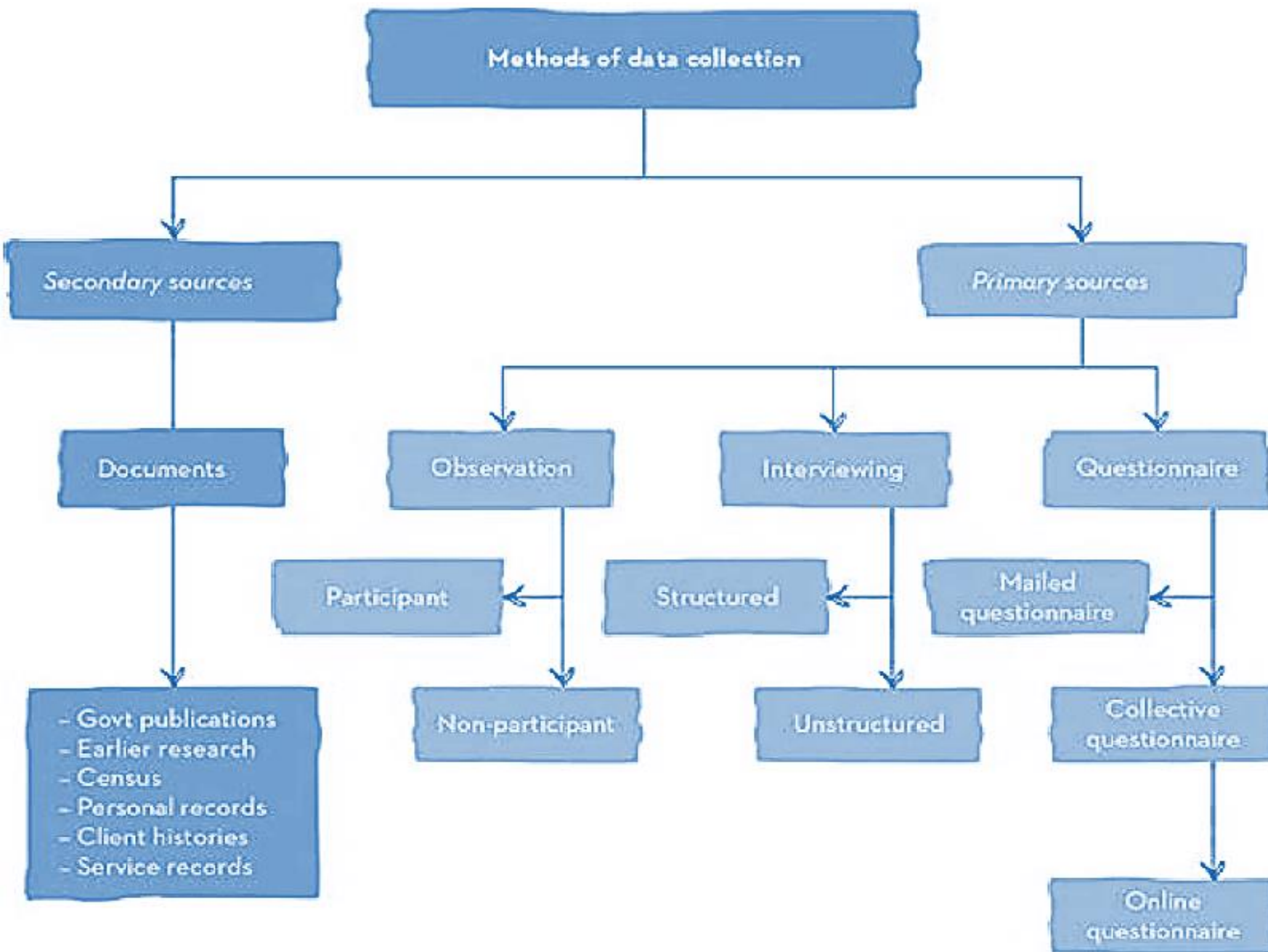
Major approaches to information gathering

Primary sources: Sources that provide primary data such as interviews, observations and questionnaires.

Secondary sources: Sources that provide secondary data are called secondary sources. Sources such as books, journals, previous research studies, records of an agency, client or patient information already collected and routine service delivery records all form secondary sources.

None of the methods of data collection provides 100 per cent accurate and reliable information. The quality of the data gathered is dependent upon a number of other factors, which we will identify as we discuss each method. Your skill as a researcher lies in your ability to take care of the factors that could affect the quality of your data. One of the main differences between experienced and amateur researchers lies in their understanding of, and ability to control, these factors. It is therefore important for a beginner to be aware of them.



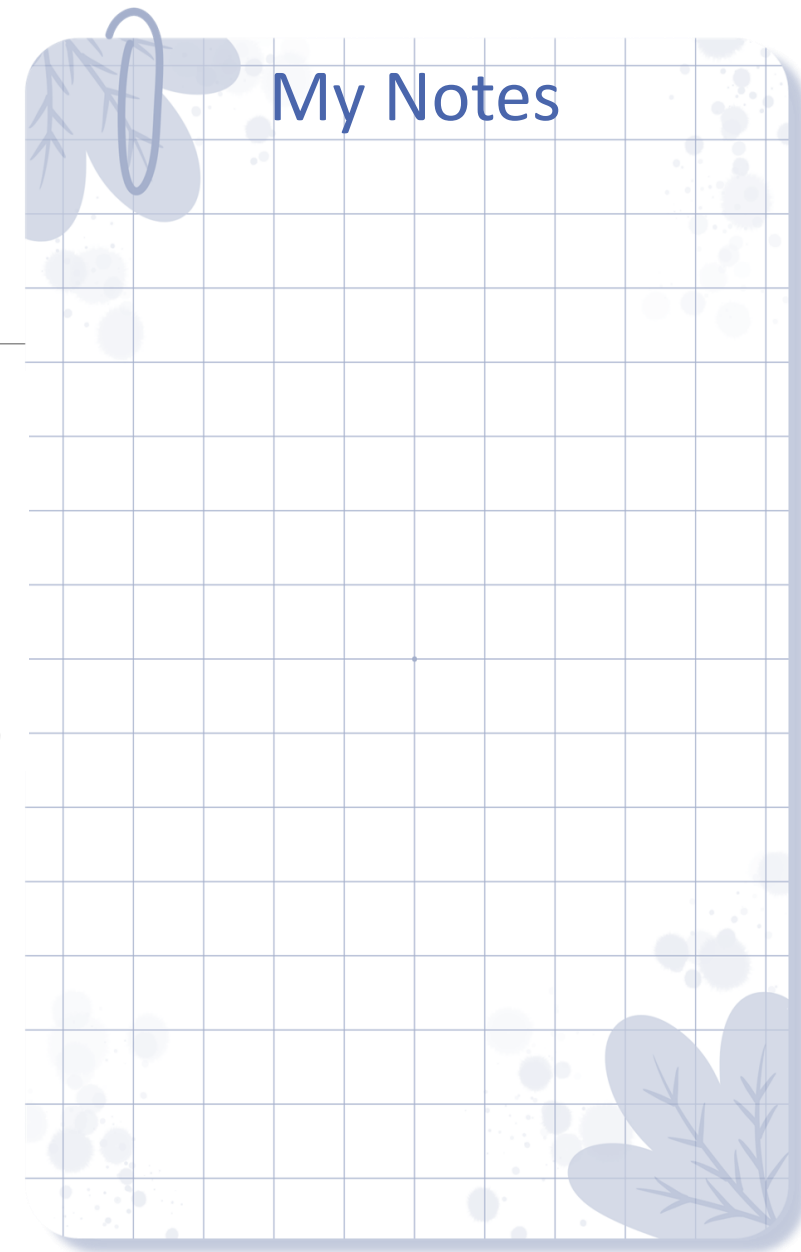


Collecting data using primary sources

Observation is one way to collect primary data. It is a purposeful, systematic and selective way of watching and listening to an interaction or phenomenon as it takes place. There are many situations in which observation is the most appropriate method of data collection; for example, when you want to learn about the interaction in a group, study the dietary patterns of a population, ascertain the functions performed by a worker, or study the behaviour or personality traits of an individual. It is also appropriate in situations where full and/or accurate information cannot be elicited by questioning, because respondents either are not co-operative or are unaware of the answers because it is difficult for them to detach themselves from the interaction. In summary, when you are more interested in the behaviour than in the perceptions of individuals, or when subjects are so involved in the interaction that they are unable to provide objective information about it, observation is the best approach to collecting the required information.



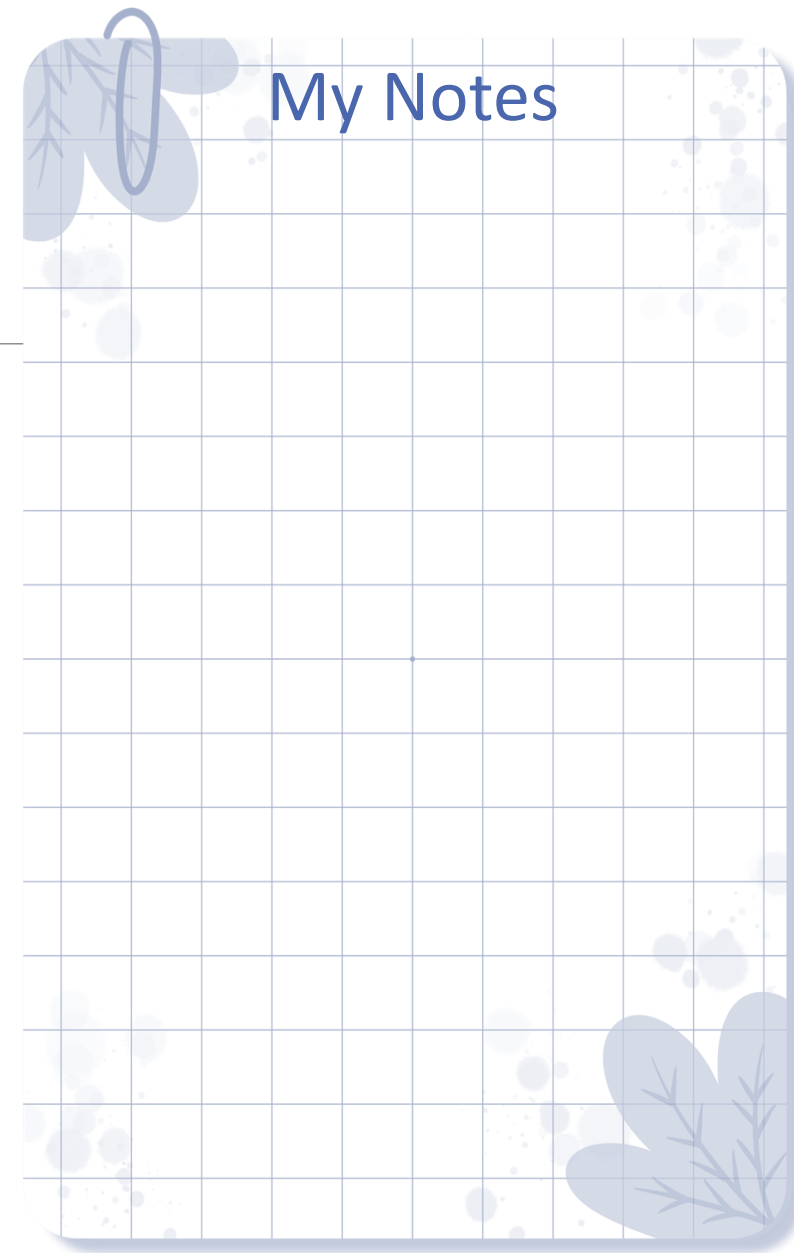
Collecting data using primary sources



Interviewing is a commonly used method of collecting information from people. There are many definitions of interviews, but it is essentially a person-to-person interaction, either face to face or otherwise, between two or more individuals with a specific purpose in mind. According to Monette et al. (1986: 156), 'an interview involves an interviewer reading questions to respondents and recording their answers'. According to Burns (1997: 329), 'an interview is a verbal interchange, often face to face, though the telephone may be used, in which an interviewer tries to elicit information, beliefs or opinions from another person'.

Interviewing: One of the commonly used methods of data collection in the social sciences. Any person-to-person interaction, either face to face or otherwise, between two or more individuals with a specific purpose in mind is called an interview. It involves asking questions of respondents and recording their answers. Interviewing spans a wide spectrum in terms of its structure, from highly structured to extremely flexible.

Collecting data using primary sources



Unstructured interviews: Interviews in which, you as an interviewer, have every flexibility in terms of questions that you ask of your respondents, explanation you provide, wording you use and the sequence in which you ask them.

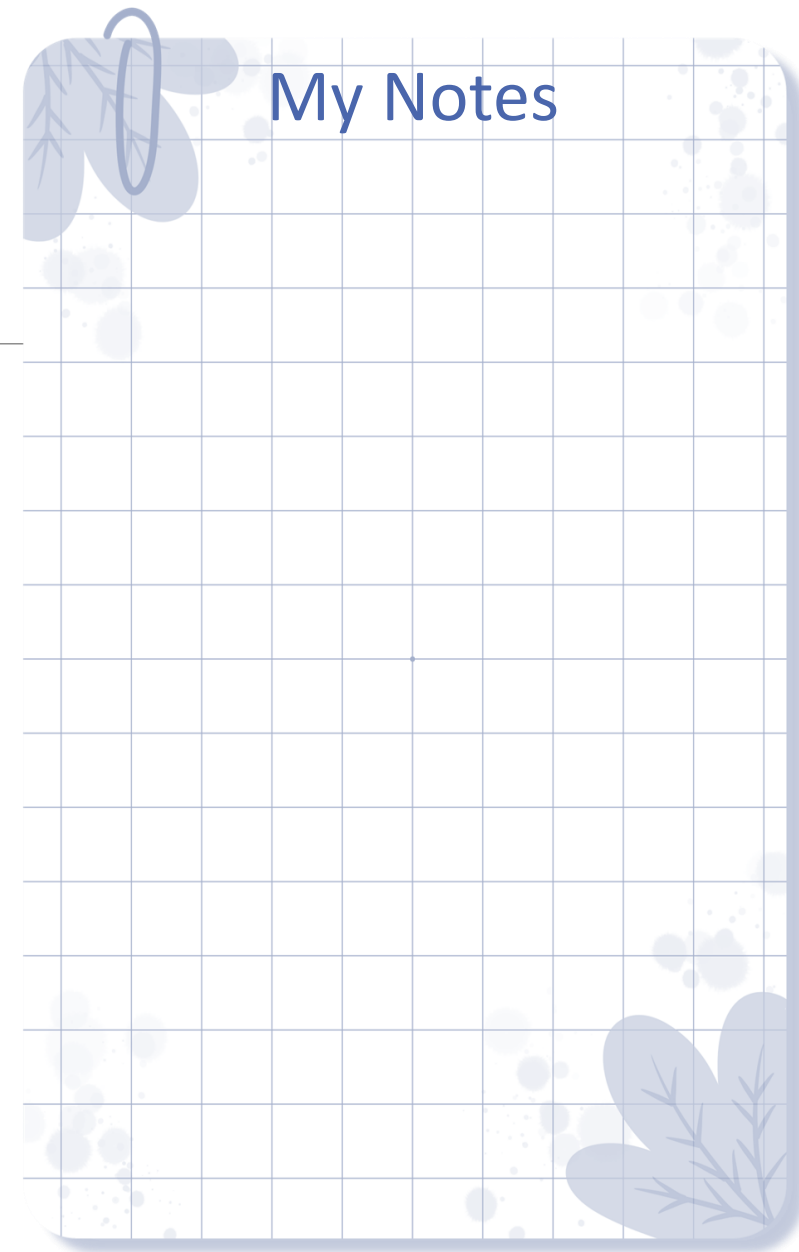
Structured interviews: Interviews in which the questions that you ask of your respondents, their wording and sequence are predetermined. Everything that forms part of the interview is fixed and predetermined and any deviation from it is not permitted.

Collecting data using primary sources

A **questionnaire** is a written list of questions, the answers to which are recorded by respondents. Thus, respondents read the questions, interpret what is expected and then write down the answers. The only difference between an interview schedule and a questionnaire is that in the former it is the interviewer who asks the questions (and if necessary, explains them) and records the respondent's replies on an interview schedule, and in the latter the replies are recorded by the respondents themselves. This distinction is important in accounting for the respective strengths and weaknesses of the two methods and their respective use in gathering data.

There are two types of question commonly used in social research:

- open-ended questions; and
- closed questions.



Collecting data using primary sources

My Notes

Open-ended question: In an open-ended question the possible responses are not given. In the case of a questionnaire the respondent writes down the answers in his/her words, whereas in the case of an interview schedule the investigator records the answers either verbatim or in a summary describing a respondent's answer.

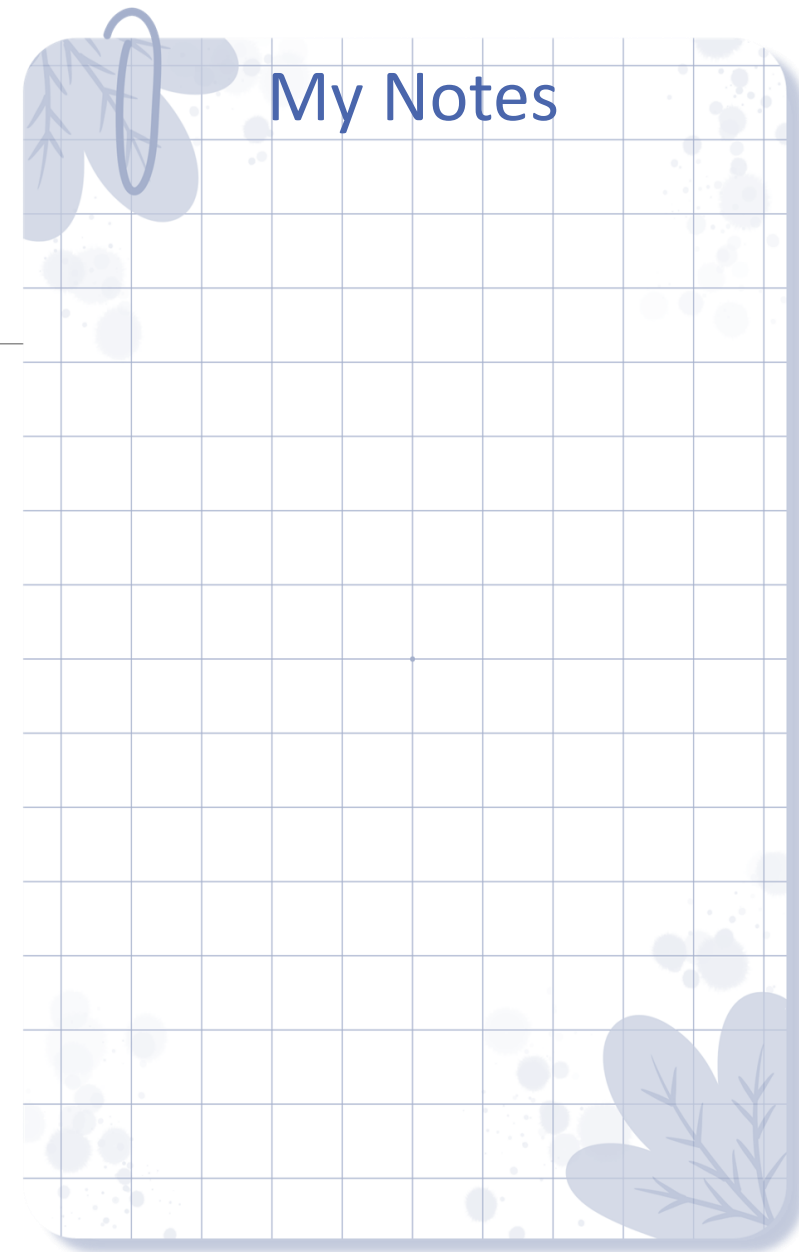
Closed question: In a closed question the possible answers are set out in the questionnaire or interview schedule and the respondent or the investigator ticks the category that best describes a respondent's answer.

Collecting data using primary sources: pros & cons

Advantages of a questionnaire

A questionnaire has the following advantages:

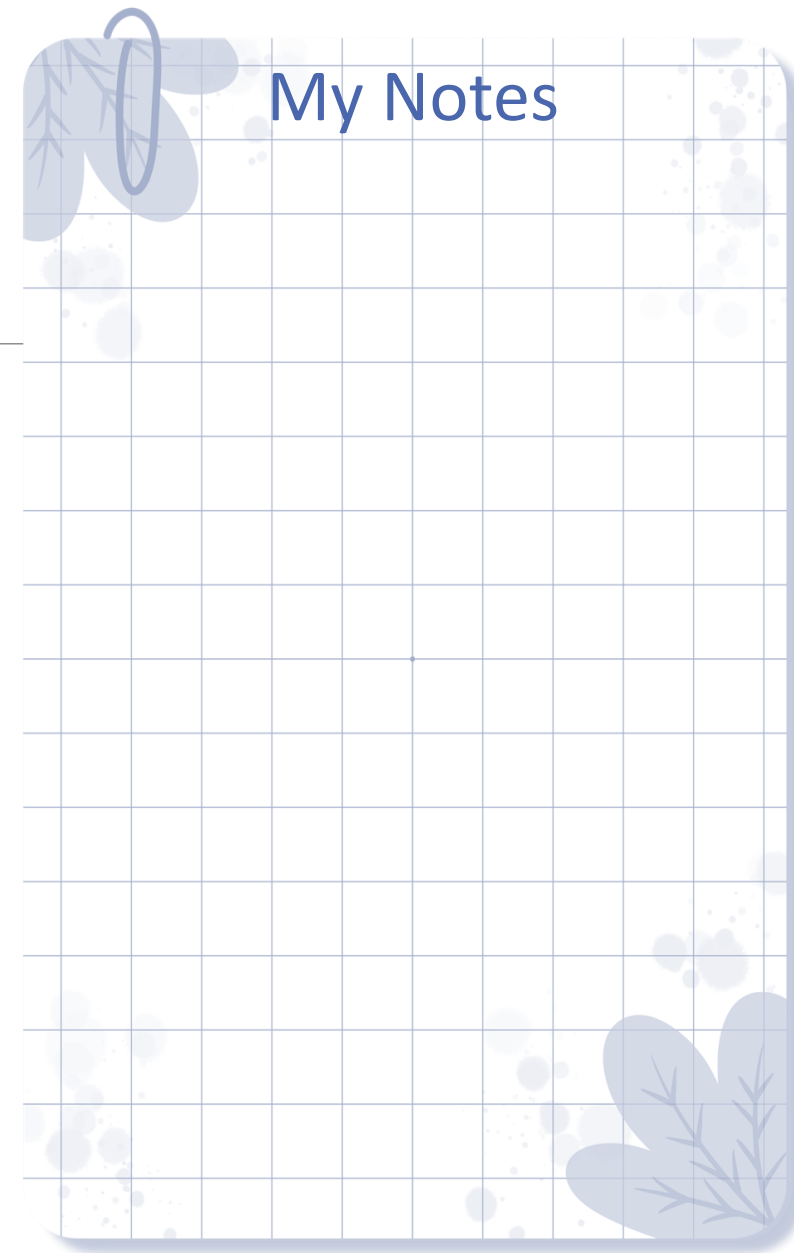
- **It is less expensive.** As you do not interview respondents, you save time, and human and financial resources. The use of a questionnaire, therefore, is comparatively convenient and inexpensive, especially when it is administered collectively to a study population.
- **It offers greater anonymity.** As there is no face-to-face interaction between respondents and interviewer, this method provides greater anonymity. In some situations where sensitive questions are asked it helps to increase the likelihood of obtaining accurate information.



Collecting data using primary sources: pros & cons

Disadvantages of a questionnaire

- **Limited application.** One main disadvantage is that application is limited to a study population that can read and write. It also cannot be used on a population that is very young, very old or handicapped.
- **Low response rate.** Questionnaires are notorious for their low response rates; that is, people fail to return them. If you plan to use a questionnaire, keep in mind that because not everyone will return their questionnaire, your sample size will in effect be reduced. The response rate depends upon a number of factors: the interest of the sample in the topic of the study; the layout and length of the questionnaire; the quality of the letter explaining the purpose and relevance of the study; and the methodology used to deliver the questionnaire. You should consider yourself lucky to obtain a 50 per cent response rate, and sometimes it may be as low as 20 per cent. However, as mentioned, the response rate is not a problem when a questionnaire is administered in a collective situation.
- **Lack of opportunity to clarify issues.** If, for any reason, respondents do not understand some questions, there is almost no opportunity for them to have the meaning clarified unless they get in touch with the researcher (which does not happen often). If different respondents interpret questions differently, this will affect the quality of the information provided.



Collecting data using primary sources: pros & cons

My Notes

Advantages of the interview

- **More appropriate for complex situations.** It is the most appropriate approach for studying complex and sensitive areas as the interviewer has the opportunity to prepare a respondent before asking sensitive questions and to explain complex ones to respondents in person.
- **Useful for collecting in-depth information.** In an interview situation it is possible for an investigator to obtain in-depth information by probing. Hence, in situations where in-depth information is required, interviewing is the preferred method of data collection.
- **Information can be supplemented.** An interviewer is able to supplement information obtained from responses with those gained from observation of non-verbal reactions.
- **Questions can be explained.** It is less likely that a question will be misunderstood as the interviewer can either repeat a question or put it in a form that is understood by the respondent.
- **Has a wider application.** An interview can be used with almost any type of population: children, the handicapped, illiterate or very old.

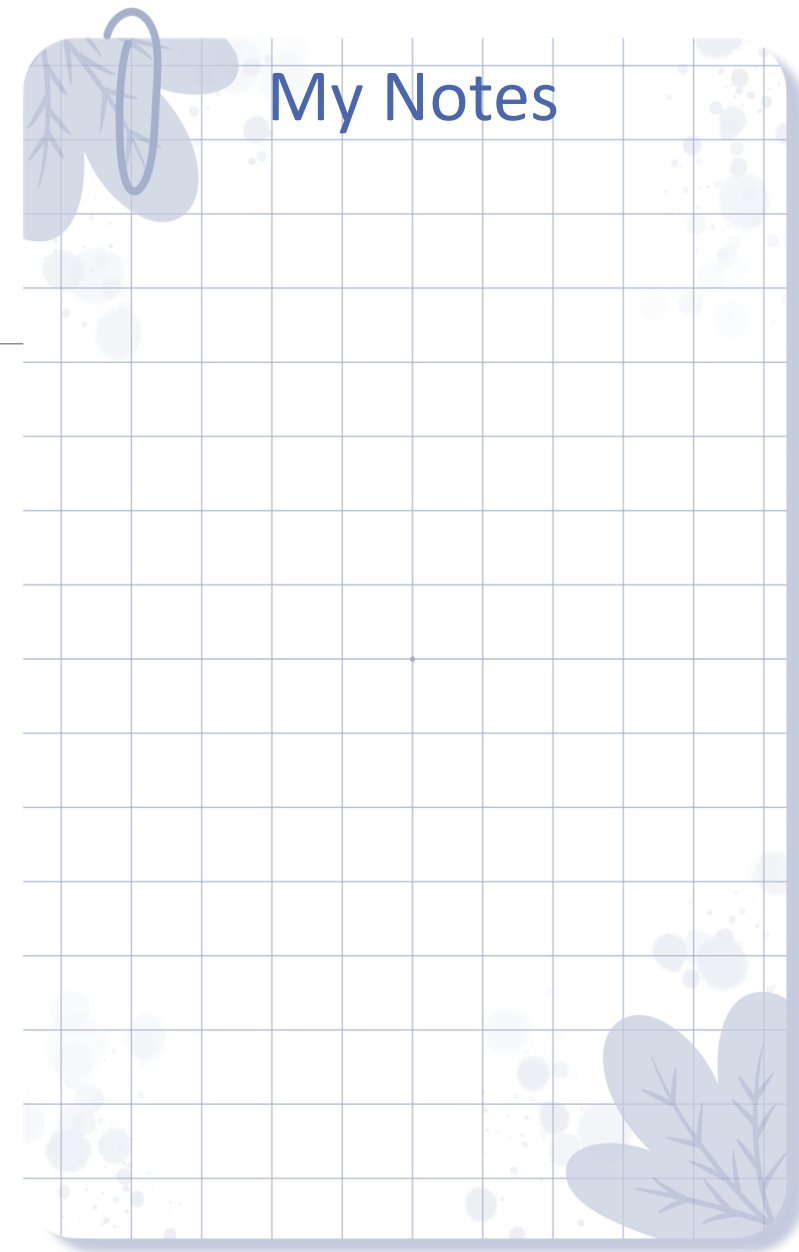
Collecting data using primary sources: pros & cons

Disadvantages of the interview

- **Time-consuming and expensive.** This is especially so when potential respondents are scattered over a wide geographical area. However, if you have a situation such as an office, a hospital or an agency where potential respondents come to obtain a service, interviewing them in that setting may be less expensive and less time-consuming.
- **The quality of data depends upon the quality of the interaction.** In an interview the quality of interaction between an interviewer and interviewee is likely to affect the quality of the information obtained. Also, because the interaction in each interview is unique, the quality of the responses obtained from different interviews may vary significantly.
- **The quality of data depends upon the quality of the interviewer.** In an interview situation the quality of the data generated is affected by the experience, skills and commitment of the interviewer.
- **The quality of data may vary when multiple interviewers are used.** Use of multiple interviewers may magnify the problems identified in the previous two points.

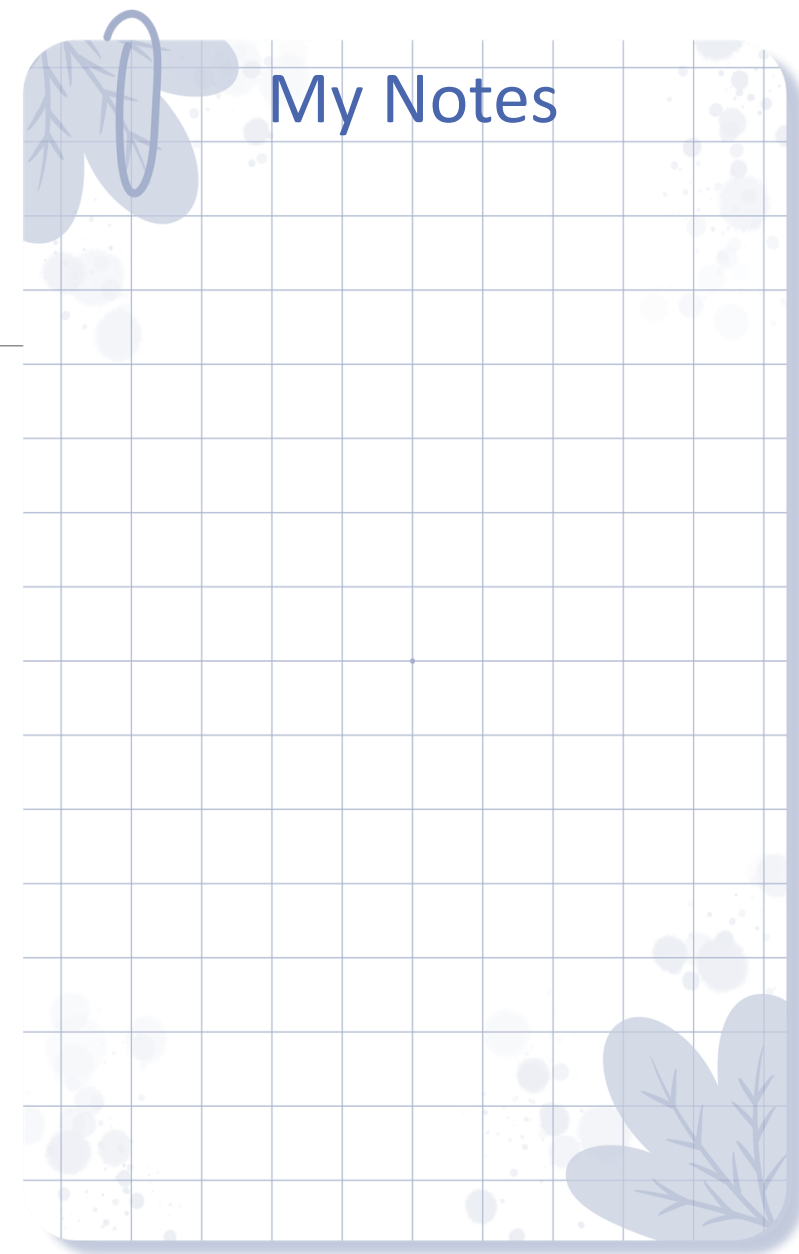
Formulating effective questions

- **Always use simple and everyday language.** Your respondents may not be highly educated, and even if they are they still may not know some of the 'simple' technical jargon that you are used to. Particularly in a questionnaire, take extra care to use words that your respondents will understand as you will have no opportunity to explain questions to them. A pre-test should show you what is and what is not understood by your respondents.
- **Do not use ambiguous questions.** An **ambiguous question** is one that contains more than one meaning and that can be interpreted differently by different respondents. This will result in different answers, making it difficult, if not impossible, to draw any valid conclusions from the information.'



Formulating effective questions

- **Do not ask double-barrelled questions.** A **double-barrelled question** is a question within a question. The main problem with this type of question is that one does not know which particular question a respondent has answered. Some respondents may answer both parts of the question and others may answer only one of them.
- **Do not ask leading questions.** A **leading question** is one which, by its contents, structure or wording, leads a respondent to answer in a certain direction. Such questions are judgemental and lead respondents to answer either positively or negatively.



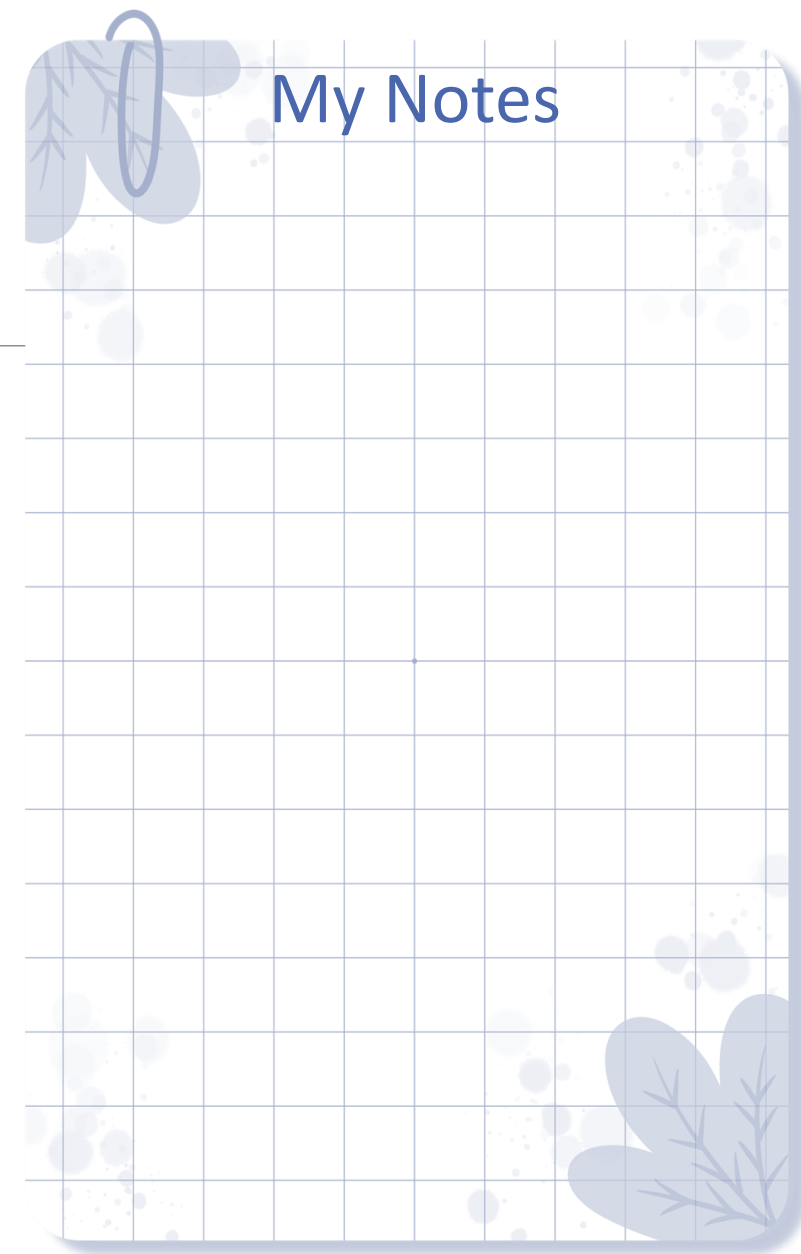
Collecting data using secondary sources

My Notes

- **Government or quasi-government publications** – There are many government and quasi-government organisations that collect data on a regular basis in a variety of areas and publish it for use by members of the public and interest groups. Some common examples are the census, vital statistics registration, labour force surveys, health reports, economic forecasts and demographic information.
- **Earlier research** – For some topics, a vast array of research studies that have already been done by others can provide you with the required information.
- **Personal records** – Some people write historical and personal records (e.g. diaries) that may provide the information you need.
- **Mass media** – Reports published in newspapers, in magazines, on the Internet, and so on, may be another good source of data.

Problems with data using secondary sources

- **Validity and reliability** – The validity of information may vary markedly from source to source. For example, information obtained from a census is likely to be more valid and reliable than that obtained from most personal diaries.
- **Personal bias** – Information from personal diaries, newspapers and magazines may have the problem of personal bias as these writers are likely to exhibit less rigour and objectivity than one would expect in research reports.
- **Availability of data** – It is common for beginning researchers to assume that the required data will be available, but you cannot and should not make this assumption. Therefore, it is important to make sure that the required data is available before you proceed further with your study.
- **Format** – Before deciding to use data from secondary sources it is equally important to ascertain that the data is available in the required format. For example, you might need to analyse age in the categories 23–33, 34–48, and so on, but, in your source, age may be categorised as 21–24, 25–29, and so on.





Success is liking yourself,
liking what you do, and
liking how you do it.

Maya Angelou

quote fancy

@SalhaAlzahrani