Assess the value of using secondary data in sociological research. (40 marks)

Secondary data is data already produced often by someone who is not a sociologist, e.g. Government, Charities, Trade Unions, letters or diaries, TV programmes and e-mails. Official statistics are a source of secondary data. The government produces them. There are both hard and soft statistics. Hard statistics include birth, death and marriage rates. These are registered when they occur by law, so these statistics are entirely objective. Another type of official statistic is called soft statistics, e.g. crime statistics and unemployment figures. These might not seem as objective as they first seem. They are open to manipulation for political ends, and can be considered to have a political use. For example, the methods used to measure unemployment have been changed over 20 times, because not everyone without a job counts as unemployed, e.g. Elderly. The picture that statistics give can depend on how they were collected, and on what parts of the statistics matters – they do not always present a valid picture of society.

Secondary data has both advantages and disadvantages, from issues such as practical and ethical for example, Laslett used secondary data, in the form of Parish records to study family structure. Secondary data can save time and money and they provide access to historical data that cannot be produced using primary research because the event took place before the current members of society were born. Probably the main advantage of using secondary data is that the data has already been collected thus saving time and money on behalf of the researcher. This time can instead be used on making the data more valid, possibly by using secondary data as a block to which to base further research on. Secondary data can also be useful as evidence to results found, a researcher can compare previous research to research found today, and can make before and after comparisons.

Historical secondary data can also cause problems. It is probable that the data collected was from the middle class. Therefore, it reflected ideologies of those who produce them, being the middle class. This would make the data unrepresentative, as it didn't reflect the society as a whole. Also, many pieces of secondary data are accused of being male stream, this being it focused mainly on men. Feminists say that official statistics are male stream.

Secondary data is usually accepted at face value to positivists, especially official statistics. It is easy to look for correlations and cause effect relationships in society. Durkheim used official statistics on crime on his research on suicide and concluded that suicide is a result of loneliness. However, the cause/effect that a particular set of statistics shows may not in fact be the most important factor. To compare two variables is to ignore all the other variables. Furthermore, to analyse figures on a table or chart is to ignore all the subjective and individual factors, which Interpretivists focus on.

Interpretivists use qualitative secondary data such as letters and diaries. These give an insight into individual opinions in society. However, John Scott said that the subjective nature of this kind of data is a problem – the researcher may misinterpret documents or find them hard to understand, especially if they are old. For example, the meaning of the word "gay" has changed over time. If this word I used in a piece of historical data, it may have been understood to mean "happy" which is obviously different to what we know it as today.

The extraction of data from documents can be very time consuming, especially as data/ideas are not always presented in a useful form i.e. definitions/categories may be different. In addition, not all documents are easily available e.g. the Black Report

(1980) was released in very select circumstances; this is even more true of personal documents.

Apart from problems of time, presentation, and access, there are a few general considerations that should be applied to any piece of secondary data. Firstly validity. Does secondary data give us a true and detailed picture of what is being measured? In relation to diaries/e-mails, it could be possible that the author lied or exaggerated about their activities. Secondly reliability. To what extent have the data been influenced by personal biases? Obviously, replication is impossible, all the researcher can do is compare one account with others and try to balance the various accounts. The general argument here is that public documents are likely to be higher in terms of reliability etc than private documents like letters etc although this may be a questionable assumption. Personal documents such as diaries and e-mails data may be fake or unrepresentative. It may not be credible; the author of the document, especially if unknown may have lied especially with diaries, where some exaggerate.

Generalisation is difficult since we cannot know how representative the documents are; Plummer argues that, "any life-history is always representative of a larger group. It always tells you something about more than that one person." Nevertheless, generalisation is a problem, not least because of the selectivity of the data. Here the problem of 'discarded data' becomes more acute. All research involves the discarding of data.

In conclusion, the main point in favour of secondary data in general is that the researcher doesn't have to do all work himself/herself. Gathering primary data by means of surveys, questionnaires and interviews can take up a lot of time, effort, and money. However, sociologists using secondary data must rely on the original researcher to collect data with reliable and repeatable methods. When using this secondary data, the user has a responsibility to interpret and analyse the data in an unbiased way.