Analysing the Fear of Crime using the British Crime Survey

Secondary Analysis for Social Scientists Workbook

Authors: Lee Williamson, Mark Brown, Jo Wathan, Vanessa Higgins, Centre for Census & Survey Research, University of Manchester.

This workbook is an updated version of the “Secondary Analysis for Social Scientist, Exemplar workbook 1 “Analysing the fear of crime using the British Crime Survey”, funded by the Economic and Social Research Council (ESRC).

Updated for the UK Data Service by Sarah King-Hele, August 2013
**Who is this workbook for?**

This workbook is designed for undergraduate and Masters Dissertation students or other people planning to start research using secondary data with an emphasis on the use of quantitative datasets. It contains:

- some material for you to read and think about
- some guidance on how to use online resources and software to conduct an analysis
- some questions to think about and answer.

Like other materials you will have received on courses, this workbook is not comprehensive. It contains suggested reading for books that discuss the process of analysing data. However, the greater part of your reading will relate to the topic of your analysis which is up to you and your imagination!

There are a lot of interesting questions that can be answered by using data which has already been collected. Many large, good quality, datasets are *archived* and made available for researchers to use – you can use them too to do your own research. It’s a great resource because the costs involved in creating the data and making them available have already been paid for on your behalf. However, as with all research you will need to read and think through the work if you are going to demonstrate your skills as a critical social scientist. You will also need to pick up some skills to ensure that you are analysing the data correctly.

This workbook is designed to help you with the skills to do such work. It will guide you through the key steps involved in analysing the British Crime Survey to answer some questions about fear of crime. You then will get the opportunity to try out some ideas yourself.

Try these skills on a research topic and dataset that really interests you - the steps you take will be the same.
Contents

Section 1: Before you start, think! ................................................................. 4
Section 2: Finding and using published tables and reports ........................................ 11
Section 3: Finding the right survey microdata to analyse ........................................... 21
Section 5: Accessing survey microdata from the UK Data Service ................................. 30
Section 6: Getting started with the BCS 2007-2008 Teaching Dataset ...................... 41
Section 7: Exploratory analysis: using the British Crime Survey 2007-2008 Teaching Dataset to explore variations in fear of crime ..................................................... 65
Section 8: Further exploratory data analysis .............................................................. 85
Section 9: Reporting exercise .................................................................................. 93
References ........................................................................................................... 99
Appendix 1: Introducing the British Crime Survey 2007-2008 ...................................... 102
Appendix 2: The BCS 2007-2008 Teaching Dataset .................................................. 106
Section 1: Before you start, think!

1.1 Research questions

The more you know about what you want to achieve in your research, the easier it will be to do it.

In order to help you to establish a research question, you can:

- Read around your subject to see what is known, and which research methods have been used
- Consider the sorts of analyses that have been done with available data

The kinds of question that you want to answer can be very broad to begin with. However, initial questions of interest may not be researchable in practice using survey data already collected and will, in any case, need to be refined and developed using the literature on the topic.

Some interesting crime research areas, amenable to secondary analysis of survey data, include:

- Gender differences in the experience of crimes
- Public Perceptions of Crime Rates
- Gender, age and class differences in the fear of crime
- The impact of fear of crime on behaviour
- Regional differences in victimisation

In this workbook we will explore fear of crime. The first step when undertaking research is to find out about your research area.

1.2 Reading to develop your research questions

Thinking critically about a research topic is not specific to secondary analysis. We will therefore not provide a full discussion here. The following extracts from real research demonstrate how reading around your topic will help to:
Identify the relevance of your topic and key issues to be considered
Understand how key concepts are understood by other researchers
Indicate any problems with the measures that you might be considering using.

Pantazis and Gordon (1999:198) provide a useful introduction to the importance of this question in social research:

“Crime and fear of crime have emerged as major public and political issues in recent decades. This may, in part, be attributed to the enormous growth in recorded crime since the 1970s, when the average annual increase has been about 5 per cent. [...] People’s anxiety about crime has also risen in this period. Surveys have repeatedly shown that crime has surpassed unemployment and health as an issue of major public concern (Jacobs and Worcester, 1991). This has led some criminologists to conclude that fear of crime poses almost as large a threat to society as crime itself (Clemente and Kleinmen, 1977). Both official statistics and research show that there is an uneven distribution in criminal victimisation and in fear of crime (Barclay et al., 1995; Mirrlees-Black et al., 1996). Some individuals and some geographical areas are more vulnerable to crime than others and certain groups fear crime more.” (Pantazis and Gordon 1999:198)

It is also important to fully understand what is meant by fear of crime in order that we can relate this to measures asked in the British Crime Survey (BCS), for example. To expand:

“The term ‘fear of crime’ has gained almost universal use. Some believe it is overused and misused (Ferraro and La Grange 1987). Pantazis (2000) points to the confusion that exists in the literature on the precise meaning of fear of crime and how it should be measured. Nevertheless, it is widely accepted as a short-hand term to refer to a whole range of attitudes feelings, reactions and emotions that [...] people have towards crime and victimisation. It can include worry and anxiety (Hough 1995, cited by Pantazis 2000); terror, panic and unease (Bannister and Fyfe 2001) (Burnett 2006:127-128)

In your own research for your dissertation you need to review a wide range of literature, however, for this workbook only a few sources have been consulted.

In this worked example we will be considering the British Crime Survey (BCS) question: ‘How safe do you feel walking alone in this area after dark?’ This information is recorded in the survey as the variable
‘walkdark’. However, before using any question as a measure of fear of crime, it is important to think critically about the question and what it is actually measuring. For example, in the first BCS report Hough and Mayhew (1983:23) comment “This question has been used extensively in American and Canadian surveys and is believed to be a sensitive measure of fear of violent crime in public places”. Furthermore, Maxfield (1984) after the first CSEW reports that:

“Skogen and Maxfield (1981) present an extended discussion of this question and its adequacy as a measure of fear of crime, concluding that, though relatively narrow scope, its nonetheless a reliable and valid indicator of fear of personal safety. Certain characteristics of the question are notable. It presents a more or less specific situation or stimulus – walking along after dark. Assessments of risk are implicit, since feelings of safely on neighbourhood streets at night presumably reflect beliefs about the likelihood of victimisation. The question asks specifically about night-time safety primarily because other studies have found that few people feel unsafe walking in their neighbourhood during the day.

There are, however, at least three limitations of this question as a measure of fear of crime. The first is that it may conceal the true extent of fear among men: whereas women may feel little reluctance to admit to feeling uneasy on the streets at night, men may be wary of expressing similar fears. Secondly, the measure obviously refers only to street crime, which is relatively rare, spatially concentrated, only one of a number of crime threats people fear. Burglary is far more common than the kinds of predations people are likely to encounter walking about in their neighbourhood (Hough and Mayhew 1983)” (Maxfield 1984:4)

We are interested in investigating fear of crime by various subpopulations (such as sex and age group). By means of an introduction to the subject the quote below is of a review of literature from the USA – however, it does provide an understanding that the characteristics of a respondent can influence perceptions of crime:

“Studies show that many demographic features affect perceptions of crime both negatively and positively. Concerning age, research has varied results. Some research shows that older people tend to have less fear of crime when compared to younger people (Chadee & Ditton, 2003; Ferraro & LaGrange, 1992; Rountree, 1998; Ziegler & Mitchell, 2003). Whereas other research states the opposite older adults report greater fear of crime (Baker et al., 1983; Weinrath & Gartell, 1996). People who are more affluent and belong to a higher social class have less fear of crime (Rountree,
In addition, women report a higher fear of crime than men (Haynie, 1998; LaGrange & Ferrano, 1989).” (Truman 2005:1
http://www.urj.ucf.edu/docs/URJmanuscript_Truman_080509.pdf [accessed 4/10/13])

In your research it is important to give critical consideration to the survey questions that underpin the specific variables used in your analysis.

Other reports on the BCS are valuable in providing further background information on this variable and how it relates to social and demographic characteristics. For example, the 1982 BCS report found differences for some subpopulations, for example:

“People in manual occupations were consistently more fearful than those in non-manual, and those whole live alone more than those living with others. […]"

| Table 5: Fears for personal safety after dark and risks of ‘street crime’ |
|-----------------|------------------|------------------|
|                  | % feeling ‘very unsafe’ | % victims of ‘street crimes’ |
| **Men**          |                  |                  |
| 16-30            | 1                | 7.7              |
| 31-60            | 2                | 1.6              |
| 61+              | 9                | 0.6              |
| **Women**        |                  |                  |
| 16-30            | 15               | 2.8              |
| 31-60            | 17               | 1.4              |
| 61+              | 34               | 1.2              |

Question: “How safe do you feel walking along in this area after dark?”

Weighted data; unweighted data n=10,905 (Source: Hough and Mayhew 1983: 25, Table 5)

The table shows that those who felt most unsafe were least often victims. These findings may suggest that the fear of crime is irrational or excessive, but some caveats should be stated. In the first place, those who are most fearful may expose themselves to least risk, thus avoiding victimisation. For example the elderly go out less than others, and consequently they are less at risk of street crime and their homes less at risk of burglary. Again, many young men are not afraid of a fight, and get into fights; older people do not. It is complicated to calculate victimisation rates which fully take into account differential exposure to risk, and fuller analysis is needed here […]
Secondly, the experience of victimisation has a different impact on different people. There are straightforward differences of physical vulnerability – the old are physically frail, and women are generally weaker than men. […]

There are also material differences: the wealthy are more cushioned against financial loss […] In other words, people’s reactions to crime – and to the prospect of being victim of crime – are very variable, and research cannot simply decree that one person’s anxiety is reasonable and another’s excessive. Nevertheless, in some areas fear of crime appears to be a serious problem which needs to be tackled separately from the incidence of crime itself.”

(Hough and Mayhew 1983:23 and 25)

All this work was based on the 1982 BCS. More recently the 1992 BCS has shown difference between fear of crime by gross household income level (grouped into quintiles):

---

![Figure 2.7](image_url)

Source: Pantazis and Gordon 1999 p.204

Source: Module 2 Theories About Crime: Public Perceptions of Crime Rates

Available online from: [http://x4L.data-archive.ac.uk/learning/module2/module2.pdf](http://x4L.data-archive.ac.uk/learning/module2/module2.pdf)

Pantazis and Gordon (1999:206) report that:

“Respondents in the poorest household income quintile experienced the highest rates of crime (49 per cent), while respondents in the richest two income quintiles had a fear of crime of only 23 per
cent. Thus the 1992 BCS demonstrates that despite experiencing higher levels of victimisation, respondents in the richest household income quintile are less fearful of crime.” (Pantazis and Gordon 1999:206)

Finally, Pantazis and Gordon (1999:209) conclude:

“The analysis [...] showed a clear link between poverty and fear of crime. People living in poverty suffered from a disproportionately high level of fear of crime regardless of whether or not they had been victimised. This fear is not irrational, as some criminologists suggest, but results from the greater impact that crime has on poor people” (Pantazis and Gordon 1999:209)

So...?
These few references have demonstrated how reading around the subject can help to firm up your research question. These extracts have helped to:

- Confirm that fear of crime is a suitable topic for research
- Provide some background on the pros and cons of the measure we are considering using
- Identify groups which experience different levels of fear of crime

When you do your own research you will need to read a wider range of sources than it has been possible to discuss here. In particular it will be helpful to read material that has arisen from qualitative sources, as this will provide contributions to theory that cannot be adequately explored using survey data alone.
The next two sections provide information about published materials and data service resources.

**Section 1 Key Points**

- The more you know about what you want to do, the better your chances of doing it
- Some topics are better suited to secondary analysis than others
- Reading around your subject will:
  - Help you to identify important areas in your topic
  - Help to identify the types of questions you can explore with data
  - Help you to become a critical user of measurement data
Section 2: Finding and using published tables and reports

There are many useful published reports and tables created by agencies that carry out survey research that can be consulted when you have a research question/topic in mind. These provide ways to obtain information and statistics from survey data without having to access the individual-level (microdata) files.

Published results are useful for several reasons:

- The answer to one or more of your questions may already have been answered.
- Reviewing published figures will form part of your literature review.
- Looking at published reports will help you to understand what sort of results you can expect and will help you to spot errors in your own work.

2.1 Finding published tables and reports

A useful way to obtain tables and reports from surveys is to go through the UK Data Service website. This site collects together useful resources for users of surveys and will be one with which you will become very familiar.

Other places to look...

If a survey you are using is not supported by the UK Data Service you can try finding published reports and tables by:

- Using information in the survey’s documentation
- Using the web site of the organisations that conducted, commissioned (often a government department) or distribute the data.
- Using a general web search (like Google)
- Using a national statistics website such as the Office for National Statistics http://www.statistics.gov.uk/hub/browse-by-theme/index.html
- Using a good library...

In this section we limit ourselves to tables and publications which have been produced from the Crime Survey for England and Wales (formerly the British Crime Survey).

Follow instructions marked with a ➔ on your computer:
Go to http://www.ukdataservice.ac.uk/ then click on Get data > Other providers > Ready-made UK Statistics.

Under UK Statistics, scroll down and click on Office for National Statistics (ONS) theme pages: Crime and Justice. This takes you to the following webpage that has useful links related to the survey, including, in this case, links to published tables.
→ Click on the tabs Summaries and Publications or Data tables for reports and data in Excel, some of which are taken from the Crime Survey for England and Wales (formerly the British Crime Survey).

→ Click on one of the statistical bulletins in the Summaries and Publications section.

This takes you to a page that summaries key points from the publication to help to decide whether you want to look further. On this page there are links to Download PDF and Data in this release for the data tables from the report in Excel.

→ Go to the Data in this release.

→ Click Open, and an excel file will open. If there are a number of spreadsheets contained in the file, you will probably see the first one which may contain a list of all the tables contained in the other spreadsheets in that file.

For example:

<table>
<thead>
<tr>
<th>Crime in England &amp; Wales, year ending March 2013 - bulletin tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>These data tables are published alongside the bulletin Crime in England &amp; Wales, year ending March 2013.</td>
</tr>
<tr>
<td>For explanatory notes on these statistics see the User Guide to Crime Statistics for England and Wales.</td>
</tr>
<tr>
<td>Data tables shown in this workbook relate to both police recorded crime and the Crime Survey for England and Wales (CSEW).</td>
</tr>
<tr>
<td>The tables contained in this file comprise:</td>
</tr>
<tr>
<td>Tables</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>
This is useful because it means that you can make charts from the tables in the report or present the tables which underlie the charts in the report.

**A** Can you use this data to calculate the percentage of all crime reported in the CSEW that is violent in 1981, 1991 and 2001/2?

**Question for you to answer**

Tip: In a cell in the B column, type the formula “=(B17/B20)*100”

Copy that formula across the same row of the other columns to produce the equivalent percentage for other years. In column C the formula should now read “=(C17/C20)*100”. In column E it should read “=(E17/E20)*100”.

➤ Go back to the ESDS website and click on ‘Crime in England 2005/6 online report (HO web site)’ link

This brings you to the following screen:
Now scroll down the page to section on 'Structure of the report', as shown below, where there are links to the Excel tables in the report.

Statistical tables from the chapters are available in Excel format:
- chapter 2 tables (324kb)
- chapter 3 tables (18kb)
- chapter 4 tables (40kb)
- chapter 5 tables (64kb)
- chapter 6 tables (116kb)
- chapter 7 tables (138kb)
- chapter 8 tables (112kb)

Note: Revised figures for distraction burglary added 01.12.06

Summary of the main statistics (pdf, 980kb)

An additional version of Table 2.01 (with additional BCS years) is available: Excel

Additional tables reporting statistics for individual Crime and Disorder Reduction Partnerships and Basic Command Units are available:
- Crime and Disorder Reduction Partnerships - Recorded Crime for Key Offences 2004/05 to 2005/06 Excel
- Basic Command Unit - Recorded Crime for Key Offences 2004/05 to 2005/06 Excel
- Basic Command Unit - Detection for Key Offences 2004/05 to 2005/06 Excel

Now click on the 'Excel' link for the table labelled 'An additional version of Table 2.01 (with additional CSEW years) is available'

Then click to open (or save and then open) the Excel file

The following Excel sheet is more detailed than the one from Social Trends 37, as it has all years.
This is an example of how you could easily save the Excel files from the Home office website and even present the information from the CSEW tables in chart form.

Can you use this data to produce a pie chart of the breakdown of different types of violent incidents in 2005/6?

Tip:

- Select the violence categories (for labels) and the associated data cells for 2005/6 – you will need to Ctrl-click to select 2 different areas. See screenshot below.
- Click on the graph button and select pie chart
2.2 Published Reports

Published reports are also a very useful way to get information to use in your dissertation. The report usually contains tables produced by the agency that carried out the survey which can be reproduced and even made into charts for your dissertation. The reports may also contain examples of analysis which the researchers have undertaken which may be of interest.

Using the CSEW/BCS we provide some examples of information from published reports to demonstrate what could be done - including reporting on statistical analysis undertaken. For demonstration purposes we will consider only three published reports: 1982, 2001 and 2005/6. The published BCS reports gives tables of some of the variables asked in the BCS, including the number of respondents that the percentages are based on.

For example, if we consider the question on personal safety: ‘How safe do you feel walking along in this area after dark?’ the 1982 BCS report (Hough and Mayhew 1983), contains the table reproduced below:
### Table 4 and 5: Fears for personal safety after dark

#### Table 4
Fears for personal safety by age, sex and area: percent feeling “very unsafe”

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Inner cities</th>
<th>Other large city areas</th>
<th>Other areas</th>
<th>% feeling ‘very unsafe’</th>
<th>% victims of ‘street crimes’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>31-60</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>61+</td>
<td>27</td>
<td>12</td>
<td>6</td>
<td>9</td>
<td>0.6</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-30</td>
<td>28</td>
<td>18</td>
<td>11</td>
<td>15</td>
<td>2.8</td>
</tr>
<tr>
<td>31-60</td>
<td>38</td>
<td>21</td>
<td>13</td>
<td>17</td>
<td>1.4</td>
</tr>
<tr>
<td>61+</td>
<td>60</td>
<td>41</td>
<td>29</td>
<td>34</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Question: “How safe do you feel walking along in this area after dark?”

Weighted data; unweighted data *n*=10,905 (Source: Hough and Mayhew 1983:23, Table4 and 25, Table5)

However, if you wanted to compare this information with more recent published reports the responses are given in a different format, as shown in the table below from the 2001 BCS report:

#### Table A5.8 Trend in concern about personal safety (1984 to 2001 BCS)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking alone in area after dark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very safe</td>
<td>31</td>
<td>28</td>
<td>27</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>Fairly safe</td>
<td>38</td>
<td>39</td>
<td>41</td>
<td>40</td>
<td>43</td>
<td>42</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>A bit unsafe</td>
<td>20</td>
<td>21</td>
<td>21</td>
<td>23</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Very unsafe</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Unweighted N</td>
<td>10,999</td>
<td>10,353</td>
<td>10,021</td>
<td>14,461</td>
<td>16,303</td>
<td>14,903</td>
<td>19,319</td>
<td>8,864</td>
</tr>
</tbody>
</table>

(Source: Kershaw et al. 2001:71, tableA5.8)
Is there a trend in how safe people feel when walking alone in their area after dark? Use the figures in table A5.8 to justify your view.

- You might want to consider whether there is a trend in those who feel safe, or a trend in those who feel unsafe, or a pattern in the ratio of those who feel safe to those who feel unsafe.
- Is any one of the ways of measuring safety using table A5.8 preferential, or is it better to compare different measures?

In addition, you may create charts in Excel from the percentage from the published reports, such as the one shown below from the 2001 BCS report:

![Line graph showing trend in concern about personal safety (1984 to 2001 BCS)](chart.png)

(Source: Kershaw et al. 2001:75, table A5.14)

Does this line graph help you to understand whether there has been any change in concern about personal safety between 1984 and 2001?

⇒ If you have time...
⇒ Can you create a line graph which contains two lines, one for those who feel safe (% very safe + % fairly safe) and not safe (% a bit unsafe + % very unsafe)?
⇒ You may need to type the data into Excel

Even where tables in reports for different years are not strictly comparable, there is often information comparing trends over time for key topics. This may (or may not) be detailed enough for your purpose.
2.3 Using other researchers’ statistical analysis

There may be cases where there has been some statistical analysis undertaken which may be of interest to your research question but the analysis is beyond the scope of your dissertation. In this case it is acceptable to report the research findings as long as they are well referenced.

Walker et al. (2006), using CSEW 2005/06, researched ‘Geographic patterns of crime’. The results of the statistical analysis showed that, once the influence of other characteristics was taken into account there was a significant but weak association between the following:

- living in an urban area and increased risk of becoming a victim of burglary,
- living in an urban area and increased risk of becoming a victim of vehicle theft;
- living in a particular ACORN area and increased risk of becoming a victim of vehicle theft, for example, households in ‘moderate means’ areas were at increased risk when compared with ‘wealthy achiever’ areas; and,
- living in a particular ACORN area and increased risk of becoming a victim of violent crime, for example, people in ‘moderate means’ areas were at increased risk when compared with ‘wealthy achiever’ areas.

(Walker et al. 2006:118)

The results of the statistical analysis could be reported in your dissertation. Journal papers may also be a valuable source of statistical analysis.

Useful online resources using the CSEW/BCS:

Home Office Website: Crime statistics
ONS website: Crime and Justice

In addition, the CSEW has been included in some of the ONS Social Trends:

Section 2: Key Points

- The answers to some of your questions may already have been produced.
- More information can be obtained by reworking and reanalysing published tables, reports and other researchers’ work.
Section 3: Finding the right survey microdata to analyse

Section 2 looked at the ways in which pre-existing tables, reports and other research results can be used to inform your research. This section starts to focus on microdata. We first consider what microdata are and why you might want to use them. We then look at resources to help you find survey microdata for studying crime.

3.1 What are microdata?

Microdata are data for each individual case in the study; the data resembles that which you might collect yourself if you conducted a study. In social science the case is often an individual person who has responded to a survey, sometimes however the case might be a household, a school, a company or some other unit.

In order to summarise information about a large number of cases, we typically analyse these data in an analysis package like SPSS. Within SPSS it is possible to see the structure of microdata, each case has values for a range of variables. Variables are the characteristics that vary from case to case - they are typically measured using survey questions. The answers to the questions are stored as values. Because the values are usually numeric they are often labelled to make them easier to understand.

The screenshot below shows the top corner of the British Crime Survey 2007-2008: Teaching dataset in SPSS:
Each row contains a case (a person in this instance). Each column contains a variable. The shaded cell is in the row for case 5 and for the livharm variable. It contains the value 1. Livharm is the name given to the variable which gives the ONS harmonised marital status of respondents. We don’t have the labels showing here, if we did we would see that this individual is married (value 1).

We have a lot of information about each case so we can use the data in a much more flexible way than with data which has already been tabulated. With microdata we can:

- Produce tables to our own specification
- Define sub-groups to analyse separately
- Manipulate the data to regroup or combine information

Microdata are a lot more flexible and powerful than pre-tabulated data.

3.2 Selecting a dataset – survey data at the UK Data Service

The principle source of microdata in the UK is the UK Data Service. The UK Data Service provides access to and support in using thousands of datasets. Access to UK Data Service data and resources are available through the website at www.ukdataservice.ac.uk.

The Crime Survey for England and Wales (formerly called the British Crime Survey) is probably the most widely used cross-sectional survey data for researching crime in the UK along with the Scottish Crime and Justice Survey (SCJS).

Data by theme and the variable and question bank

The UK Data Service has a variables search that allows you to search for variables and questions in UK surveys. You can also use ‘Subject’ in the left hand menu in the UK Data Service’s Catalogue search engine Discover and select ‘Law Crime and Legal Systems’ to get a list of relevant datasets. If you just want an overview of the key datasets, use the left hand menu to select ‘Series’ and click Refine. You will then see something like the following list:
Note that each of the series has an icon in front of it that identifies it as a series record. The icons to identify other kinds of records can be found at the top of the search results. You can obtain those kinds of records by using the left-hand menu ‘Type’.

Theme pages

You can also find theme pages about Crime data on the UK Data Service website via Get data > Data by theme. The Crime theme provides a list of key crime data available via the UK Data Service as well as advice about analysing crime data, research examples and useful resources.

Useful, eh? You can use these references to better understand the measurement of the topic you are interested in. Understanding measures and their strengths and weaknesses will help you to become a critical user of survey data.
Other ideas for finding data:

1. Searching your local data archive: the UK Data Service catalogue is the most comprehensive way of searching for UK data click, you can do this from the UK Data Service home page at http://ukdataservice.ac.uk/. The CESSDA (for Europe, www.cessda.org) or IFDO websites (www.ifdo.org) are helpful in locating data archives in other countries.

2. Other guides to data available for specific themes can be found on the UK Data Service website. A crime theme area on the UK Data Service website can be found under Get data > Data by theme > Crime and Social Control:
   http://ukdataservice.ac.uk/get-data/themes/crime.aspx

3. UK Census data is available through the UK Data Service Census support pages:
   http://census.ukdataservice.ac.uk/

4. Looking for data from data collectors: If you have drawn a blank, national statistics organisations and other major social survey organisations may be able to provide you with information about datasets that they have collected.

5. Does a dataset exist? A literature search on your research topic is a good way to find out about data availability. If a major data source is available it is likely that someone will already have used it. A good knowledge of the literature will help you to identify the data sources that are available.

Try these steps to find a UK survey dataset for a secondary analysis project of your choice
- Look on the UK Data Service website to see if any theme pages or guides to your research topic have been produced.
- Look up a keyword for the topic that you are interested in, using the UK Data Service catalogue search engine Discover.

➔ Go to the UK Data Service website at www.ukdataservice.ac.uk

➔ Type a keyword into the DISCOVER UK DATA SERVICE search on the right hand side and click GO.

➔ You will get a lot of hits! There are a lot of data in the UK Data Service catalogue! To refine your results, you can use the Type or Date menus on the left, or to sort the results by Date, Title or Most downloaded data, use the sorting box on the right. You can select only Series in the Type tab and this will select only the widely used survey series.

Section 3 Key Points

- Microdata contain detail for all units from which data were collected (typically individual respondents to a survey). They are more flexible than summary tables.
- Most major social surveys are available through the UK Data Service, which has a website at: www.ukdataservice.ac.uk
- UK Data Service provide useful theme pages
  - You can also find survey data by looking at the UK Data Service catalogue, Discover.

Throughout this workbook we will be using the British Crime Survey 2007-2008 Teaching Dataset. This is a subset of data from the British Crime Survey dataset which is available for researchers. It is simply a cut down simplified version of the data which is easier for learners. The data are real data collected from respondents to the survey and will produce valid results.

In order to be a critical researcher it is important to understand where the data come from. The documentation that comes with the Teaching Dataset will help a little, but it will also help to look at how the original dataset was created. Appendix 1 contains summary information about the BCS 2007-2008; you should read this before analysing the data. This information is drawn from the documentation that is available from the UK Data Service. Each UK Data Service dataset has an online catalogue entry and may also have a series web page. Through these, you can access documentation. You can access the catalogue entry by searching Discover, the UK Data Service search engine, or you can take a shortcut.

➔ Go to the UK Data Service website at www.ukdataservice.ac.uk
➔ Go to Get data > Key data and scroll down to find the Crime Survey for England and Wales
➔ Click on the link to open a page for the whole Crime Survey for England and Wales (CSEW) series including the British Crime Survey (BCS) (the old name for the CSEW)

This will take you to a webpage which lists all the CSEW/BCS datasets available for downloading (see the DATA ACCESS section). This will lead you to the catalogue page for each CSEW/BCS survey so you can explore all the documentation before downloading any data (for example to check if a certain question is asked or if a variable is available for any/all years). The series webpage also gives information on RELATED RESOURCES.

➔ Under DATA ACCESS, find the British Crime Survey 2007-2008 and click on the link to go to its catalogue page
➔ Look at the Primary investigators, sponsors, abstract, coverage, universe sampled and methodology sections to answer the following:

• Who were the primary investigators on the BCS 2007-2008?
• Who paid for it?
Respondents are asked about their experiences of crime for what period?

Which countries are covered by the BCS 2007-2008?

How many people were interviewed in each sample household?

What methods of data collection were used?

Scroll down to the bottom of the screen to the documentation box

Open up the technical documentation part 1 and see if you can find the sampling information

Scroll to the end of the technical documentation part 1 to find the questionnaire and scroll through the administrative information to the first question respondents are asked (on p3)

What is the first question asked?

Return to the Crime Survey for England and Wales series page

Scroll to the bottom of the page to RELATED RESOURCES

Click on the link for the BCS 2007-2008: Teaching Dataset (SN 6561)

You should see a web page that looks like this:

---

**British Crime Survey 2007-2008: Teaching Dataset**

**Persistent identifier:** 10.5255/UKDA-SN-6561-1

**Depositor:** University of Manchester, Cathie Marsh Centre for Census and Survey Research, ESOS Government

**Principal Investigator:** Cathie Marsh Centre for Census and Survey Research, ESOS Government

**Original data producers:** Home Office Research, Development and Statistics Directorate, NFER, Social Research

**Acknowledgements:** The Economic and Social Data Service (ESDS) were funded to compile this teaching dataset by the Economic and Social Research Council (ESRC) and Joint Information Systems Committee (JISC).

The Home Office funded the main British Crime Survey, 2007-2008, (BCS) from which this dataset has been created.
Look through the catalogue entry to answer the following:

Does the catalogue entry explain which data file the teaching dataset is drawn from? What is the Serial Number (starting SN) for the dataset from which the dataset is drawn?

How many cases are there in the teaching dataset?

Is it necessary to use weights? This is explained further in a moment!

What documentation is available at the bottom of the page?

Open the PDF user guide by clicking on its link

Search the documentation to answer the following questions:

Search on “weight”: why are weights provided? Will results be biased if weights are not applied?

Search on “missing values”: how might missing values arise in the data?

Search on “fear of crime” to find fear of crime in the codebook. What is the name of the variable which contains information on whether the respondent feels safe home alone at night?

Which weight variable should be used to weight this variable?

Search on the name of the variable which measures fear of being home alone at night to find details of this variable.

You should have found something that looks like this:

<table>
<thead>
<tr>
<th>Number</th>
<th>Home alone</th>
<th>Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>homealone</td>
<td>how safe do you feel when alone in home at night</td>
<td>7480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very safe</td>
<td>3301</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fairly safe</td>
<td>681</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a bit unsafe</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very unsafe</td>
<td>6</td>
</tr>
</tbody>
</table>

Some key points about using the Teaching Dataset are in Appendix 2 for easy reference.
You can access the documentation for a dataset of your choice by going to the relevant UK Data Service catalogue entry.

For your choice of dataset:

- Locate the dataset in the catalogue
- Read the study description in the catalogue entry
- Locate documentation and think through the following questions:
  - Will the population sampled in the survey allow you to describe the population that wish to describe?
  - Is the sample population appropriate?
  - What questions are asked that will be of use to your research question?
  - Are the questions you want answered, answered by the group you want to describe?
  - How are the answers to questions held as variables in the data? (i.e. What variables are in the data, what values do they have)
  - Do the variables come directly from the questionnaire – or have they been derived (if so, can you work out how?)
  - Do you have a large enough sample?
  - Do you need to weight?

**Section 4 Key Points**

- UK Data Service datasets have a catalogue entry that provides a short description of the data
- A fuller explanation of what is in the data and how it is created will be available in the documentation
- The documentation describes the type of sampling, whether weighting is necessary and lists the variables.
Section 5: Accessing survey microdata from the UK Data Service

This section gives information on accessing and downloading surveys from the UK Data Service. For demonstration purposes we will be accessing the BCS 2007-2008 Teaching Dataset. However, the process is the same for the other surveys.

5.1 Downloading the survey data

To access and download UK Data Service survey data, all users must Login/register with the UK Data Service.

Registered users can download the datasets direct from the UK Data Service web site www.ukdataservice.ac.uk (usually in SPSS, STATA or tab-delimited formats) via each dataset’s catalogue page. You can find the catalogue page either by using the UK Data Service search engine Discover, or for some of the large, long-running dataset, via the series pages that you can find via Get data> Key data.

Why do I have to register?

CSEW/BCS data, like most other datasets used by secondary analysts is only available to users who agree to certain conditions of use. These conditions require you to act responsibly and ethically with the data – in particular you promise to respect the anonymity of respondents. You must make this agreement personally, so you cannot pass the data to someone who has not agreed to these conditions. In other words, you must not share data with friends.

Registering also helps the services who make the data available to understand who is using the data, this helps them to ensure that data continue to be made available.

Registration is part of your ethical responsibility in secondary analysis.

You can register to use the data by trying to access a dataset that you are interested in. We will register to use the British Crime Survey 2007-2008: Teaching Dataset. You can do this by going to the catalogue entry in the same way that we did in Section 4 to read the documentation.

You can access this page by:

Going to the UK Data Service website at www.ukdataservice.ac.uk
Using the UK Data Service catalogue search engine Discover to search for “british crime survey teaching dataset”. Scroll down until you find the British Crime Survey 2007-2008: Teaching Dataset.

Clicking on the Explore online link allows you to explore the data using Nesstar.

Clicking on the dataset title takes you to the dataset’s catalogue page.

On the catalogue page, there are links to Download/Order the dataset. There are also links to KEYWORDS, DOCUMENTATION, RELATED STUDIES AND GUIDES, PUBLICATIONS and a list of VARIABLES.

This page gives lots of useful information on the dataset, including information on the Principal Investigators, Depositors, Original Producers, Sponsors, an Abstract about the CSEW/BCS, the Main Topics, coverage and geography, observation Units, the survey population and some methodology details, such as number of units obtained, method of data collection and weighting. Detail about the how the data were collected, the variables and their coding and many other things are in the DOCUMENTATION section normally in searchable (Control-F) pdfs.

It is very important to read through this information before downloading or accessing any survey data in order to get a feel for the data to make sure you understand broadly when and how it was collected, and the main topics asked and how many observations were obtained.

Click on link next to the shopping basket icon labelled Download/Order near the top right of the catalogue page.

This will take you to the login screen

Enter your username and password

This brings you to a screen which lists all the datasets held at the UK Data Archive that are associated with your account.

Click on ‘register new use of data’

To register a new use of data fill in the boxes that are required, for example, Intended use of data, use the drop down box to select Non-commercial.
In Usage Title you can enter anything that gives an acceptable concise description of what you will be doing with the data. For example, you could simply add in ‘Secondary Data Analysis for Dissertations’.

In Brief Description of Usage, again, anything sensible and relevant will do.

Register a new use of data

You must register all new uses of the data, including re-use of data already supplied to you, but for a different purpose. Please note that each usage will expire two years from the date that it was registered. However, you can apply to extend the expiry date.

If you are working on an existing usage and need additional data go straight to usage details.

Intended use of data

Commercial use is usage undertaken within or under the direction of a commercial organisation for commercial or profit-making purposes. Please note that if you do not choose the correct intended use, you may be in breach of the End User Licence that you agreed to when registered with the ESRC/UKDA.

Usage title

Secondary Data Analysis for Dissertations

Once you have finished registering the new use of data you will be able to add the BCS 2007-2008 Teaching Dataset to your order.

Click on Download (we will consider Explore Online on the next page)

This will bring you to the screen below, where you will have to choose the format in which you want the dataset in.
Click on SPSS

You can choose Open or Save

It is important to read through all the documentation before starting any secondary analysis using SPSS. For example, if you do not read through the documentation before you create tables or present statistics then all your findings will be wrong as you will not know the correct weighting variable to use.

5.2 Exploring survey data online using Nesstar

An increasing number of datasets are also available in the Nesstar system, which will allow you to explore the data online before registering. You can look at frequencies for variables without registration, but after registration you can weight the data, look at crosstabulations, produce graphs, download the data or a subset of the data and conduct simple analyses. Nesstar is a useful resource particularly for those who do not have access to a statistics package like SPSS.

Search for the dataset in the UK Data Service’s search engine Discover and in the search result for that dataset, click on Explore online.
This takes you to the Nesstar Catalogue where you can browse all the datasets that can be explored online.

This takes you to the following page:

In the left hand pane, under the British Crime Survey, 2007-2008: Teaching Dataset (not the unrestricted access version though!), click on Variable Description (or press the + button to its left) for a list of groups of variables in that dataset. Keep selecting the groups until you get to the variables themselves and then click on a variable. You will see information about that variable appear in the right hand pane.

For demonstration purposes only we will have a brief investigation of the variable ‘crimerat’ from the question on: Change in crime rate over the last two years.

Click on the ‘+’ to expand the menu for Respondent Opinion About Anti-social Behaviour and Crime in Their Area.

Click on ‘how much crime rate has changed in the area since 2 years ago’

This produces the following frequency table which gives the counts for all the different categories, including information on missing values and who the question was asked to (Universe). It also provides some documentation about the source of the question.
What proportion of responses indicated that they thought that the crime rate in their area had risen a lot in the last two years?

What is the wording of the question that was asked?

Was the question asked of everyone in the survey, or just a specified group? (Hint: look at the “Universe”)

You can click on any of the variables to get a frequency of how it was answered; however, the frequencies are not produced using the weighting variable (details of how to weight using Nesstar are given next).

(2) Accessing & downloading the survey data (when registered) via Nesstar

Now we know a little about the BCS 2007-2008 Teaching Dataset, we can do some real analyses. By way of demonstration we will look at the responses to the crime rate perception question, but this time we will apply a weight to ensure that the results are unbiased. For more information about weighting surveys, see the What is Weighting? guide.

⇒ Click on the small scales symbol on the top right hand side of the screen.
This brings up the following screen which lists all weighting variables associated with the dataset (as there is more than one) and gives information on the weights.

![Weighting variables screen]

- Click on *Weight to be used when analyzing individual-level data (mean=1)* in the ‘Weighting variables defined in the dataset’ box.

- Click on the ‘>’ to move it over into the ‘Weighting variables selected’ box and click on *Ok*.

This will bring up the Tabulation screen shown below in which we will create a table. (If you don’t see the tabulation screen, click the blue tab at the top of the screen labelled TABULATION).

![Tabulation screen]

In general, frequencies and tables should be weighted before quoting percentages from them. To apply weights, select the Weight icon and choose one or more weight variables to use. Users should consult the associated documentation for further information on whether, and how, to correctly weight the data.
→ Click the ‘+’ menu for Respondent Opinions About Anti-social Behaviour and Crime in Their Area to expand it and then click on how much change in crime rate since 2 years ago.

This brings up the options to ‘Add to row’, ‘Add to Column’ etc.

→ Click on ‘Add to Row’

This takes you to the screen where you are now required to enter your username and password. Once you have logged in the Tabulation screen below should appear again updated with the information for the variable.

Note that the words “Weight is on” appears under the table.

Can you spot any differences between the percentages given in this table and those you produced earlier without weights?

For demonstration purposes only (to give an example of using Nesstar) we will investigate this by sex, in order to shown that you can easily create two-way tables using Nesstar without having to download the dataset and using SPSS or the microdata.

→ Break down the responses to the crime rate perception question by sex; by clicking on the ‘+’ menu for Socio-demographic Variables to expand it and then click on ‘Respondent sex’
This time click on ‘Add to Column’

This will produce the following table:

From the table you can see that the weighting variable is still on and you are able to report the percentages (and the counts in which they were based) in your dissertation.

Further to this you can also easily export the table to Excel:

Click on the Excel icon on the top right hand side of the screen

This will bring up the dialogue box below:

Click on Save, and then save the file with an appropriate name to an appropriate location.
The resulting Excel file looks like this:

![Excel file screenshot]

You can easily amend the Excel file to produce a suitable table or chart for inclusion in your dissertation.

In Nesstar you can also produce graphs.

→ Click on the chart button 📊 to produce a graph of the information in your table.

![Graph screenshot]
Furthermore, if you are a registered user you are able to download all, or a subset of, the data. Nesstar can save data into formats suitable for SPSS, STATA, SAS, Statistica, DIF (suitable for use in Excel), Dbase and NSDStat formats.

⇒ Click on the Disk icon on the top right hand side of the screen to bring up the screen below where you can select the format for the dataset and also create and download a subset of the data.

However, we have already described how to download the dataset from UK Data Service so we will not give an example of how to download in Nesstar.

For further information on using Nesstar search Discover for ‘Nesstar’ and in the Type menu, select Support/ how to guides.

Nesstar is a great way of exploring a dataset and its documentation, and of producing very simple analyses quickly. You could use it to get simple statistics for a set of datasets over time to produce a line graph to look for trends. Don’t forget however, you will still need to read the documentation and understand how data is collected for each dataset you use!
Once you have registered for the UK Data Service and registered a use you can access other datasets.

If you have not previously registered for the UK Data Service and you try to access a dataset you will automatically be directed to the registration system.

Try to download a dataset you want to explore for your dissertation:

- Go to the catalogue entry for that dataset then start exploring it:
  - Click on Explore online in the top right hand corner of the catalogue record if Nesstar is available (not all datasets are available in this format)
  - Click on Download/order if you would like to download the entire file
  - Fill in your username and password, or complete the registration process if you have not yet done so
  - Remember that if you download a file you may have to unzip it in order to make it work
  - If you discover that the SPSS file is a portable file (.por) rather than a .sav file – you will need to open it in SPSS by ensuring that file type is set to .por rather than the default .sav (it won’t make any other difference)
- You can now start looking at variables to check that you have:
  - The data that you were expecting
  - Enough cases

### Section 5 Key Points

- Users of the UK Data Service and many other secondary datasets need to register to access the data
- UK Data Service registration is online and quick
- You can download a copy of the data in SPSS (or Stata) format from the catalogue entry for that dataset
- Many datasets can be explored online using Nesstar, it is possible to look at frequencies and documentation without first registering for the data.
Section 6: Getting started with the BCS 2007-2008 Teaching Dataset

This part of the workbook will provide practical examples of how to start secondary analysis of the BCS 2007-2008 Teaching Dataset in SPSS.

This section illustrates how to open a dataset and how to start exploring variables. While it is assumed that you will have some familiarity of using SPSS, screen shots are provided to demonstrate what is required. For a more detailed introduction to SPSS see Chapter 2 on The SPSS Environment in Discovering Statistics Using SPSS by Field (2005) or SPSS for Social Scientists using the 1998 BCS by Miller et al (2002).

6.1 Opening the BCS 2007-2008 Teaching Dataset

There are several options available to open the dataset once it is downloaded and unzipped, these include:

- Double-clicking on the icon in the folder where it was unzipped to (this open may require choosing SPSS as the program).

- Using Windows> Start> Programs> SPSS, this option usually brings up the dialogue box shown below, where you can click to find the dataset (it also displays recently used datasets). You can simply close this by clicking the Cancel button.
Ensure that the Data View tab at the bottom of the window is highlighted and then click on the Open File Button to open your dataset. This will open the open file dialogue box (similar to opening files in Word or Excel).

6.2 SPSS Basics

SPSS runs in two or more windows. The two main windows are the Data Editor window and the Output Viewer window. More serious users will also use the Syntax Editor window. Other windows will open for specific purposes such as for editing a graph.

The Data Editor window is where you will type in data (if for example, typing in questionnaire results once codes have been applied) or where datasets are displayed. There are two views, the Data View and the Variable View. The Data View is shown below:

Menu Toolbar

Rows:
• Cases
• Observations

Columns: • Variables • Attributes

<table>
<thead>
<tr>
<th></th>
<th>r label</th>
<th>r adults</th>
<th>r child</th>
<th>sex</th>
<th>age</th>
<th>agegrp7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61302140</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>61384010</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>63880500</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>63890200</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>63845810</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>38</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>71109020</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>71109140</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>86</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>71109020</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>58</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>71108080</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>38</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>71108100</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>29</td>
<td>2</td>
</tr>
</tbody>
</table>
Ensure that you have the Data View selected

The rows in the Data Editor window are usually referred to as cases (or observations) in statistical software packages. Here (in the BCS 2007-2008 Teaching Dataset) each case/observation is an individual who has responded to the survey (but in other circumstances could be a family or other unit). Each case in this survey has an ID number called rowlabel. Each row is also numbered in SPSS.

Press Control-↓ to go to the last case in the dataset

How many cases are there in the dataset you have opened?

Does this accord with the number of cases that the documentation for the dataset says should be there?

Each column relates to a variable. These variables (or attributes) might be each person’s age, sex or income. In the BCS 2007-2008 Teaching Dataset, for example, shown above we have variables ‘nadults’ (number of adults in household), ‘nchil’ (number of children under 16 in household), and ‘sex’. The names of the variables are at the top of the columns; you can see the variable label by running your mouse over the variable name you are interested in:

<table>
<thead>
<tr>
<th>nchil</th>
<th>sex</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of children under 16 in household</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

What is the variable label for the variable rubbcomm? Use the go to variable icon.

Scroll across to the last variable – what is the name and label of the last variable in the dataset?

In each cell are values which represent answers given to each question (or information derived from answers). These are typically stored in SPSS numerically but value labels are defined so us non-computers can see what’s what. By default your software probably shows the labels rather than the values. You can toggle between value labels and values by clicking the label button in the menu. You can also change the widths of the columns if you want to see more or less, this is particularly useful if the current width of the column is narrower than the value labels.
What is the value for rubbcomm for the last case in the data?
What is the associated value label for rubbcomm for the last case in the data?

More information about how the variables have been set up is in the variable view.

Click on the Variable View tab

In the Variable View tab, information can be obtained about each variable in the dataset. Each row includes information about a variable, for example its name (invariably a ‘shorthand’ description), data type, width (number of characters displayed in the column in the Data View), decimal places being displayed, the variable label (a more meaningful description than the name) and the value labels assigned to each level of a categorical variable (e.g. 1 = male, 2 = female).

The missing column is particularly useful. It contains information about which values have been set to missing. Values that have been set to missing will be excluded in most analyses.

What values have been set to missing for the variable came?
What values have been set to missing for the variable work?

Click on the Output Viewer window
In the Output Viewer window (shown below) there is a slightly different choice of menus and tools and there are two panels in the main part of the screen. The left hand panel has a windows explorer-like tree which chronologically lists the operations/commands carried out in any session using SPSS. You can use this to navigate around your output. The right hand panel will contain any tables, graphs or results of statistical analyses which are output as a result of any commands. All of these can be edited, deleted or copied and pasted into other packages (for example, so that you can include parts of your output in a Word document).

You can perform most analyses from either the output window or the data editor. So let us take an opportunity to undertake a procedure! We’ll look at respondents’ sex in the first instance.

→ From the menu click Analyze
→ Select: Descriptive Statistics> Frequencies...
The frequencies dialogue box appears, scroll down the variable list to find the variable sex. Click on it.

Press the arrow button to transfer the variable to the selected list on the right hand side.

Click OK.

As you are already in the output viewer you should be able to see the result of running this procedure immediately.
Note that:

- The output has been added to what was already in the output viewer
- New bookmarks have been added to the left hand pane to help you navigate around your output as it becomes bulkier

If you want to save your output you need to click on the save button. Just like a word file save output frequently and remember to save before you quit the programme (SPSS will prompt you when you exit).

### 6.3 SPSS Syntax

You will have noticed that SPSS commands are run using a Windows point and click interface. Each procedure you request using the menu activates a command which can be recorded in command language called **syntax**. You may well already have seen some of this without knowing it (depending on how your computer has been set up).

Your output window should now look like that on the previous page. Above the results a record of the command run is available. It says:

```
FREQUENCIES VARIABLES=sex
/ORDER= ANALYSIS .
```
Recording your commands is useful to help you understand your output. You can also paste the commands into a window called the syntax editor to re-run them.

If your computer does not record the commands as you go, you can fix this by changing an option.

1. From the menu select: Edit> Options...
2. Make sure that the Viewer tab is selected
3. In the Initial Output State section check “display commands in the log”

We will now use the SPSS syntax editor now to run a command.

1. Open a syntax window by clicking on File> New> Syntax in the menu

2. Type or paste the following commands into the empty syntax editor
FREQUENCIES VARIABLES=sex
/OPTION= ANALYSIS.

Hold your mouse button down to select the 2 commands (as long as part of each line is selected it’s OK)

Hit the run button

With any luck you’ll have created a frequency table in the output viewer.

New users of SPSS generally prefer to work by using menus. More experienced users prefer to write (or generate, using the menus) syntax. The advantage with that way of working is that it allows you to keep all of your workings together and re-run them when you (invariably) change your mind or spot a mistake.

The following information on SPSS syntax in the box is optional, while it is really useful to do when using SPSS it is not a necessary requirement. At various stages throughout the rest of this workbook examples of using SPSS syntax will be provided in boxes labelled Quick Tip.
Syntax Quick tip!

The Syntax Editor window will only be present if you have asked for the commands to be pasted when using a dialogue box, or if you have asked opened a SPSS syntax file (with the extension .sps).

The Syntax Editor is a useful way of making a record of any commands that you have made. This ranges from opening datasets, recording variables to statistical analysis.

Syntax is very useful, in cases where you have recoded a variable, for example the variable ‘Agegrp7’ (age grouped) from 7 into 5 age groups and then have run some exploratory data analysis (EDA) using the new age variable and then later decide that it would have been better to recode ‘AgeGrp’ into 4 age groups. When using syntax in the Syntax Editor you can simply edit the lines of syntax and ask SPSS to run the lines of commands. This save trying to remember what was previously done in the recode and also saves time having to go through using the menus and reselecting many different options all over again.

The figure below illustrates how to get SPSS to paste commands into the Syntax Editor (using the example of creating a frequency table of the variable ‘sex’ (more on this later!). In the normal dialogue box there is a Paste button (under the Open button).

Once Paste has been pressed this brings up the Syntax Editor which contains the command to create the frequency table (and a pie-chart)!:
To run the command you will need to press the run button or you can use the Run menu to get SPSS to run the command. If you have many commands then you can simply highlight the selected commands and press run, or go to the Run menu and there are four options: All, Selection, Current or To End.

If the SPSS syntax file is saved then it can be used to record all commands made. You can also write comments to remind yourself why you were performing the command/what a command is doing. This is really useful because when you look back at your syntax you know exactly how you prepared the dataset and any EDA or statistical analysis performed. To make a comment, as oppose to a command, it is necessary to enclose the text within a * (as shown below).

It cannot be stressed enough that SPSS syntax is an incredibly useful way of keeping a record of commands and saving time when rerunning recodes or analysis. In addition, command language can be used instead of using the Paste button in the dialogue boxes. Information on all SPSS commands are included in the SPSS user guide Command Syntax Reference found in the Help menu. This gives detailed information on all commands.
6.4 Getting ready for EDA using the BCS 2007-2008 Teaching Dataset

Before exploring the dataset it is useful to know about the Variables tool which opens up the Variables window which displays all the information on the variables including how the variable has been coded (including information on missing values).

To display this information either click on the icon in the tool bar or go to the menu: Utilities> Variables...

The Variable window is shown below for the variable ‘homealon’

What is the value label for value 9?
Are higher valid (i.e. non-missing) values associated with more or less fear about being home alone at night?
Which values have been set to missing?

Display the variable information for wmugged
What is the variable name and label?
What values have been set to missing?
This window does not explain where the data came from, or what it means. How would you obtain this information?

Remember that you can find out about the data by looking at the documentation. For datasets which are teaching subsets, you may also need to look through the documentation for the dataset from which the teaching data were drawn.

Looking at unweighted frequencies

➔ Use the same approach described previously to:
➔ Obtain the unweighted frequencies for ‘cartot’,

Normally you would expect the unweighted frequencies in your data to match those in the documentation (if not, check that you’ve got the right file and haven’t advertently selected only part of the data – if you get inexplicable differences contact the service who made the data available, they may be able to help – or you may have stumbled across a problem with the data that needs fixing!)

Do these frequencies you have produced match those given in the documentation for the BCS teaching dataset?

Watch out for variables with a large number of missing values. These variables may only be applicable to a subset of cases, or there may have been a large number of respondents who refused to answer the question.

Which of the variables have the lowest number of missing values?
Which of the variables have the highest number of missing values?
In order to check why a variable has a large number of missing values you can look at the questionnaire. Remember that the questionnaire may be in the documentation for the BCS 2007-2008 dataset (rather than the teaching dataset).

To whom is the ‘cartot’ variable applicable?

It is possible to make systematic changes to the data in order to deal with problems like this. We will discuss this in further sections.

6.5 Weighting a dataset

**NOTE:**
Before starting any EDA it is very important to remember to add the correct weighting variable. For the purpose of this work book the analysis is at the individual level so the weight we use is ‘tcindwt’ with the Variable Label ‘to be used when analyzing individual-level data (mean=1).’

To apply the weighting variable either press the weighting button or select the menu: Data> Weight Cases... and the weight cases dialogue will appear where you can scroll down the list of variables to find ‘tcindwt’.
To check that this has worked (using the menus or syntax), look at the bottom right corner of the Data Editor where it should display ‘Weight On’.

Rerun the frequencies of cartot

Are there any changes in the percentages obtained now that weights have been applied (as compared with before)?

Be aware of the effect of grossing weights!

In this teaching dataset, the weight has been adjusted so that its mean is equal to 1. In the original BCS 2007-2008, the weight when applied grosses frequencies up to the population of England and Wales. When this happens, weighting can give us the impression that the data are more reliable than they actually are because we see large numbers in each category. In order to avoid the inflation of counts due to grossing it is useful to rescale the weight back down to the correct sample size. This is done by
finding the mean of the weights and then for each of the individual cases we need to divide the individual case weight by the mean weight value.

*Sounds complicated? Don’t worry!*

SPSS can calculate means for you. We will find out later how we can compute variables.

---

**Syntax Quick tip!**

Instead of clicking OK, try clicking on the *Paste* button and making a record of what you have done so far (including adding in a note using ‘*’ text ‘.’ to distinguish the note from the commands).

You can highlight the Weight command and press the run button.

---

It is always advisable to undertake some Exploratory Data Analysis (EDA) on a dataset in order to get a feel for the data before considering any recodes or further data analysis.

### 6.6 Properties of variables: looking at variable distributions

What is a Distribution? The distribution for a variable is the pattern formed by the values of that variable. The first tasks in exploratory data analysis is to look at the ‘distributions’ for key variables of interest. This refers to the way individual observations are distributed across the possible values of a variable (e.g. for variable sex and age).

- Distributions can be described and summarised using various techniques:
  - with frequency tables and descriptive statistics
  - graphically
- The choice depends on the type of variable (categorical or continuous/interval) and what aspects of the distribution you want to show.

Assuming that the weighting variable (‘tcindwt’) has been applied the EDA can begin. All analyses will be weighted until you turn the weight off.
We will now focus on categorical variables (i.e. those with values stored as categories). These are common in survey data because so many questionnaires are of the tick-box type.

6.7 Tabulating Categorical variables

- We have produced frequencies before. Use the same approach as before:

  Analyze > Descriptive Statistics > Frequencies and a dialogue box will appear.

Given that we are interested in fear of crime and influencing factors, it is important to explore the possible variables before considering whether there is a relationship between them.

For exploratory purposes we will consider the 6 variables: ‘sex’, ‘ethgrp2’, ‘agegrp7’, ‘walkdark’ and ‘homealon’.

- These variables can be selected in the frequencies dialogue box and press OK to run these frequencies
Syntax Quick tip!

Instead of clicking OK, try clicking on the Paste button and making a record of the frequency that you have run. If you want to run these frequencies again next time you use SPSS you need only highlight the commands and press the run button.

The first table displayed in the Output window is information on the valid and missing counts.

<table>
<thead>
<tr>
<th></th>
<th>Respondent sex</th>
<th>Respondent ethnic origin (5 categories)</th>
<th>age group (7 bands)</th>
<th>how safe do you feel walking alone after dark</th>
<th>how safe do you feel when alone in home at night</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Valid</td>
<td>11676</td>
<td>11673</td>
<td>11660</td>
<td>11625</td>
</tr>
<tr>
<td></td>
<td>N Missing</td>
<td>0</td>
<td>3</td>
<td>18</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the table it is apparent there are no missing values for the variable sex, yet there are some missing values for the other variables.
Firstly, considering the variable sex (‘sex’) - there are no missing variables so every individual in the BCS 2007-2008 is recorded as either male or female. Cases (respondents) are distributed between these two outcomes. The number who are male and female, together with their associate percentages, are recorded in the frequency table for sex. You can see that there are 45.5% males and 54.5% female respondents.

**Respondent sex**

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>5307</td>
<td>45.5</td>
<td>45.5</td>
<td>45.5</td>
</tr>
<tr>
<td>female</td>
<td>6369</td>
<td>54.5</td>
<td>54.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>11676</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Next, considering the question on the respondent's ethnic group (‘ethgrp2’), which is self-rated from the list of 16 options from the original BCS 2007-2008 ‘show card’ below:

From the frequency table (below) for the variable ‘ethgrp2’ it is apparent that some of the original ethnic groups have been amalgamated (from 16 categories on the ‘show card’ to 5 different categories in the frequency table). Moreover, from the frequency table it can be seen that over 93% of the respondents in the BCS 2007-2008 Teaching Dataset are ‘White’. It is important to consider the number of respondents from different ethnic minority groups – for example there are only 134 respondents classified as Chinese or other.
From the frequency table it can be seen that the Percent and the Valid Percent are slightly different, this is due to the Missing Values. The valid percent column omits the missing values.

It is generally useful to check what the categories of the missing values are. For example, data may be missing due to the routing of the original questionnaire and if an investigation of the missing values is not undertaken then the interpretation may be wrong. In this case the number of missing cases is quite small so we need not be too concerned with these cases. However, it is always worth reporting information on the missing values in your dissertation and explaining why cases are missing as best you can.

Now we will consider some of the answers to the questions which consider fear (including fear of crime).

The first question is on ‘walkdark’ (how safe you feel walking alone after dark), the frequency table is below:

<table>
<thead>
<tr>
<th>how safe do you feel walking alone after dark</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very safe</td>
<td>3002</td>
<td>25.7</td>
<td>25.0</td>
</tr>
<tr>
<td>fairly safe</td>
<td>4718</td>
<td>40.4</td>
<td>40.6</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>2604</td>
<td>22.3</td>
<td>22.4</td>
</tr>
<tr>
<td>very unsafe</td>
<td>1301</td>
<td>11.1</td>
<td>11.2</td>
</tr>
<tr>
<td>Total</td>
<td>11625</td>
<td>99.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
don't know                                  | 51        | .4      |               |
|Total                                       | 11676     | 100.0   |               |

Again, as with the variable ‘ethgrp2’ there are a small percentage of missing values, for which most of them are ‘don’t know’ (rather than ‘refusal’).
Regarding the question on ‘homealon’ (how safe you feel when alone in home at night) from the frequency table below it can be seen that there are a small percentage of missing values, for which all were ‘don’t know’ replies.

<table>
<thead>
<tr>
<th>how safe do you feel when alone in home at night</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Valid</td>
</tr>
<tr>
<td>very safe</td>
</tr>
<tr>
<td>fairly safe</td>
</tr>
<tr>
<td>a bit unsafe</td>
</tr>
<tr>
<td>very unsafe</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Missing</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

6.8 Graphing a Single Categorical Variable

Further to running a frequency it can also be useful to visually explore the frequencies of the variable that you are interested in. Chart options can be added when running frequencies in SPSS.

To get SPSS to graph the output:

⇒ Return to the Frequencies dialogue box by repeating the previous series of selections (or click on the Dialogue recall button).

⇒ In the Frequencies dialogue box click on the Charts... button.

The following dialogue box will appear:
Both *Bar charts* and *Pie charts* are applicable to categorical data. For this example a Bar chart will be produced.

- Select *Bar charts* as Chart Type
- Select *Frequencies* as the axis measurement in the Chart Values area
- Click on *Continue* and *OK* when back in the Frequencies dialogue box (or *Paste*).

The resulting bar chart displayed in the *Output Viewer* can than be copied and pasted into a Word document.

From the chart you can see the distribution of the respondents according to ethnic group.

### Section 6 Key Points:

- SPSS works within several windows
  - A *data editor* with two views
  - An *output viewer* that contains results
  - An optional ‘*syntax*’ editor in which you can run and save commands
- SPSS materials are available online if you want more help
- Start your analysis by checking that your data looks like you expect it to
- Remember that SPSS will not tell you what the variables mean, you will have to keep referring to the documentation to inform your analysis.
- Check whether there are a lot of missing values in your variables of interest
- Apply an appropriate weight if the documentation tells you to
  - If you apply a weight and it substantially changes the counts you can rescale the weight
- Produce frequency tables to look at categorical variables
- You can also ask for bar charts to show the distribution of categorical variables
You’re registered for the UK Data Service.
You’ve got a dataset and research question.
You can start looking at your data!

Check your documentation to see whether you need to weight the data
Look at frequencies for variables of interest
• Do they look like the frequencies given in the documentation or other research sources?
• Are there a lot of missing values?
Apply the weight, if appropriate;
• Do the total counts change a lot if you weight the data?
• If the counts increase a lot, you may want to rescale the weight – or simply use the weighted percentages but use the unweighted totals as the counts.
Section 7: Exploratory analysis: using the British Crime Survey 2007-2008 Teaching Dataset to explore variations in fear of crime

The literature suggests that fear of crime (specifically personal safety) differs for different sub-populations - i.e. by sex and by age groups. For example the 1982 BCS report by Hough and Mayhew (1983) showed distinct differences in feeling safe (as measured be feeling ‘very unsafe’) for age groups.

This example requires you to undertake some data manipulation of the BCS teaching dataset to help understand what is going on.

The variables you need in order to do this are:

‘walkdark’ How safe do you feel walking alone after dark
‘agegrp7’ Age group (7 bands)
‘sex’ Respondent sex

In some cases these variables will have to be recoded to make them more manageable for checking relationships or associations between the variables. For example, recoding the variable ‘agegrp7’ into fewer age groups in order to investigate the relationship between fear (for personal safety) and area, broken down by age.

7.1 Running a frequency for the variable ‘walkdark’

This first thing to do before carrying out any secondary analysis is to explore the variables of interest. For example, running frequencies on the variables and deciding if any recoding is necessary.

Before you start

All outputs below assume that you have applied the weighting variable tcindwt.

It is always advisable to check exactly how any question was asked in the questionnaire before running a frequency on it. For the variable ‘walkdark’ the question is shown below:

65
You should be able to find this sort of information in the documentation for your survey.

Now run a frequency to check what the variable looks like.

To run a frequency, as described earlier in ‘Tabulating Categorical variables’, go to the menu Analyze > Descriptive Statistics > Frequencies...

The frequency table for the variable ‘walkdark’ is shown below:

<table>
<thead>
<tr>
<th>Valid</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>very safe</td>
<td>2915</td>
<td>25.0</td>
<td>25.1</td>
<td>25.1</td>
</tr>
<tr>
<td>fairly safe</td>
<td>4045</td>
<td>41.5</td>
<td>41.7</td>
<td>66.7</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>2604</td>
<td>22.3</td>
<td>22.4</td>
<td>89.1</td>
</tr>
<tr>
<td>very unsafe</td>
<td>1252</td>
<td>10.8</td>
<td>10.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>11625</td>
<td>99.6</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>don't know</td>
<td>46</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11672</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the table it can be seen that there are a small number of respondents (remember that the dataset is weighted) who are classified as ‘don’t know’ (accounting overall for 0.4% of responses). For the purposes of our investigation we want to keep these responses as missing values.

However, if there are a large number of ‘don’t know’ or ‘refusal’ responses and the question is an attitudinal question on a sensitive topic we may sometimes want to retain these as valid responses (as the reasons for the replies could be important in relation to the question asked). While this is not the case here, it is always worth considering the number of ‘refusals’ or ‘don’t know’ responses and the nature of the question being asked.
We will consider the distribution of the responses by creating charts.

To do this, as demonstrated in the previous section, this can be done by going to the menu Analyze > Descriptive Statistics > Frequencies… and further to this you may select Charts… > Bar Charts and selecting the Radio button for Frequencies (as opposed to percentages).
As with the frequencies table it can be seen that the largest number (count) of responses is for the category ‘fairly safe’.

Alternatively, another way to describe the variable ‘walkdark’ is to show the percentages in each age category using a pie chart.

⇒ To create a pie chart use the dialogue recall button to bring up frequencies dialogue box again, but this time in Charts... select Pie Charts and Percentages.

The pie chart above also shows the small percentage of missing values for this question.

There is not one correct ‘type’ of chart to be made when exploring the variables. For example, it may be useful to show that there are only a small number of missing values by using a pie chart and then give the overall count of cases it is based on (i.e. the number of responses). Alternatively, a bar chart of the raw counts can be useful to show how many people have answered the question.

7.2 Selecting data: Do men and women have the same fears about walking alone at night?

So far, we have looked at the fear of walking alone after dark but we have not distinguished this by sex. We were unable to distinguish respondent’s sex. If we want to compare male and female patterns it is useful to produce separate graphs for men and women. A simple way of achieving this is to use only males or females via Select Cases.
To select males only:

➔ In the menu select **Data > Select Cases**...

➔ Click the radio button for **Select If condition is satisfied** and press the **If** button

In the resulting dialogue box you can define the selection criterion

➔ As we will be using a condition on the sex variable we should first check the values of the sex variable. A quick way of doing this is to find the sex variable in the variable list, right click and click **Variable Information**
This tells us that male is coded as 1 and female is coded as 2.

If we want to select males only we need to select the cases where sex = 1.

In the select if box you can use the variable list and calculator interface to create the criterion in the criterion box. However, you may find it easier to just type “sex=1”!

Click on Continue

Notice that the condition has appeared next to the If... button.
Ensure that you don’t accidentally delete cases by making sure that you have “Filter out selected cases” in the Output area at the bottom of this dialogue box.

Click OK to apply the filter.

Once you have done this you should notice 3 changes in the data editor.
1. The row numbers of unselected cases have been struck out
2. The words “Filter On” appear in the status bar at the bottom of the screen
3. A new filter variable has appeared at the end of the data file.

Run a frequency of the variable sex to ensure that the procedures you are now running only apply to males.

Obtain a bar chart for the percentage of males who feel safe walking alone at night.

You should get something like this:

You can now repeat this procedure selecting only females to compare male and female anxieties about walking alone after dark.

Are women more likely to be afraid of walking alone after dark than men?

Remember to set select cases to select All before continuing!
7.3 Recoding ‘agegrp7’ into 3 age groups

- Before working with ‘agegrp7’ use the approach shown above to look at the distribution of ‘agegrp7’. If you’re unsure of anything remember you can
- Look at the variable view in SPSS to see how the variable has been defined
- Look at the documentation that came with the teaching dataset and original BCS 2007-2008 file to find out where data came from

You should now know that age ranges from 16 to 75+. However, the age variable has many categories and can be made more manageable for exploratory analysis if grouped into age groups.

To recode the variable ‘agegrp7’ it is useful to think carefully about how many categories you would like while keeping the categories meaningful. When recoding ‘agegrp7’ you want to keep together the age groups who are likely to share common experiences of the fear of crime. If we were to break the cases into just 2 groups, for example we might mask distinctions related to age. While there is definitely no ‘correct’ number of groups, options such as 10 year age bands, or young adult/entering the labour force, working ages and retired ages would be sensible recode options.

We will follow the approach of Hough and Mayhew (1983), by using 3 comparable age groups - young adult, other pre-retirement and 65+.

- To recode a variable select the menu Transform > Recode into Different Variables...

- Select ‘agegrp7’ from the Variable list and move it across into the Numeric Variable box (as shown above)
Then type the new name ‘age3grps’ into the Output Variable box.

Type ‘age group (3 broad groups)’ as the Label.

Hit the Change button and ‘age3grps’ will be added as the Output Variable.

You now need to specify values relevant to the recode, so click on the Old and New values button. This opens a sub-dialogue box.

Specifically for this example:

It is useful to check the value labels for each code (in other words the age groups represented by each code) before recoding.

It is also useful to write down a list of the old and new codes to save confusion, especially in cases like this when the thing that you are recoding is numerical, for example:
To start recoding from old to new values

- Click the Range radio button on the left and add the values 1 and 2 as shown below.
- In the New Value area type 1 as shown below.
- Click Add.

You’ve now defined the first new value; the new variable will take the value 1 where agegrp7 equals 1 to 2.

- For the next group type click on Range and insert 3 and 5 in the Old Value panel, ‘2’ as the New Value and click on Add.
- Repeat this for the oldest age group for 6 and 7, then set all other values to system missing as shown below:
When all recodes have been added to the Old → New box, click *Continue* which will take you back to the Recode into Different Variables dialogue box where you may click on *OK* (or *Paste* if you are working with SPSS syntax).

The new variable ‘age3grps’ should now be visible at the end of your data in the *Data Editor*.

Whenever you derive a variable you should check that the derivation has worked in the way you expected it to. A simple way to check whether a straightforward recode has worked is simply to cross-tabulate the old and new variables (cross-tabulations are described in more detail later).

To perform this check go to the menu *Analyze > Descriptive Statistics > Crosstabs...* This will produce a cross-tabulation of ‘agegrp7’ by ‘age3grps’, move the variable ‘agegrp7’ in to the *Row(s)* and ‘age3grps’ into the *Column(s)*.
From the Case Processing Summary it can be seen there are no missing values (we set values that were not in the range 1 to 7 as missing values in the recode though in this case this wasn’t necessary).

### Case Processing Summary

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Valid</td>
<td>Missing</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
<td>Percent</td>
<td>N</td>
<td>Percent</td>
</tr>
<tr>
<td>age group (7 bands) *</td>
<td>11672</td>
<td>100.0%</td>
<td>0</td>
<td>0.0%</td>
<td>11671.891</td>
<td>100.0%</td>
</tr>
<tr>
<td>age group (3 broad groups)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Number of valid cases is different from the total count in the crosstabulation table because the cell counts have been rounded.

The cross-tabulation of the 2 age group variables is shown below:

<table>
<thead>
<tr>
<th>age group (7 bands) * age group (3 broad groups) Crosstabulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>age group (3 broad groups)</th>
<th>1.00</th>
<th>2.00</th>
<th>3.00</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>age group (7 bands)</td>
<td>1692</td>
<td>0</td>
<td>0</td>
<td>1692</td>
</tr>
<tr>
<td>25-34</td>
<td>1819</td>
<td>0</td>
<td>0</td>
<td>1819</td>
</tr>
<tr>
<td>35-44</td>
<td>0</td>
<td>2153</td>
<td>0</td>
<td>2153</td>
</tr>
<tr>
<td>45-54</td>
<td>0</td>
<td>1938</td>
<td>0</td>
<td>1938</td>
</tr>
<tr>
<td>55-64</td>
<td>0</td>
<td>1842</td>
<td>0</td>
<td>1842</td>
</tr>
<tr>
<td>65-74</td>
<td>0</td>
<td>0</td>
<td>1194</td>
<td>1194</td>
</tr>
<tr>
<td>75+</td>
<td>0</td>
<td>0</td>
<td>1034</td>
<td>1034</td>
</tr>
<tr>
<td>Total</td>
<td>3511</td>
<td>5933</td>
<td>2228</td>
<td>11672</td>
</tr>
</tbody>
</table>
From this cross-tabulation you can check that all the ages have been correctly recoded into the new values for the new variable ‘age3grps’. Having created ‘age3grps’, you will need to label it to show what each code represents.

➔ Go to the Variable view in the Data Editor
➔ Click on the cell in the ‘age3grps’ variable row in the Values column.

![Value Labels dialogue box]

The value labels dialogue box will appear - it will be empty. To attach value labels to values: type the value of the variable in the Value field and the corresponding label in the Value Label field:

➔ For the first age group in ‘age3grps’ type ‘1’ in the Value box
➔ Type 16-34 in the Value Label field
➔ Then click on Add.
➔ Repeat this process for all 3 values making sure you click Add after completing the final value before clicking on OK.

If you make any mistakes you can use the Change and Remove buttons.
**Syntax Quick tip!**

To recode a variable use the ‘recode’ command:

RECODE
agegrp7
(1 thru 2=1) (3 thru 5=2) (6 thru 7=3) (ELSE=SYSMIS) INTO age3grps .

VARIABLE LABELS age3grps 'age group (3 broad groups)'.

EXECUTE .

You can also use SPSS syntax to assign the value labels, this is given below:

VALUE LABELS age3grps
1 '16-34'
2 '35-64'
3 '65 and over'.

Next, as with any recode, it is useful to check the distribution of the new variable ‘age3grps’ in order to for you to understand how the cases are distributed among the 3 age groups and also so that you can describe your data.

⇒ To do this, as demonstrated earlier, this can be done by going to the menu Analyze> Descriptive Statistics> Frequencies... and further to this you may select Charts...> Bar Charts and selecting the Radio button for Frequencies (as oppose to percentages).

The frequency table shows that the majority of the cases (50.8%) are in the 35-64 age group. Whereas, there are only 30.1% in the 16-34 age group and 19.1% in the 65 and over group.

<table>
<thead>
<tr>
<th>age group (3 broad groups)</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>3511</td>
<td>30.1</td>
<td>30.1</td>
<td>30.1</td>
</tr>
<tr>
<td>16-34</td>
<td>3511</td>
<td>30.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-64</td>
<td>5933</td>
<td>60.8</td>
<td>60.8</td>
<td>60.9</td>
</tr>
<tr>
<td>65 and over</td>
<td>2228</td>
<td>19.1</td>
<td>19.1</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>11672</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This is shown visually by the bar chart:

\[\text{age group (3 broad groups)}\]

Alternatively, another way to describe the new ‘age3grps’ variable would be to show the percentages in each age category using a pie chart.

To do this, use the dialogue recall button to bring up frequencies dialogue box again, but this time in Charts select Pie Charts and percentages.
It should be further stressed that there is no correct way to recode the variable ‘agegrp7’ and it is quite reasonable to recommend recoding it again into 4 categories to breakdown the 35-64 (established working age) group. However, for the purposes of this example this will not be done since these were the age groups used by Hough et al. (1983).

**Question to think about!** Do you agree with the recode of the variable ‘agegrp7’? How would you have recoded it?

**Question to think about!**
Do you agree with leaving the variable as it is? How would you have recoded the variable ‘walkdark’? Would you have considered grouping other responses – for example, ‘very’ and ‘fairly’ into ‘safe’ and ‘a bit unsafe’ and ‘very unsafe’ into an ‘unsafe’ category? Think about the consequences of recoding into fewer categories.

Next we will create a summary variable (using the `compute` command) which summarises the level of fear of crime across a range of variables. However, first we need to carry out some more recodes.

- **Recode the variables wburgl and wattack to take the value 1 if the respondent is concerned or very concerned and 0 if not.**
- **Remember to start by obtaining the frequencies for each variable.**
- **Remember you can check the values and associated labels in the variable view.**

Use the following recode:

```
wburgl 1,2 = 1; All other values = 0 into wburgyn
wattack 1, 2 = 1; All other values = 0 into wattyn
```

Use this scheme for now so that you can check your results are the same as ours. However, you may feel that there is a better way to produce a yes/no classification of whether the respondent is worried. Note that we have treated don’t knows and not applicable in the same way as ‘not worried’. Can you think of another way that we could have handled this?
Remember to check that the recode has worked by cross-tabulating the old and new variables.

Syntax Quick tip!
To recode a variable use the 'recode' command:

```spss
RECODE agegrp7
(1 thru 2=1) (3 thru 5=2) (6 thru 7=3) (ELSE=SYSMIS) INTO age3grps .
VARIABLE LABELS age3grps 'age group (3 broad groups)'.
EXECUTE .
```

You can also use SPSS syntax to assign the value labels, this is given below:

```spss
VALUE LABELS age3grps
1 '16-34'
2 '35-64'
3 '65 and over'.
```
7.4 Using compute to create a variable based on multiple characteristics

We have already conducted a compute procedure in Section 6. We were able to perform an arithmetic function on the weighting variable. Now we will use the compute procedure to combine information from more than one variable.

We have already created two variables each with a yes/no indicator of the fear of crime. Each variable took the value 1 if the respondent was worried. We can combine these variables to indicate whether the respondent was concerned about 1, 2 or neither of the crimes.

To do this:

➔ From the menu select Transform > Compute Variable...

Another dialogue box appears to allow you to build the function that you want to perform. In this case, we’ll be able to create the variable that we want simply by adding the two variables together. This will give us a 0 if they were not worried by either burglary or being attacked, a 1 if they were worried by one but not the other and 2 if they were worried by both.

The dialogue box has an area on the left to give a name and label to the new variable that you are creating.

➔ Type “fearnum” as the name and “count of burg/attac” as the label.

➔ On the right hand side, build (or type) the function wburgyn + wattyn

➔ Click on OK to run this.

You should now have a new variable fearnum!
Syntax Quick tip!

To compute this variable in syntax:

```plaintext
COMPUTE fearnum = wattyn + wburgyn.
VARIABLE LABELS fearnum 'count of fear burg/attac'.
EXECUTE.
```

Can you think of a way that you can check that your variable compute worked OK? (Hint you can produce 3-way cross-tabulations!)

Can you find a way to add value labels for 0 ‘neither’ 1 ‘1’ and 2 ‘both’?
Use this variable and the select cases procedure to work out whether men and women are more or less likely to be concerned about both types of crime.

Can you find away to summarise how safe people feel by combining information from the walkdark and homealon variables?

Watch out!
When you combine variables in this way, you need to be careful about what happens when cases are missing. In this example we had very few missing cases, and we treated them as not worried. You should ensure that cases without valid values for the original constituent variables behave as you want them to.
You’ve got a dataset and research question.
You’ve identified some variables that are useful.
You can work with these variables to make them more appropriate.

- Check your documentation to see where the variables have come from.
- Check the values and value labels for your variables. Be particularly aware of the presence of missing values.
- If a variable has too many categories, think hard about how you would like the categories to look before you attempt to recode it.
- If you want to do some sort of calculation you can use the Compute command.
- Remember that its possible to do recodes and computes for some of the cases (as defined by some criterion).
- Microdata gives you a lot of flexibility – keep your research question in mind, and work out what you’d like to be able to show with your graphs and tables. This will help you to work out what you would like your data to look like!

**Section 7 Key Points:**

- To compare groups, or limit your analyses to a subgroup you can use the Select Cases procedure.
- Remember that if you select only a small group you may make your analyses unreliable – avoid analyses of small groups.
- Manipulating data allows you to perform more powerful analyses.
- Recoding allows you to change the values of categories
- Computing allows you to create variables on the basis of some function.
- Always create a new variable so that you can check it against the old variables you used to create it.
- Be especially careful around missing values!
Section 8: Further exploratory data analysis

In previous sections we have seen how to:

- Apply weighting
- Produce one-way frequencies and bar charts
- Produce one way tables for subsets
- Undertake data manipulation to recode and compute new variables
- Cross-tabulate old and new variables

In this section we will

- Produce 2 and 3-way tables
- Produce descriptive statistics for an interval variable

**Before you start**

*All outputs below assume that you have applied the weighting variable WghtA_RS.*

*Apply this weight. If you don’t have this weight, you can find the syntax to produce it in Section 6*

8.1 Two- and three-way tables

Remember! You should always be familiar with a variable’s characteristics before attempting bi-variate and multivariate analyses (i.e. analyses using two or more variables respectively). In this section we assume that we already know the characteristics of fear of crime and sex. You may want to check Section 7 to remind you of this.

We previously looked at the impact of age and sex on the fear of crime. This time we will look at the impact of sex on the fear of crime and also whether the respondent has experienced violent crime in the reference period.

Remember from Section 7 that we can produce two-way tables by using the crosstabs command. You can either:

- Go to the menu:

  Select **Analyze > Descriptives > Crosstabs...**

  Select which variable to place in the row and column
• By using the command crosstabs <rowvar> by <colvar>

Adding /cells=col count at the end of your command means that you get the column percentages & counts in your cells.

Before progressing, let us look at the one way table of delibvio (indicating if anyone has deliberately used force/violence on respondent).

**How can you check where this variable comes from?**

**How can you obtain the following one way frequency of the variable?**

<table>
<thead>
<tr>
<th>if anyone has deliberately used force/violence on respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Valid yes</td>
</tr>
<tr>
<td>Valid no</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Missing don't know</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Let us first find out whether respondents are more or less likely to report an incidence of someone having deliberately used force/violence on respondent if they were male or female.

We wish to break down the figures by male and female (i.e. sex). So we put the variable sex in the columns of a table in which delibvio is in the rows. We would like to compare responses to delibvio between the two sexes so we can also ask for column percentages and counts to be placed in the cells of the chart.

⇒ Use the menus to go to the crosstabs dialogue box (select Analyze> Descriptives> Crosstabs...)
⇒ In the dialogue box select delibvio as the Row and sex as the Column variable
Click on Cells... to check column percentages.

Click continue, then OK to run the procedure.

Are men more or less likely to report that someone deliberately used force/violence on them?

Now produce a 2-way table that looks at whether those who have been deliberately hit or violated are more or less likely to have anxieties about walking alone after dark.

Does an experience of someone deliberately used force/violence on them have a substantial affect on fear of walking alone after dark? (hint you will need to compare the percentages in the columns across the columns. Are these very different or quite similar?)

If you have time try putting this information into a stacked bar chart using the Graphs menu.
  o Choose Graphs> Chart builder...
  o Drag a stacked bar chart to the design area
Drag delibviol to the x axis

Drag walkdark to the stack variable field

Choose Percentage (of x variable) as the statistic using the Element Properties box that appears when you open the Chart Builder. Select Total for Each X-Axis Category for the denominator.

Return to the crosstabs dialogue box and add sex in as a layer.

Is the relationship between fear of walking alone after dark and experience of violence in the previous year the same for men and women?

Could you have produced these same results using select if?

What command could you have typed to produce a 3-way table in syntax? (hint, if in doubt return to the dialog box and hit paste)
8.2 Producing descriptive statistics for an interval variable

The variable ndelibv counts how many incidences of someone deliberately using force/violence on the respondent. We will now look at summarising the characteristics of this variable. This variable has some tricky features arising from the routing of the questionnaire and the way in which the variable has been coded. These are worth exploring as these are features that may arise in other datasets.

To produce descriptive statistics of a variable we need to select from the menus: Analyze > Descriptive statistics > Descriptives

Select ndelibv from the list and click OK to get the summary statistics shown below:

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>how many times has this happened (delibv)</td>
<td>308</td>
<td>1</td>
<td>75</td>
<td>2.15</td>
<td>5.662</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>308</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What are the minimum and maximum values for this variable? – do any cases report 0 incidents of deliberate hitting or violation?

How many cases have valid values of this variable?

Thinking about the frequencies of the delibv, does 308 sound like a familiar number?

Can you think of a way that you could check who the question was relevant to by;

- Looking at the data
- Looking at the questionnaire

So, which group do the descriptive statistics relate to?

How would you interpret the mean value of 2.15 for this variable?
A good way to look at the distribution of a continuous variable is to produce a histogram. The histogram groups contiguous values together in bins to produce bars for a range of values.

The histogram has been produced using the legacy dialogue boxes. By selecting **Graphs > Legacy dialogs > Histogram**, then selecting `ndelibv` from the variable list. You can also use the Chart builder which is context sensitive and require you to drag the appropriate components to a the interactive area. You will need to have the variable type set to scale before you will be allowed to produce a histogram.

Looking at the right hand side of this distribution can you see something odd about this distribution?

Check the questionnaire documentation and search for `ndelibv` – why do we have a small peak on the right hand side?

Can you think of a way you might exclude the group with the value 97 recorded?

Re-run your descriptive analysis and produce a histogram with 97 filtered out – you may want to limit your scale to a smaller maximum this time.

How would you now report your results?
You might also consider grouping this continuous variable into a categorical variable with ordering to it (often referred to as an ‘ordinal’ variable).

How could you use the recode command to generate a variable that has the categories:
- 0 experience,
- 1 experience,
- 2+ experiences (inc. too many to recall).

One solution would be to recode sysmiss to 0, 1 to 1, 2 to 97 = 2, other = sysmiss.

What could you do to ensure that people who did not respond to the delibv variable are set to missing? (hint – use the If button on the main recode dialog)

What alternative is there to setting other values to system missing, to ensure that the values are recorded as missing but still excluded from analyses.

How could you check that the new variable has been derived correctly?

What issues may arise if you were to split those who had experienced 10 or more incidences of deliberate hitting or violation into a separate subset for analysis?

**Section 8 Key Points:**
- Build up from one way frequencies for categorical data.
- Two or three way analyses help to identify relationships between characteristics.
- Look for differences in the percentages between groups.
- Check to see if other factors have an effect on relationships you see.
- Remember to keep group counts available to your reader to help to assess reliability.
- Simple analyses of scale variables can be achieved by looking at descriptive statistics and histograms.
What could you do next?

Remember, the best you can do with sample data is to make estimates!
If you want to know how accurate your estimates are, and you have more than 30 cases drawn from a representative sample, you can use statistical theory to tell you how accurate your estimates are.

If you want a measure of how accurate an estimate such as a percentage is you can use a measure called the standard error. This measure can be used to calculate the range within which you can be 95% confident that the true value lies for the whole population. This range is called the 95% confidence interval.

Statistical theory also allows you do other good stuff (on the basis that we can estimate the sampling error) like testing whether variables are related to each other (typically with a test called the chi-squared test).
Section 9: Reporting exercise

You are supervising an undergraduate student. She needs help and has asked you whether her reporting is up to the right standard for a dissertation. She will have access to a black and white printer.

- What changes do you think could be made to improve the following report? Bear in mind that the report should be clear, accurate and understandable to a reader who is not familiar with the dataset.
- Does the report contain any mistakes?
- What suggestions for developing the analysis can you suggest?
- The syntax used to produce the analysis is available at the end, so that you can see what was done (and rerun it on the CSEW teaching dataset should you think it might help!)

Report for Secondary Analysis for Social Scientists

In this report I will do an analysis of a dataset using exploratory data analysis and SPSS. The data are taken from the UK Data Service and was downloaded from the website. The data was appropriate for the analysis because it is crime data.

Most people feel fairly or very safe while walking alone after dark.

![Pie chart showing the distribution of how safe people feel walking alone after dark, with a majority feeling very safe or fairly safe.]

People feel less safe if vandalism is more common. Similarly they feel less safe if other problems are present in the area as shown in the tables below.
Feel safe while walking alone after dark * How common - vandalism, graffiti or deliberate damage to property Crosstabulation

<table>
<thead>
<tr>
<th>% within how common is vandalism graffiti or damage to property</th>
<th>very common</th>
<th>fairly common</th>
<th>not very common</th>
<th>not at all common</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>how safe do you feel walking alone after dark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very safe</td>
<td>16.9%</td>
<td>16.5%</td>
<td>19.4%</td>
<td>29.9%</td>
<td>25.3%</td>
</tr>
<tr>
<td>fairly safe</td>
<td>30.1%</td>
<td>34.8%</td>
<td>41.0%</td>
<td>42.3%</td>
<td>41.5%</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>28.9%</td>
<td>29.7%</td>
<td>26.6%</td>
<td>19.7%</td>
<td>22.4%</td>
</tr>
<tr>
<td>very unsafe</td>
<td>24.1%</td>
<td>19.0%</td>
<td>13.0%</td>
<td>9.1%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Feel safe while walking alone after dark * How common - homes in poor conditions Crosstabulation

<table>
<thead>
<tr>
<th>% within how common are homes in poor condition run down</th>
<th>very common</th>
<th>fairly common</th>
<th>not very common</th>
<th>not at all common</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>how safe do you feel walking alone after dark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very safe</td>
<td>16.5%</td>
<td>16.6%</td>
<td>21.1%</td>
<td>28.9%</td>
<td>25.3%</td>
</tr>
<tr>
<td>fairly safe</td>
<td>34.6%</td>
<td>31.5%</td>
<td>42.5%</td>
<td>41.9%</td>
<td>41.5%</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>22.6%</td>
<td>29.7%</td>
<td>24.9%</td>
<td>19.8%</td>
<td>22.3%</td>
</tr>
<tr>
<td>very unsafe</td>
<td>27.4%</td>
<td>19.2%</td>
<td>11.5%</td>
<td>9.4%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Feel safe while walking alone after dark * How common - litter or rubbish lying around in the immediate area Crosstabulation

<table>
<thead>
<tr>
<th>% within in the immediate area how common is litter/rubbish</th>
<th>very common</th>
<th>fairly common</th>
<th>not very common</th>
<th>not at all common</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>how safe do you feel walking alone after dark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very safe</td>
<td>17.9%</td>
<td>18.4%</td>
<td>20.7%</td>
<td>31.1%</td>
<td>25.3%</td>
</tr>
<tr>
<td>fairly safe</td>
<td>32.3%</td>
<td>37.2%</td>
<td>42.9%</td>
<td>41.8%</td>
<td>41.5%</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>28.7%</td>
<td>28.7%</td>
<td>24.9%</td>
<td>19.1%</td>
<td>22.3%</td>
</tr>
<tr>
<td>very unsafe</td>
<td>21.1%</td>
<td>15.8%</td>
<td>12.5%</td>
<td>8.0%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

The difference in those who feel unsafe by feature of area is most marked for those areas with graffiti and vandalism. There is a twenty percent difference between the sense of safety of those in areas with very common and not at all common graffiti and vandalism.
Old people are typically more scared when they walk alone at night. Because of this we check whether this effect is related to age. Two tables were done; one for older people one for younger people respectively to look how many problems they had and whether this made them more scared.

**Feel safe while walking alone after dark * numprobs**

Crosstabulation

<table>
<thead>
<tr>
<th>how safe do you feel walking alone after dark</th>
<th>numprobs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>very safe</td>
<td>799</td>
<td>42</td>
</tr>
<tr>
<td>% within numprobs</td>
<td>22.1%</td>
<td>16.3%</td>
</tr>
<tr>
<td>fairly safe</td>
<td>1435</td>
<td>83</td>
</tr>
<tr>
<td>% within numprobs</td>
<td>39.7%</td>
<td>32.3%</td>
</tr>
<tr>
<td>a bit unsafe</td>
<td>851</td>
<td>72</td>
</tr>
<tr>
<td>% within numprobs</td>
<td>23.5%</td>
<td>28.6%</td>
</tr>
<tr>
<td>very unsafe</td>
<td>531</td>
<td>0</td>
</tr>
<tr>
<td>% within numprobs</td>
<td>14.7%</td>
<td>23.3%</td>
</tr>
<tr>
<td>Total</td>
<td>3615</td>
<td>257</td>
</tr>
<tr>
<td>% within numprobs</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

From the tables we can see that the number of problems affects both young and old people’s concerns about walking alone at night. Old people are more likely to be concerned about walking around at night. Both old and young people are most likely to be concerned about walking alone at night if there are lots of problems. We can tell that the differences in age are not solely down to differences in the places in which older and younger people live.
Syntax for analysis behind the report

*start by weighting the data so that we don’t need to worry about the affect of the sample design.
weight by tcindwt.

*1 do one way tables of the variables will be using.
freq rubbcomm poorhou vandcomm walkdark sex.

*lets get out the pie charts for walkdark.
GRAPH
/PIE=COUNT BY walkdark .
*decided to add 3D affect using menus at this point to show I know how to do this - marker will be impressed.

*look at these areas and see whether these problems individual affect anxiety about walking home.
crosstabs walkdark by rubcomm.
crosstabs walkdark by poorhou.
crosstabs walkdark by vandcomm.

*whoops i forgot to put the percentages on these, and this would be better for making comparisons.
crosstabs walkdark by rubbcomm /cells = col.
crosstabs walkdark by poorhou /cells = col.
crosstabs walkdark by vandcomm /cells=col.

*looks like all 3 household problems are related to sense of safety in walking alone after dark.
*does the number of problems an area have relate to fear of walking alone after dark?
*produce a measure of area problems.
*take the 3 indicators and recode them so they are eaily counted (1 = very or fairly common problem).
*not this counts the num of problems declared so no answer is treated the same as a no.
recode rubbcomm ( 1 thru 2 = 1) (else = 0) into rubarea.
crosstabs rubbcomm by rubarea.
recode poorhou ( 1 thru 2 = 1) (else = 0) into houarea.
crosstabs poorhou by houarea.
recode vandcomm (1 thru 2 = 1 ) (else = 0) into vanarea.
crosstabs vandcomm by vanarea.

*now combine these to make a count from 0 to 3 of how many problems the area has.
compute numprobs = rubarea + houarea + vanarea.
freq numprobs.

* now lets see if this is related to walkdark.
crosstabs walkdark by numprobs /cells = col count.

* lets look at the relationship between agegroup and fear of walking alone after dark.
* first check the univariate distribution of age group.
freq agegrp7.
crosstabs walkdark by agegrp7 /cells = count col.
*the distribution of fear of walking alone goes up steeply after 65, so break here?.
* look at the relationship between type of area and fear of walking alone.

* doing the selection with syntax so need to use the temporary command first.
* select those 65 or over.
temporary.
select if (agegrp ge 5).
crosstabs walkdark by numprobs /cells = col count.

* repeat for those aged less than 65.
temporary.
select if (agegrp7 lt 5).
crosstabs walkdark by numprobs /cells = col count.
You know how to produce and report good 1-way, 2 way and 3 way tables. You can produce graphs for categorical variables. You can produce histograms and summary statistics for interval data.

You already know how to check variables, assess the fitness of a dataset, derive new variables and select subsets.

You have the key skills for a straightforward piece of analysis of real data from a major survey. Identify your research question, think about what you are aiming to do – you have the skills, you can do it!
References


Appendix 1: Introducing the British Crime Survey 2007-2008

The British Crime Survey (BCS) is one of the largest social surveys conducted in Britain. It is primarily a ‘victimisation’ survey in which respondents are asked about the experiences of property crimes of the household (e.g. burglary) and personal crimes (e.g. theft from the person), which they themselves have experienced in the 12 months prior to the date of interview. The wording of the series of questions which are asked to elicit victimisation experiences have been held constant throughout the series of BCS surveys.

Because members of the public are asked directly about victimisation, the BCS provides a record of the experience of crime which is unaffected by variations in the behaviour of victims about reporting the incident to the police, and variations over time and between places in the police practices about recording crime. The scope of the BCS goes well beyond the counting of criminal incidents, although it is for this estimate that it has become established as a definitive source of information. In order to classify incidents, the BCS collects extensive information about the victims of crime, the circumstances in which incidents occur and the behaviour of offenders in committing crimes. In this way, the survey provides information to inform crime reduction measures and to gauge their effectiveness. The BCS has been successful at developing special measures to estimate the extent of domestic violence, stalking and sexual victimisation, which are probably the least-reported to the police but among the most serious of crimes in their impact on victims.

The BCS data is deposited at the UK Data Archive at the University of Essex, which holds the data for each BCS carried out since 1982. Further information about the methodology of the BCS can be found in the BCS Technical Report on the UK Data Archive website <www.data-archive.co.uk>.

Historical outlook

The BCS has been carried out in England and Wales 17 times, in 1982, 1984, 1988, 1992, 1994, 1996, 1998, and annually from 2000. Since April 2001, interviewing has been carried out continually and reported on in financial year cycles. It was conducted by a consortium of the National Centre for Social Research (formerly SCPR) and the Office for National Statistics. The 1982 and 1988 surveys were also conducted in Scotland. Users should note that the 1988 Scottish survey was also known as the Scottish Areas Crime Survey. Since 1993, separate Scottish Crime Surveys have been conducted approximately once every three years.
Aims

The main aim of the BCS is to estimate the extent of crime against individuals and their private property. It provides an alternative measure of crime to that provided by the recorded crime statistics. The BCS asks adults aged 16 years and over living in private households in England and Wales about their experiences of criminal victimization over the previous 12 months, regardless of whether or not they reported the incident to the police. To enable comparisons between surveys, the core questions on victimization have remained unchanged since the first sweep.

There are three BCS measures of the extent of crime in England and Wales:

- incident rates: the number of crimes per 10,000 adults (aged 16 plus) or households in England and Wales
- prevalence rates (also known as risks): the percentage chance of an adult or household being a victim once or more
- crime count: the total number of crimes (calculated by applying the incident rates to population figures)

All three measures are calculated for the financial year. So, the 2007-2008 BCS measures crime from April 2007 to March 2008. It should be noted that the figures derived from the BCS are estimates. As with any sample survey the BCS estimates are subject to sampling error and a range of other methodological limitations.

The survey includes demographic and lifestyle variables, both for the respondent and the head of household, which allow the identification of relative risks of victimisation by socio-economic, demographic and lifestyle factors. The BCS also collects information on the nature of crime, including where and when offences occur and the emotional, physical and financial impact of crime. On an ad-hoc basis it also covers various other crime-related issues such as concern about crime, attitudes to the police and drug misuse. The survey is also used to measure non-crime issues, such as experiences of fires, on an ad-hoc basis.

For some topics, trend analysis may prove difficult due to the fact that topics are not always covered by the same questions each time. For non-crime sections there is the concern that responses will be affected by the overall crime context of the survey. However, great care is taken to minimize contextual effects by choosing the most appropriate place in the survey to place non-crime topics.
Questionnaire development

The 2007-2008 BCS interview comprised 6 main sections. These were as follows:

- introductory questions about the household
- screener - questions used to identify victimization experiences
- victim forms for any incidents identified at the screeners (up to a limit of six forms)
- follow-up modules: experience of the police, attitudes to the criminal justice system, crime prevention and disorder, and an ad hoc module
- modules on night-time economy, anti-social behaviour, crime and disorder in town centres and on public transport and domestic violence and sexual victimization
- self-completion sections on drugs and drinking behaviour

The entire interview was administered using computer-assisted personal interviewing (CAPI). All respondents were asked the introductory questions and the screeners, although within these sections particular questions were asked of sub-groups and in some cases two variants of the same questions were each asked of half the sample. Victim Forms were automatically included in the interview by the CAPI program, to collect details of incidents identified at the screeners. A maximum of six Victim Forms could be asked. A long version of the Victim Form was asked for the first three incidents and a shorter version for the fourth to sixth incident and for any that had occurred outside of England and Wales.

The victimisation screening questions were designed to ensure that all incidents of crime that fit within the scope of the BCS, including relatively minor ones, are mentioned. The screener questions also aim to ensure that each incident is only counted once.

Sample design

The BCS sample is designed to give, after appropriate weighting, both a representative cross-section of private households in England and Wales, and of individuals aged 16 and over living in them. Since 1992 the Small Users Postcode Address File (PAF) has been used as the sampling frame. The PAF, listing all postal delivery points in the country, represents the fullest register of household addresses as almost all households have one delivery point, or letterbox.
A stratified multi-stage random probability design is used to select the sample of addresses. As with all large-scale surveys the BCS sample is clustered to keep costs at an acceptable level without compromising the quality of the sample. Since 1992 the procedure has been as follows. Postcode sectors are sorted into 10 standard regions. Within each region, sectors are listed in order of population density and divided into three roughly equal-sized bands, in terms of the number of delivery points. Within each of the 30 strata, sectors are ordered in increasing order of the percentage of households with head of household’s socio-economic group (SEG) defined as professional and managerial, other non-manual or skilled manual. By sampling systematically down the ordered list, using a random starting point, postcode sectors are selected with a probability proportional to size (number of PAF addresses). Within each of the postcode sectors selected, the list of delivery points is divided into four equal-sized segments. One of these segments is chosen at random.

Where one address has more than one household, a single household is selected using random selection procedures. One adult aged 16 or over in each selected household is identified for interview using similar random-selection procedures. No substitution of respondents is allowed.

Inner city areas are oversampled by a factor of about two. Inner city areas are selected on the basis of classifying postcode sectors according to population density, level of owner-occupied tenure, and social class profile.

**Sampling errors and design effects**

If the BCS sample was a simple random sample of dwelling units in Britain, the estimates produced at this stage would represent victimization estimates covering England and Wales for the time period of the survey. However, the sample is clustered within postcode sectors and different individuals are selected with different levels of probability. While weighting removes these differences in selection, the weighted results are not based on a simple random sample. An estimation procedure is used to calculate the extent to which the estimates need to be set within wider confidence intervals, due to the complex nature of the sample design.

**Weights**

The BCS contains a number of weights that should be applied for different types of analyses. However, for the purpose of the teaching dataset only two weights are available: tcindwt for individual-level analyses and tchhdwt for household-level analyses. This is explained in more detail in the following section.
Appendix 2: The BCS 2007-2008 Teaching Dataset

The data and variables

The data file contains 11,676 adults aged 16 years and over in England and Wales. This represents roughly a quarter of the core sample in 2007-2008, only those individuals who were randomly assigned to answer the Module B questions. The teaching dataset includes 149 variables.

Weights within the dataset

Different units of analysis (households and individuals) in the BCS Dataset have different probabilities of inclusion in the sample. These differences arise from a number of sources: the over-sampling of small Police Force Areas (PFAs), the sub-selection of one dwelling unit at an address, the selection of one individual within a dwelling and differential response rates within subgroups. It is necessary to correct for these differences, by weighting, in order that estimates will be unbiased. Otherwise, the sample would over-represent small PFAs, single-dwelling addresses and people living alone.

The sample is designed to be representative of the entire household population of England and Wales, so use of tcindwt (for analysis of individual-level data) or tchhdwt (for analysis of household-level data) will provide total population estimates. These weight variables are the individual and household weights provided in the full 2007-2008 BCS (indivwgt and hhdwgt) each divided by their mean value.

Most of the variables included within the teaching dataset are individual variables, and require individual-based analysis. However, there are a few household variables such as mottheft, motstole or yrdeface. When analysing these variables the household weight variable (tchhdwt) should be used. The table on pages 10 to 16 indicates which weight should be applied to each variable. In interpreting and presenting the analysis the weighted percentages, means etc should be used. Unweighted information is only used for the sample size.

Missing values within the dataset

Respondents on the BCS are not usually explicitly given the options of “refusal” or “don’t know” when asked a question. However, respondents may say they do not know or they may refuse to answer a question or they may feel that the question does not apply to them. The code for “not applicable’ is ‘7’ and for “refusal” is ‘8’ for code frames up to 6, ‘97’ and ‘98’ for code frames up to 96, and so on. The code for “don’t know” is ‘9’, ‘99’ and so on.
It is often useful to run frequencies on the variables as a first stage in any analysis to examine the
distribution of responses and the proportion of ‘don’t know’ and ‘refusal’ responses. In most analysis
‘refusal’ codes are excluded. ‘Don’t know’ codes are also usually excluded unless there is interest in
‘don’t know’ responses, for example in the case of attitudinal questions. ‘Refusal/don’t know’
responses have been dealt with slightly differently within the two versions of the teaching datasets.

**Missing values in the SPSS teaching dataset:** The SPSS teaching dataset has all ‘refusal’ and ‘don’t
know’ responses set as missing values. Some variables within the dataset also contain ‘system missing’
cases - SPSS automatically codes an empty cell as ‘system missing’ which is denoted by a dot (.).

**Missing values in the STATA teaching dataset:** The STATA teaching dataset includes all refusal and don’t
know responses as missing responses. The equivalent of ‘system missing’ data are denoted by a dot (.),
while .a denotes ‘not applicable’, .b denotes ‘refused’ and .c denotes ‘don’t know’ responses.

---

**Listed links**