

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.