Design Principals for the Basic Fact Sheets and Annual Statistical Reports

Jeremy Broughton, Jackie Knowles and Louise Walter, TRL

1 Introduction

One of the achievements of the SafetyNet Integrated Project was to develop a series of Basic Fact Sheets (BFS) and an Annual Statistical Report (ASR). These were designed to make CARE data available to a general audience with an interest in road safety in Europe. They were widely accessed, indeed proved to be one of the most successful outputs of SafetyNet. Consequently, when the current DaCoTA project was set up, one of the tasks covered the production of new and updated versions of the BFS and ASR. The ASR produced during the first year of the DaCoTA project included 56 tables of data, and 29 figures. Seventeen topics were covered in the BFS.

Basic Fact Sheet	
Update and expansion of existing BFS	Main figures
	Children (aged<15)
	Young people (aged 18-24)
	The Elderly (aged>64)
	Pedestrians
	Cyclists
	Motorcycles and mopeds
	Car occupants
	Heavy Goods Vehicles and Buses
	Motorways
	Junctions
	Urban areas
New BFS	Youngsters (aged 15-17)
	Roads outside urban areas
	Seasonality
	Single vehicle accidents
	Gender

Six partners took responsibility for preparing separate BFS and individually worked on these documents, as well as contributing to the ASR, so the final drafts showed a certain

variety of styles. It was subsequently agreed that a more uniform style was desirable for future versions.

This note sets out guidelines that are to be followed when a BFS is drafted or redrafted in future. Some of these issues also arise with the ASR, and section 5 brings together the guidelines that apply to the ASR. The guidelines are based as far as possible on scientific principles. In order to achieve consistency, however, it has been necessary to make rather arbitrary choices in some cases.

The annex provides guidance about the colours to be used, so that the ASR and BFS have a consistent appearance.

1.1 Purpose of the BFS and ASR, and the target audience

When designing a document, attention needs to be paid to its purpose and intended audience. The Annual Statistical Report and the Basic Fact Sheets have complementary purposes, but the same intended audience.

The intended audience in each case is those people who take an interest in road safety in Europe, both professionally and as active citizens (the broader public). The CARE database is a primary source of information about the casualties and accidents that occur on European roads, yet access to the database is tightly controlled and hitherto few publications have gone beyond "the headline figures" when presenting CARE data. The ASR and BFS are designed to fill this gap. At present, only data on fatalities and fatal accidents are included in these publications, for reasons explained in section 3.1.

- The **ASR** consists mainly of a series of tables of counts of CARE fatality data, illustrated by a series of figures derived from these tables. The purpose of this is to provide a comprehensive reference book of fatality counts.

- The **BFS** are more descriptive and provide an interpretation to a non-technical reader of the CARE data; they present an overview highlighting the main facts for a specific topic. Wherever possible, measures of risk are calculated by relating the number of fatalities from CARE to exposure data available from other sources.

As the BFS are more complex than the ASR, a wider range of issues arise – involving design as well as presentation. The name "Basic Fact Sheet" rules out complex statistical analyses, which would not be understood by many readers. Nevertheless, a BFS author should be aware of sound statistical principles when presenting results, such as the likely effects of chance when dealing with small numbers. Section 4 discusses statistical principles in more detail. The issues relating to the ASR are largely presentational and section 5 summarises the guidelines set out for the BFS that also apply to the ASR.

1.2 What does the typical BFS reader expect?

The typical BFS reader wants to access information about road safety quickly and succinctly, so text, tables and figures should be as simple as possible. Tables and figures should not attempt to convey too many messages, and the text should highlight the main ones. Percentages and rates should be used rather than counts as this allows countries to be compared easily. Large tables of numerical data and complicated graphs should not be used because they are difficult to interpret. The ASR presents the basic count data, so the BFS does not need to present these data as well.

2 The BFS template

A template has been developed for the BFS. The document has been divided into three columns with the central column containing the main text, tables and graphics. The left hand column contains at least one note per page highlighting the key points of the results that are presented in the central column. The right hand column consists of tabs with the titles of all BFS, with the title of the current BFS emphasised. The final part of the BFS consists of a common text; this includes a list of country abbreviations which should be adjusted for each BFS. Finally there is a list of authors; the author(s) of the current BFS appear in the first row.

The continental system of numbering should be used, where 1.000 represents one thousand and 1,00 represents one to two decimal places. This can be achieved simply in Word files but not in Excel charts, as Figure 8 below illustrates.

To achieve a common style, the following colours and fonts should be used:

Main heading		Ariel 26pt Dark Blue underlined
Sub heading (BFS title)		Ariel 20pt Dark Blue
Section heading		Ariel 12pt Dark Blue **Bold**
Text (side notes and main Text)		Black
Captions (Tables and Figures)		Ariel Narrow 10pt Dark Blue **Bold**
Tables	Labels	Ariel Narrow 10pt Black **Bold**
	Values	Ariel Narrow 10pt Black
	Colours	Refer to annex for the standard colours
All labels in Figures		Ariel 10pt Black

Column breaks and page breaks are used to format the pages. They should be used on every page; the alternative of inserting blank lines should be avoided.

2.1 Table design

Tables are normally prepared in Excel worksheets and copied into the Word template. In order to preserve the formatting of the Word document, it is important that each table is entered/copied directly into the Word file as text and not pasted as a picture (as the scaling affects and distorts the size of font, and Excel warnings may be left in).

Particular design details are listed below:

- The font size is Ariel Narrow 10pt black.
- If it is impossible to fit a table into the central column using Ariel Narrow 10pt font then the answer is to simplify the table – not to reduce the font size.
- Table headings appear above the table using Ariel Narrow 10pt Dark Blue **Bold**.
- The first row and the first column contain the column and row headings in bold; 15% shading should be used.

- All cells should have borders, normally single lines (solid, not broken); significant divisions of the table are marked with double lines (e.g. the EU totals should be separated with double lines at the end of the table). A small cell margin, e.g. 0,05cm, should be used so that numbers are not attached to the margin.
- Right justify numbers in columns.
- A hyphen shows that a value is unknown and a '0' denotes a zero count.
- Notes should follow each table, acknowledging the data source and date when the data were extracted. These should be presented in a uniform way, any remarks regarding the data on the left (e.g. exceptions to year of data) and the data source(s) and the date of the query on the right.
- Where tables present data by country, the following codes should be used - in this order (although some countries may appear below the EU-xx total as described in section 3.3)

Order	State Name	Abbreviation
1	Belgium	BE
2	Bulgaria	BG
3	Czech Republic	CZ
4	Denmark	DK
5	Germany	DE
6	Estonia	EE
7	Ireland	IE
8	Greece	EL
9	Spain	ES
10	France	FR
11	Italy	IT
12	Cyprus	CY
13	Latvia	LV
14	Lithuania	LT
15	Luxembourg	LU
16	Hungary	HU
17	Malta	MT
18	Netherlands	NL
19	Austria	AT
20	Poland	PL
21	Portugal	PT
22	Romania	RO
23	Slovenia	SI
24	Slovakia	SK
25	Finland	FI
26	Sweden	SE
27	United Kingdom (GB+NI)	UK

- If there are no data available in CARE for a particular country (e.g. there were no CARE data for Bulgaria or Lithuania for any year for the 2010 BFS and only 2004

data for Cyprus) then these countries should be omitted from the tables and graphs.

- Each table should have a line showing the EU total; as data will probably be missing for at least one country, this should be labelled "EU-xx". See section 3.3 for advice on calculating and presenting this total. Note: estimated values should not appear in the table, which should only show actual CARE data.
- The final page of the BFS contains a list of the abbreviations and a "definition of EU-level" that specifies "EU-xx". The reader should be referred to this list on the final page by a footnote to page 1.
- If data for particular countries are not from the year identified in the table heading, the countries should be marked and the exceptions listed in the table notes. The latest available data should be used for each country.
- If a significant proportion of data is unknown for a particular country then this country should also be omitted from the relevant tables and figures. For example, Ireland had fatal accident data for each year from 1999 to 2008, but in 82% of cases it is unknown whether or not the accident occurred at a junction and so the data for this topic are unreliable and were omitted from the junction BFS.
- If there are indications that the CARE data for a particular country may not be reliable, that country should be excluded from the BFS.

2.2 Figure design

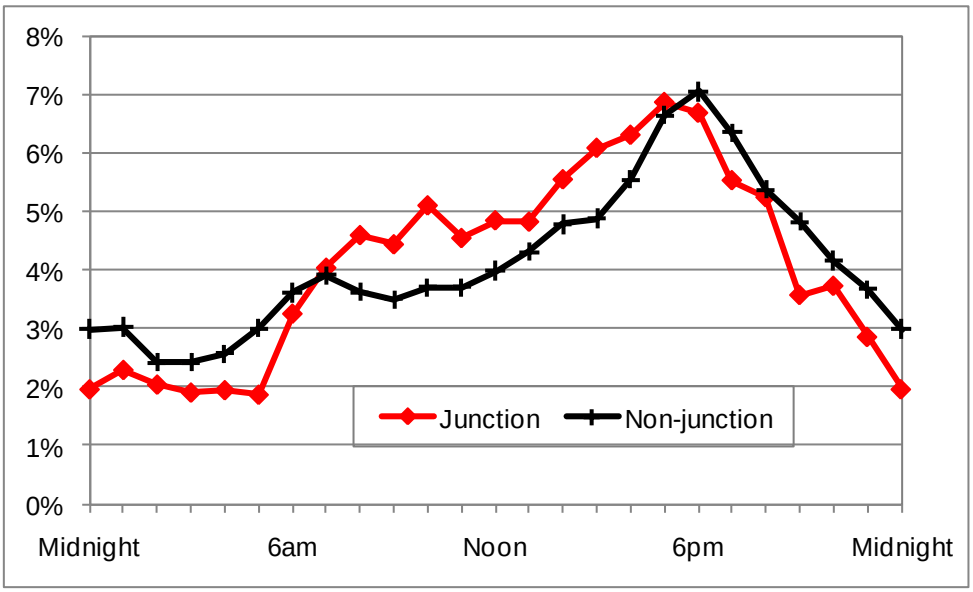
Figures are normally prepared in Excel worksheets and copied into the Word template. These need to be copied and inserted as graphical images, taking care that the images are readable and are not distorted.

Particular design details are listed below:

- Each Figure should have a border to the chart area and to the plot area.
- The plot area should be white, not shaded.
- The bars and lines should be coloured according to the standard BFS colour palette described in the annex. (Note: the figures in this document have not been prepared using the palette as they are examples from old BFSs).
- When preparing a bar chart, countries should be sorted in order according to the primary variable being displayed. A bar for the EU-xx should be included among the national bars so that a country's values can be compared directly with the EU value.
- When two Figures are to be compared directly, e.g. one for fatalities inside urban areas and another for fatalities outside urban areas, the same country order should be used in both Figures to facilitate comparisons.
- Data labels **should not** be used in the Figure to communicate the precise values (especially not to 1 or 2 significant figures). If this level of detail is needed then a table should be used (section 4.4.1).
- Do **not** use 3D charts (see section 4.3.1).
- Start the axis at zero as otherwise differences will be exaggerated (section 4.3.3).
- For line charts, position the data points above the x-axis tick marks as illustrated in Figure 1 (select format axis option 'position axis on tick marks')
- For line charts display the axis labels horizontally (select format axis alignment text direction 'Horizontal')

- When displaying data by time of day or day of week, the first values should be repeated after the final values so that their relationship can be seen, as illustrated in Figure 1. This should not be done when displaying data by month unless the January values for the following year are available, since there may well be trends in the time series. Thus, the January values are not repeated in Figure 2.
- Each Figure should acknowledge the source of the data and specify the month when the data were extracted.

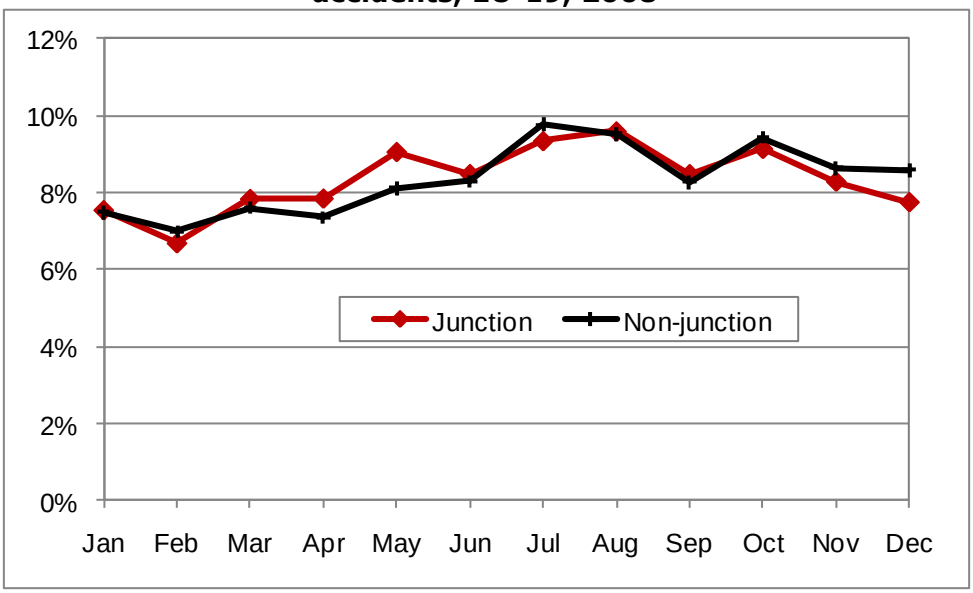
Figure 1: Distribution of fatalities by hour, EU-19, 2008



Source: CARE Database / EC

Date of query: October 2010

Figure 2: Distribution of fatalities by month in junction and non-junction accidents, EU-19, 2008



Source: CARE Database / EC

Date of query: October 2010

3 BFS author guidelines

3.1 Present fatality data only

The CARE database holds information about road accidents throughout the Member States that involved personal injury. Although the reporting of fatal accidents is generally good throughout Europe, the same cannot be said of non-fatal accidents. The severity definitions applied by national reporting systems differ, as do the proportions of accidents that are reported. Consequently, at present the ASR and BFS only present data for fatal accidents and those who died in them.

The preparation of the CARE database involves various technicalities such as the use of correction factors for countries that do not apply the 30 day definition. These are presented in the CARE documentation that is available through the EC website. The typical BFS reader is unlikely to be interested in these issues, so it is unnecessary to refer to them in the BFS.

3.2 Do not duplicate large tables of count data that can be found in the ASR

The BFS have been developed to provide a simple presentation of the CARE data on specific topics. The ASR presents the basic count data in large tables, so the BFS does not need to present these data as well. If the table you wish to include does not appear in the current ASR then suggest it as a new ASR table. Note that if it is impossible to fit a table into the central column of the BFS template using Ariel Narrow 10pt font then the table is too large and it needs to be simplified e.g. by combining categories.

3.3 Calculating the EU summary total row

There are missing values in the CARE database. Data are entirely missing for some countries for some years, while certain variables may be missing even when the data are present. The effects of this are shown in Table 1, where a hyphen shows that a value is unknown: for example, data are missing for Germany (DE) in 1999 and for Estonia (EE) in 1999-2004.

In general each table should contain a summary EU row, to indicate the European trend over the decade. This should contain as many countries as possible, to be as representative as possible of the actual EU total, but this is complicated by the absence of certain data. It would be possible to include only those countries with complete time series, fourteen in the case of Table 1. It would be possible to estimate the missing Latvian data but with only three known values the result would be questionable. A compromise is proposed where only one or two values are missing for a country: estimate each unknown value as the nearest known value. Note that these estimated values do not appear in the table, which should only show actual CARE data; the accompanying text should explain the method of estimation.

Five EU Member States with CARE data for some years were excluded from Table 1 for the reasons summarised in section 2.1. The time series are complete or nearly complete for eighteen countries, and these appear above the EU-18 row to identify which countries contribute to this total. Data for the other four EU States appear below the EU-18 row, as they have not contributed to the EU total.

There are data in CARE for Switzerland, but only for 2004. Switzerland is a member of the European Economic Area but not of the European Union, so it would be difficult to explain why Swiss data for 2004 should be included. When CARE data become available for an EEA state for several years, however, it could well be appropriate to include that state; this would be a collective decision covering the ASR and all BFS.

Table 1: Number of fatalities in junctions accident per country, 1999-2008

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
BE	302	334	357	315	272	221	210	207	195	167
CZ	297	283	241	289	303	327	267	222	218	238
DK	155	150	122	130	128	122	94	101	129	126
DE	-	1.739	1.643	1.577	1.578	1.359	1.293	1.249	1.153	1.073
EL	162	141	148	168	139	122	118	159	146	147
ES	930	914	856	805	806	764	750	754	721	577
FR	1.444	1.375	1.364	1.238	971	822	664	593	565	475
IT	1.354	1.528	2.013	2.000	1.837	1.761	1.674	1.654	1.550	1.372
LU	2	11	8	8	11	8	3	3	7	8
NL	404	401	327	321	324	247	249	276	253	227
AT	189	153	146	167	161	145	148	128	123	115
PL	-	-	934	934	983	1.014	898	768	840	834
PT	251	225	236	196	187	213	196	131	161	140
RO	53	59	71	94	64	61	236	238	272	269
SI	-	21	28	28	17	19	28	23	24	-
FI	91	85	104	93	83	65	73	65	62	72
SE	171	155	155	171	115	125	98	99	115	97
UK	1.340	1.318	1.325	1.287	1.289	1.189	1.152	1.115	1.089	907
EU-18	9.839	9.826	10.077	9.821	9.269	8.584	8.151	7.785	7.623	6.868
Yearly reduction		0%	-3%	3%	6%	7%	5%	4%	2%	10%
EE	-	-	-	-	-	-	33	38	54	38
LV	-	-	-	-	-	-	-	45	53	20
HU	-	-	-	-	316	280	260	266	268	246
SK	-	-	-	-	-	-	72	75	61	70

Source: CARE Database / EC

Date of query: October 2010

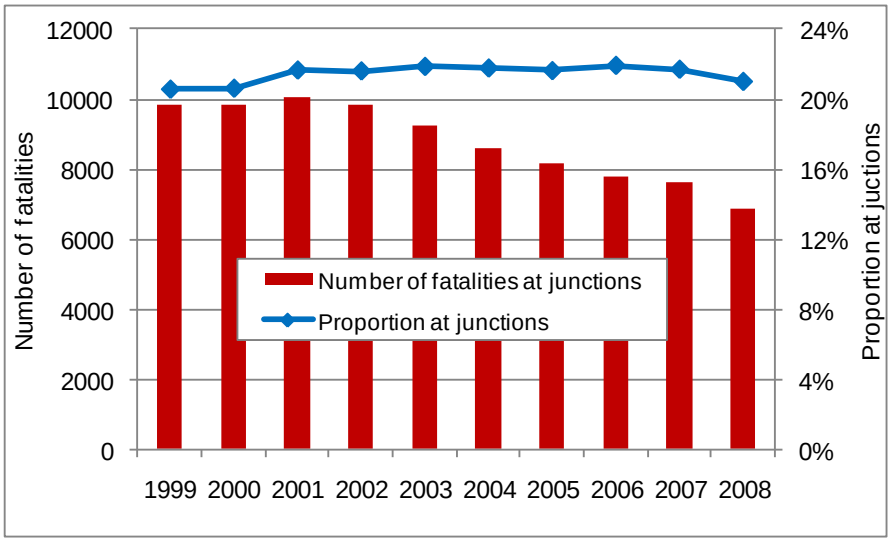
3.4 Content guidance

The BFS is not a technical report, rather an accessible presentation of numerical data. A mixture of simple graphics and tables should be used, especially as many readers have difficulty interpreting large tables of numerical data. The use of percentages and rates rather than the counts is encouraged as this allows easy comparisons to be made between the countries.

Figures should be used to illustrate aspects of the data, not to present detailed data. A Figure can illustrate data that are presented in a Table, but data labels should not be used in the Figure to communicate the precise values (especially not to 1 or 2 significant figures). If this level of detail is needed then a table should be used (refer to section 4.4.1).

Each BFS should begin with a table of national fatality counts for the past 10 years to give a feel for the basic data; e.g. Table 1 from the Junction BFS. There should also be a Figure showing the EU totals over the decade, preferably relating this to the EU fatality total. Figure 3 from the Junction BFS illustrates this point.

Figure 3: Number and proportion of fatalities in EU-18 in road accidents at junctions



A BFS generally studies a subset of fatalities, so it is important to examine the relationship between the subset and the full set. This is achieved in Figure 3 by including the proportion of fatalities occurring at junctions with the annual numbers. This shows that, while the number of fatalities at junctions has fallen in recent years, the proportion of accidents at junctions has remained fairly constant. Thus the decline of the number of fatalities at junctions has been in line with the general decline of fatalities rather than a change from junction accidents to non-junction accidents. From a scientific standpoint, it would be preferable to make the comparison with the disjoint set, i.e. fatalities not at junctions, but this would probably be more difficult for the general reader to understand.

A BFS aims to provide an overview of a specific aspect of road safety, so when dealing with developments over 10 years the text should focus on broad trends rather than the changes from one year to the next, which are often due to random variation.

Following on from this, the fatality numbers should be related to an exposure measure to calculate fatality rates whenever possible. Examples include fatality rates per million population or per billion vehicle km travelled. This allows better comparisons to be made between countries.

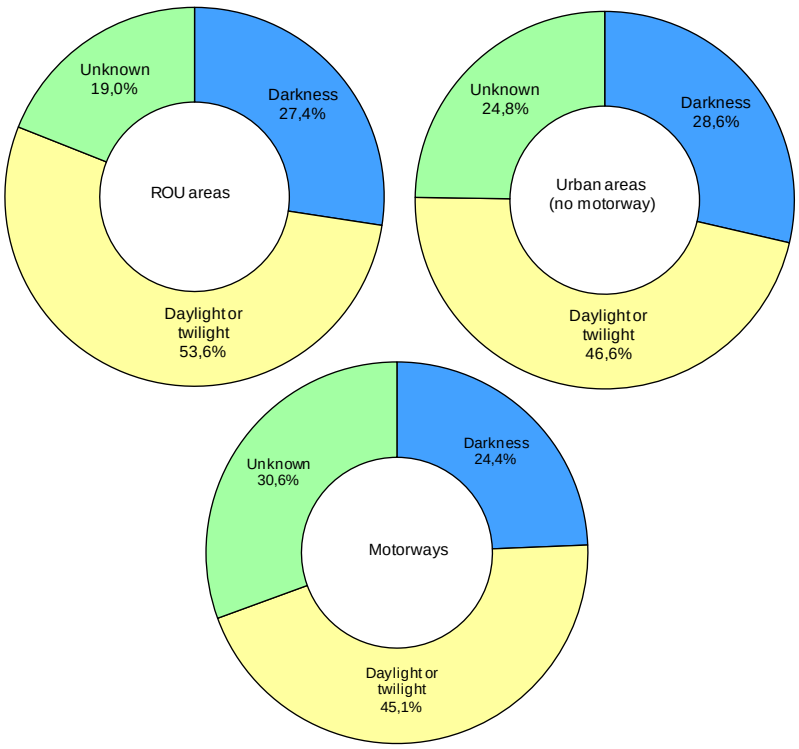
The BFS may then consider who was involved by looking at age and gender distributions. Details of where the fatalities occurred should be included such as road type, junction/no junction, urban/rural, etc. Seasonal patterns may also be of interest such as distributions by month, day of week, time of day. If the BFS is looking at a particular vehicle type it may be interesting to compare it to other vehicle types; similarly if the BFS is looking at a particular age group then comparisons may be made with other age groups. NOTE: only include tables and figures on topics that show interesting facts. When considering

variables such as manoeuvre it may be advisable to group manoeuvres together to ensure each cell has at least 5 observations.

The facts that appear in the first column of the template should be self-explanatory. They should also highlight the most important aspect of the page. It is likely that these basic facts do not vary from year to year, so changes that may be due to small numbers should not be highlighted.

Beware of displaying figures and making comparisons where there are large and unequal proportions of unknowns. For example any differences seen in Figure 4 may be due to the unknowns; it would be more meaningful to exclude the unknowns before calculating the proportions – with an explanatory note. Alternatively, those countries with a significant incidence of unknown data could be excluded and the proportion calculated for a smaller EU-xx. It may be better to display the information of Figure 4 in a table (N.B. ROU stands for Roads Outside Urban).

Figure 4: Fatalities on ROU areas, urban areas and motorways by lighting conditions in EU-22, 2008



4 Statistical principles when presenting data

Pictures are often appealing. If designed properly, they are easy and quick to understand, but if designed incorrectly they can be misleading and confusing. In general the aim of the tables and graphs in the BFS is to present simple comparisons between the EU countries or long term trends or seasonal patterns for the EU-xx total.

Tables and graphs should be simple and self-explanatory – the reader should be able to understand them without making detailed reference to the text. On the other hand, if there is no need to describe and discuss the graphs or table in the text then there is generally no need for the table or graph at all.

4.1 Choosing graphs or tables

Data can be displayed in many ways to suit the audience, data and the message. Graphs and tables should be chosen carefully to emphasize and reveal important facts, comparisons and relationships in the best possible way.

In general tables are better than graphs for giving structured numeric information, whereas graphs are better for indicating trends and making broad comparisons or showing relationships. If it is necessary to include the precise values as data labels then a table would be more appropriate.

4.2 Tables

Firstly consider if you need to show actual data values, or whether a summary of the data would be more informative. As discussed earlier, the first table in the BFS should present the counts but it is often more appropriate in the following tables to use rates or percentages as this allows countries to be compared easily. When presenting percentages, always show the sample sizes on which the percentages are based.

Rounding percentages and rates effectively makes it much easier to see the pattern. In most cases the percentages should be presented with no decimal places, as the use of too many significant digits suggests a misleading level of confidence in the estimates. Table 2 is a good example of presenting the data as percentages rather than counts as this allows easy comparisons to be made between the countries. The total is given in the final column to provide the sample size. As a general rule, percentages should not be calculated for cells of less than 5 as such percentages could well differ appreciably from year to year purely as a result of chance (see section 4.2.1).

This creates a difficulty for small countries such as LU and MT where the fatality figures will always be small. As these countries are part of the EU, the figures should be provided – but without comment in the text as conclusions cannot be drawn because of the likely effects of chance.

In general for the BFS, the row percentages add up to 100% as in Table 2. However, it may be appropriate to present the tables where the whole or parts of the table sum to 100%. For example Table 3 begins with the numbers of fatalities in junction and non-junction accidents (giving the sample sizes) and then shows the gender distributions for junction and non-junction accidents where each subset total equals 100%. Using this presentation it is easier to see that 34% of fatalities in non-junction accidents were males aged 25-49 compared with 27% in junction accidents. This would not be clear if we looked at just the count data in the top half of the table. Finally the table presents

the proportion of each group of fatalities that was killed at a junction. For example 16% of female child fatalities occurred at a junction (i.e. 60/365).

Table 2: Distribution of junction fatalities per country by mode of transport, 2008

	Car or Taxi	Pedestrian	Motor Cycle	Pedal Cycle	Moped	Lorry	Other	Total
BE	38%	13%	20%	19%	5%	4%	1%	167
CZ	44%	21%	18%	12%	1%	3%	1%	238
DK	29%	13%	17%	26%	10%	4%	1%	126
EE	58%	24%	3%	5%	5%	3%	3%	38
EL	36%	24%	36%	1%	2%	1%	0%	147
ES	32%	19%	25%	3%	12%	6%	3%	577
FR	33%	19%	24%	9%	12%	2%	2%	475
IT	38%	10%	29%	10%	9%	1%	3%	1.372
LV	60%	40%	0%	0%	0%	0%	0%	20
LU	63%	38%	0%	0%	0%	0%	0%	8
HU	38%	27%	9%	17%	5%	3%	1%	246
NL	28%	8%	11%	37%	11%	1%	4%	227
PL	37%	39%	5%	12%	3%	2%	1%	834
PT	25%	15%	24%	8%	13%	12%	2%	140
RO	35%	37%	3%	10%	7%	4%	4%	269
SI*	21%	8%	42%	21%	8%	0%	0%	24
SK	34%	39%	3%	20%	0%	4%	0%	70
FI	54%	18%	10%	8%	4%	3%	3%	72
UK	36%	30%	24%	5%	1%	2%	2%	907
EU-19	36%	22%	20%	11%	7%	3%	2%	5.957

* data for 2007

Source: CARE Database / EC

Date of query: October 2010

Table 3: Distribution of junction fatalities by age and gender, EU-19, 2008

		<15	15-17	18-24	25-49	50-64	65+	not known	Total
Number of fatalities in:									
junction accidents	female	60	64	160	372	296	614	29	1.594
	male	113	192	639	1.635	729	990	56	4.355
non-junction accidents	female	305	235	845	1.657	934	1.666	56	5.697
	male	423	532	3.638	8.689	3.418	2.823	229	19.751
Distribution of fatalities in:									
junction accidents	female	1%	1%	3%	6%	5%	10%	0%	27%
	male	2%	3%	11%	27%	12%	17%	1%	73%
non-junction accidents	female	1%	1%	3%	7%	4%	7%	0%	22%
	male	2%	2%	14%	34%	13%	11%	1%	78%
Proportion of fatalities	female	16%	21%	16%	18%	24%	27%	35%	22%
occurring at junctions	male	21%	27%	15%	16%	18%	26%	20%	18%

Source: CARE Database / EC

Date of query: October 2010

If the data include some unknown values then it is also important to mention these, and show whether these values are included in the percentages. When producing a BFS, it is desirable to preserve the overall fatality totals in the table and this could be done in two ways. The first way would be to calculate the percentages including an 'unknown' column/percentage and the second way would be to apportion the unknown values between the known values by weighting the known values accordingly to ensure the totals equal 100%. The weighting assumes that the missing data are unbiased which is generally a reasonable assumption. Notes should appear in all figures and tables that present data that have been adjusted in this way.

Tables in the BFS should not be too complicated. For example Table 4 contains so much information that it is hard to read the numbers and is unclear what pattern or feature the author is trying to convey to the reader. An alternative and simpler presentation would be to show the information for a single year (for example, 2008) with gender shown in two separate columns. If the time series was particularly interesting the trend could be shown in a graph for the EU-xx total for males and females separately. Note this example also includes imputed numbers used to calculate the EU-xx total which is not recommended (refer to section 3.3).

Table 4: Fatalities per million inhabitants by country, 1999-2008

		1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
BE	female	72	70	73	65	57	47	49	46	40	41
	male	204	220	220	191	180	178	161	159	138	137
CZ	female	59	62	57	60	55	54	50	37	46	44
	male	204	206	185	202	209	190	176	152	152	151
DK	female	53	54	40	39	45	36	29	32	38	39
	male	142	126	111	117	116	101	94	82	111	110
DE	female	49	49	46	45	42	37	34	33	31	29
	male	136	136	126	123	120	106	97	92	90	81
EE	female	57	57	171	57	57	173	58	69	62	47
	male	198	199	0	201	202	0	203	245	157	157
IE	female	63	55	53	51	40	50	49	46	39	34
	male	155	161	160	134	125	129	140	125	92	91
GR	female	83	80	75	63	52	65	63	52	60	54
	male	307	294	269	235	241	238	236	247	225	224
ES	female	57	56	55	52	50	45	37	34	31	24
	male	191	197	185	178	177	153	145	131	96	95
FR	female	75	63	66	59	45	42	41	36	34	31
	male	211	209	206	194	154	139	132	116	106	105
IT	female	51	56	60	54	48	42	41	42	33	31
	male	188	197	194	196	186	174	161	154	131	130
LV	female	75	75	76	77	77	78	78	78	81	63
	male	280	282	284	286	288	289	291	292	227	228
LU	female	74	77	76	53	44	69	73	42	29	33
	male	200	276	240	228	195	151	131	142	115	113
HU	female	59	59	59	60	60	57	58	55	59	45
	male	206	207	208	208	209	207	201	211	158	159
NL	female	35	37	30	30	32	27	25	26	23	22
	male	103	100	94	92	95	71	67	64	62	61
AT	female	78	60	61	66	60	51	46	45	38	40
	male	197	189	182	174	173	168	144	135	126	126
PL	female	67	67	67	69	67	66	63	63	68	67
	male	224	224	227	240	232	238	226	216	222	223
PT	female	74	66	68	65	56	49	43	33	34	38
	male	327	306	264	266	247	203	197	152	131	131
RO	female	52	53	50	53	45	54	62	60	67	69
	male	170	169	171	170	162	174	184	183	219	219
SI	female	72	72	50	66	53	53	70	41	58	35
	male	250	248	233	207	193	225	191	224	180	180
SK	female	48	48	48	48	48	48	48	46	54	36
	male	166	166	167	167	167	167	167	175	175	175
FI	female	50	50	52	43	40	44	36	35	38	29
	male	119	104	117	119	107	101	110	94	103	102
SE	female	35	34	33	30	31	26	25	25	28	24
	male	96	100	99	96	88	82	73	74	63	63
UK	female	32	30	29	30	30	27	27	26	24	22
	male	91	93	95	92	94	87	86	84	77	65
EU 23	female	55	53	53	51	46	43	41	39	38	35
	male	173	174	168	165	157	146	139	132	130	118
EU 16	female	55	53	53	50	44	41	39	37	35	32
	male	175	176	170	165	155	142	136	127	124	112
CH	female	0	0	0	0	0	31	0	0	0	0
	male	0	0	0	0	0	110	0	0	0	0

Tables showing time trends should only be used if the data vary significantly over time, rather than showing random fluctuations due to small numbers and volatile data. For most countries in Table 5, for example, the proportions are fairly stable, so it may be sufficient to present the figures for 2008 only – or possibly in a graph with 1999 figures as shown in Figure 5. This table also shows the problems of calculating percentages of very small numbers for LU which has large variations over time.

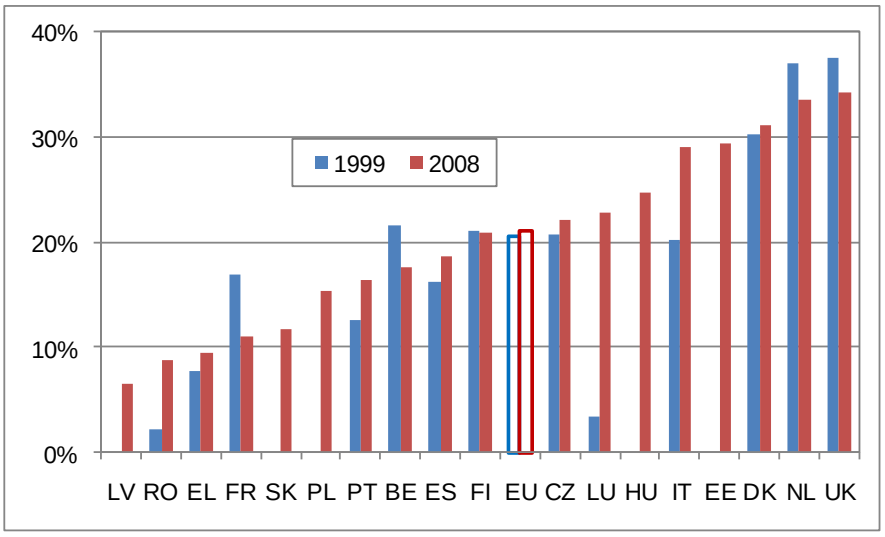
Table 5: Proportion of fatalities in junction accidents per country, 1999-2008

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
BE	22%	23%	24%	24%	22%	19%	19%	19%	18%	18%
CZ	21%	19%	18%	20%	21%	24%	21%	21%	18%	22%
DK	30%	30%	29%	28%	30%	33%	28%	33%	32%	31%
EE	-	-	-	-	-	-	22%	21%	31%	29%
EL	8%	7%	8%	10%	9%	7%	7%	10%	9%	9%
ES	16%	16%	16%	15%	15%	16%	17%	18%	19%	19%
FR	17%	17%	17%	16%	16%	15%	12%	13%	12%	11%
IT	20%	22%	28%	29%	28%	29%	29%	29%	30%	29%
LV	-	-	-	-	-	-	-	11%	13%	7%
LU	3%	14%	11%	13%	21%	16%	6%	7%	15%	23%
HU	-	-	-	-	24%	22%	20%	20%	22%	25%
NL	37%	37%	33%	33%	32%	31%	33%	38%	36%	34%
PL	-	-	17%	16%	17%	18%	16%	15%	15%	15%
PT	13%	12%	14%	12%	12%	20%	20%	17%	20%	16%
RO	2%	2%	3%	4%	3%	2%	9%	9%	10%	9%
SI	-	7%	10%	11%	7%	7%	11%	9%	8%	0%
SK	-	-	-	-	-	-	-	12%	9%	12%
FI	21%	21%	24%	22%	22%	18%	20%	20%	16%	21%
UK	38%	37%	37%	36%	35%	35%	35%	34%	36%	34%
EU-18	21%	21%	22%	22%	22%	22%	22%	22%	22%	21%

Source: CARE Database / EC

Date of query: October 2010

Figure 5: An alternative presentation of Table 5



4.2.1 Variability of annual percentages

The percentages in Table 5 for smaller countries are very variable, particularly for LU. These variations are probably random and are unlikely to represent any real change in safety at junctions, so including these intrinsically variable percentages in a BFS table is potentially misleading. Table 6 shows the 95% confidence intervals for various combinations of sample size (in this case national total of fatalities) and proportion. For

example, consider a country with 50 fatalities per year where on average 5 (10%) occur at junctions; to judge that a significant change had occurred in a particular year, the number that occurred at junctions would have be less than 2 (4%) or more than 10 (20%).

Table 6: Confidence Intervals for Percentages

Sample Size	Percentage	95% Confidence Interval
10	50%	(23%, 77%)
20	50%	(30%, 70%)
30	50%	(33%, 67%)
40	50%	(35%, 65%)
50	50%	(37%, 63%)
10	10%	(2%, 41%)
20	10%	(3%, 30%)
30	10%	(4%, 26%)
40	10%	(4%, 23%)
50	10%	(4%, 20%)

This suggests that percentages should only be included if the fatality total is at least 50 and the individual count is at least 5 (although the final row suggests a higher limit may be appropriate).

4.2.2 General guidance for tables

- Round data consistently in summary tables
- Use consistent units
- Use captions and row and column headings to show what the units are and what the numbers mean
- Right justify numbers in columns (or at least make all units, tens, hundreds etc. line up)
- Make sure that all tables in the BFS are in a similar format
- Show time either from left to right or top to bottom
- Show row totals to the right and column totals to the bottom
- Put data to be compared in columns rather than rows
- Keep tables as simple as possible

4.3 Graphs

Graphs are often the best way of presenting trend data or large amounts of data. Before deciding on a graphic, the following need to be considered:

- The type of data to be presented
- The key feature to be portrayed
- How the information will be used
- The intended audience (non-specialist)

The following sections should help to design a graph that is easy to understand and difficult to misinterpret!

4.3.1 3-D graphs

The general statistical guidance on 3D graphs is **DO NOT USE THEM** unless you are presenting three dimensions of data! For example in Figure 6 the data used was Good = 1, Average = 2, Poor = 3.

All three bars are below the gridlines representing 1, 2 and 3.

Figure 6: An example of a 3D bar chart

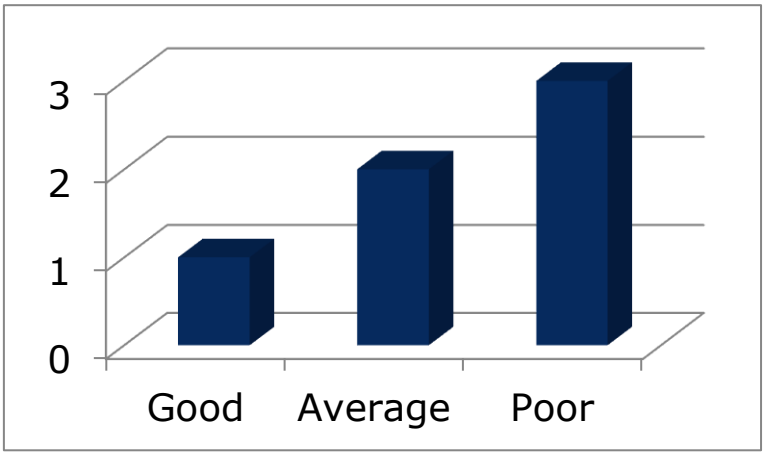
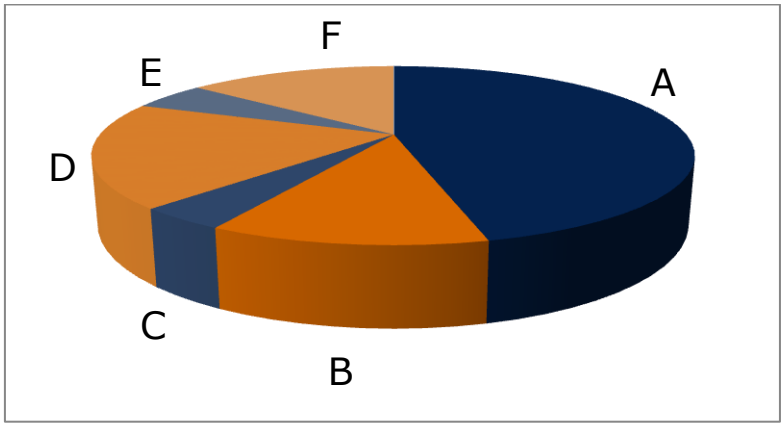


Figure 7 is also misleading. Segment B represents a proportion which is less than the proportion represented by segment F. The additional depth in the front segments and the angle that the chart is set at makes the apparent area of the segments differ from the proportions they actually represent.

Figure 7: An example of a 3D pie chart



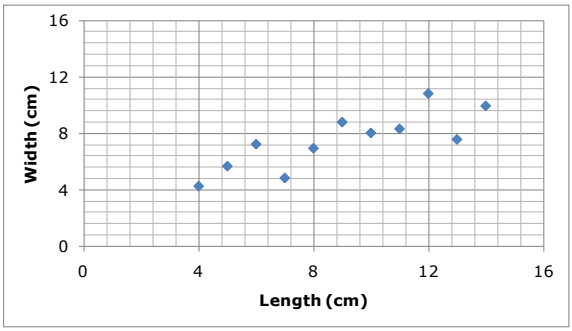
4.3.2 Gridlines

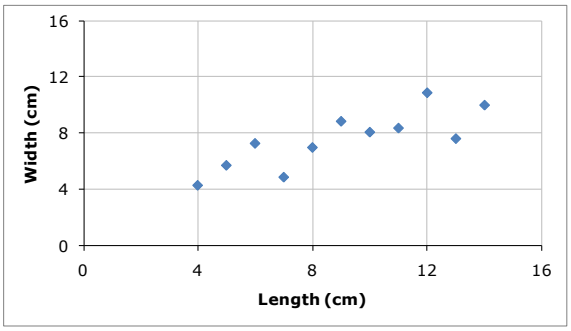
Gridlines are lines which divide graph area into segments. The idea is that they help to define the location of the data points. The important message in a graph should come from the data itself, not 'graph furniture', so a balance has to be struck between helpful guidance and distracting from the data.

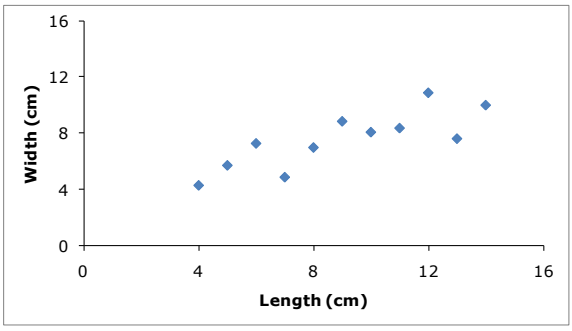
The graphs show some different choices of the use of gridlines. A compromise between too many (top) and no gridlines (bottom) seems to be the best option in this case.

The general guidelines for gridlines are:

- Use only as many that are needed to get an approximate idea of the value of any given data point in chart
- The axes of the graph should be heavier than the grid lines







4.3.3 Axes

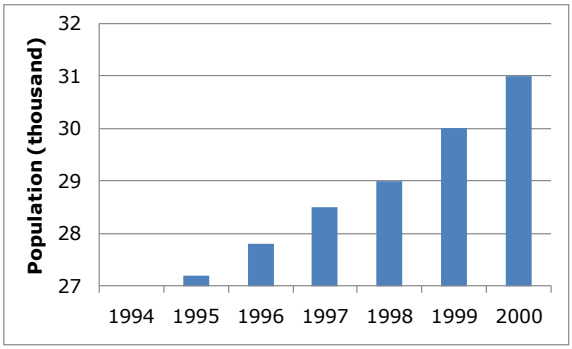
The scale on the axes of graphs is usually the part which causes the most confusion and is the part which can lead to misinterpretation.

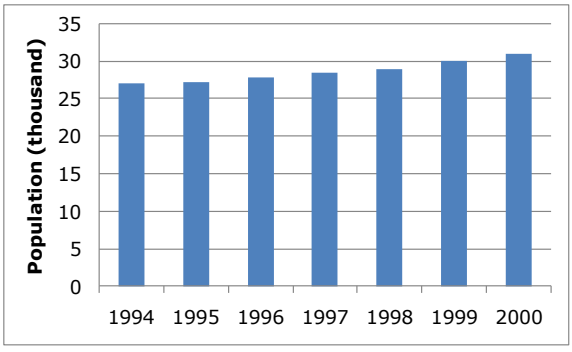
The graphs on the right show the same data – the population of a town in Berkshire from 1994 – 2000. The top graph looks far more interesting, but without taking proper notice of the scale on the y-axis (and most readers won't), you could make the following interpretations:

The population in 1994 was 0

The population doubled from 1998 to 2000

When displaying numbers that need to be interpreted as absolute numbers, the axis should start at 0.





For relative comparisons between different series consider indexing your numbers to a baseline instead of displaying two lines on the same graph using two different scaled axes (Figure 8). This example also has axes that do not start at zero and no border around the chart. An alternative presentation could be Figure 9: each index value is computed by dividing the count by the baseline figure, which, in this case, is the number of fatalities in 1999. This presentation would complement the text "The number of people killed in single vehicle accidents in 2008 was 35% less than in 1999. The total number of fatalities also fell by 34% in the 16 European Union countries over the same period."

Figure 8: An example of two axes

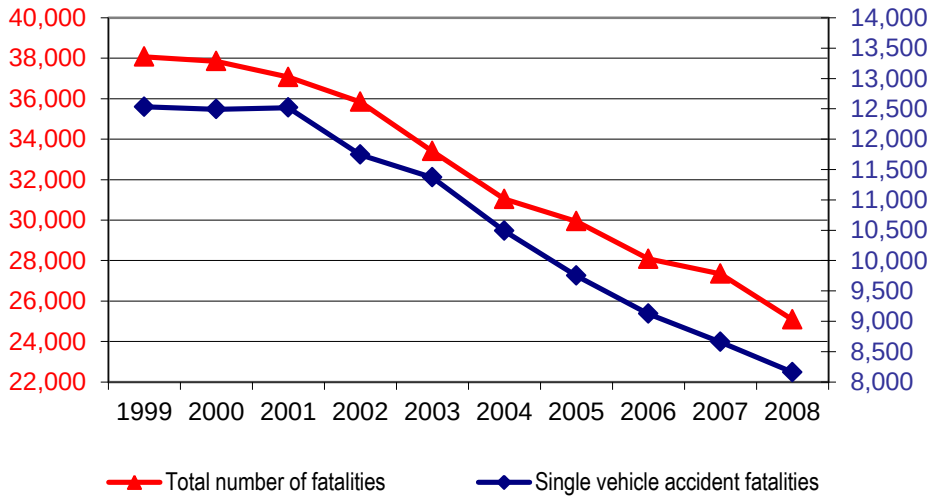
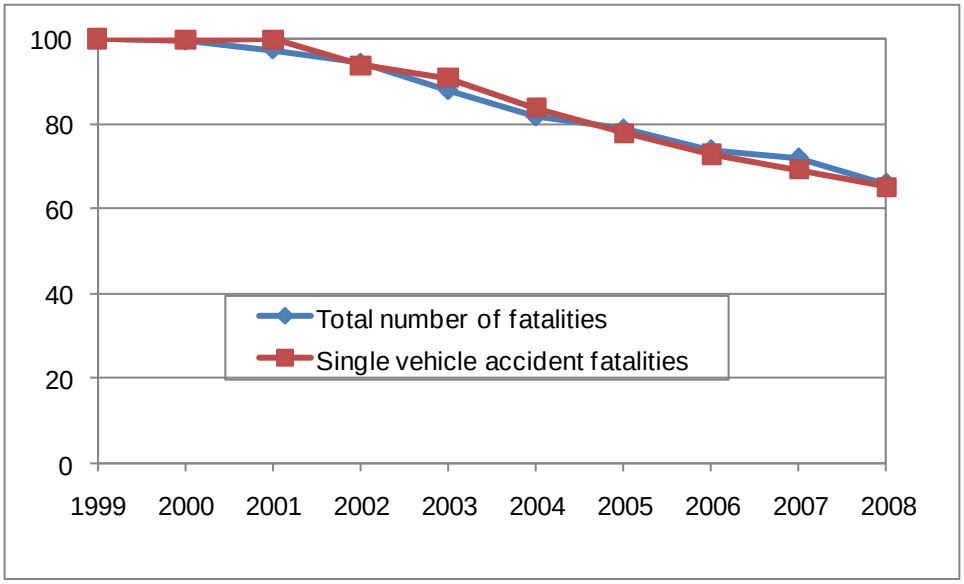


Figure 9: Relative fatality figures for single vehicle accidents indexed to 1999 values



4.3.4 Change v Reduction

An important conclusion from Figure 9 is that the number of fatalities fell between 1999 and 2008, overall as well as in single vehicle accidents. This type of information can be presented in various ways, in graphs as well as tables, but one approach should be avoided.

This approach presents the change in the number over a certain period, so that a fatality increase is shown as a positive figure or percentage, while a reduction is shown as a negative figure or percentage. While this is mathematically correct, experience shows that this is often misunderstood by readers – perhaps because people tend to interpret positive figures as "good news" whereas in this case a positive figure represents more fatalities. It is better for these graphs and tables to present fatality reductions so that a reduction (safety improvement) is shown as a positive number; if the data include negative numbers then add a note explaining that a negative number denotes an increase.

4.3.5 General guidance for graphs

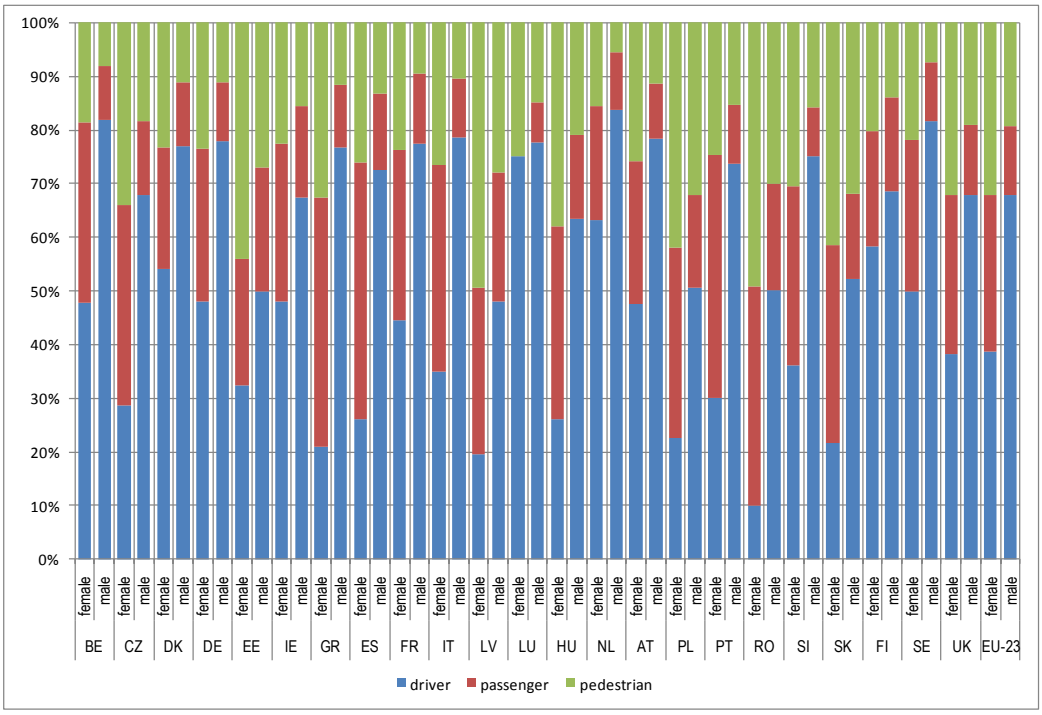
- Never use 3D (except for 3D data)
- Always put the zero point on the scale when graphing absolute numbers
- Label diagrams effectively
- Show gridlines to aid interpretation
- Round the numbers on the axes appropriately
- Ensure the picture accurately represents the data
- Make the diagram simple enough so that the reader can quickly assimilate the message

4.4 Specific graph types

4.4.1 General guidance for bar charts:

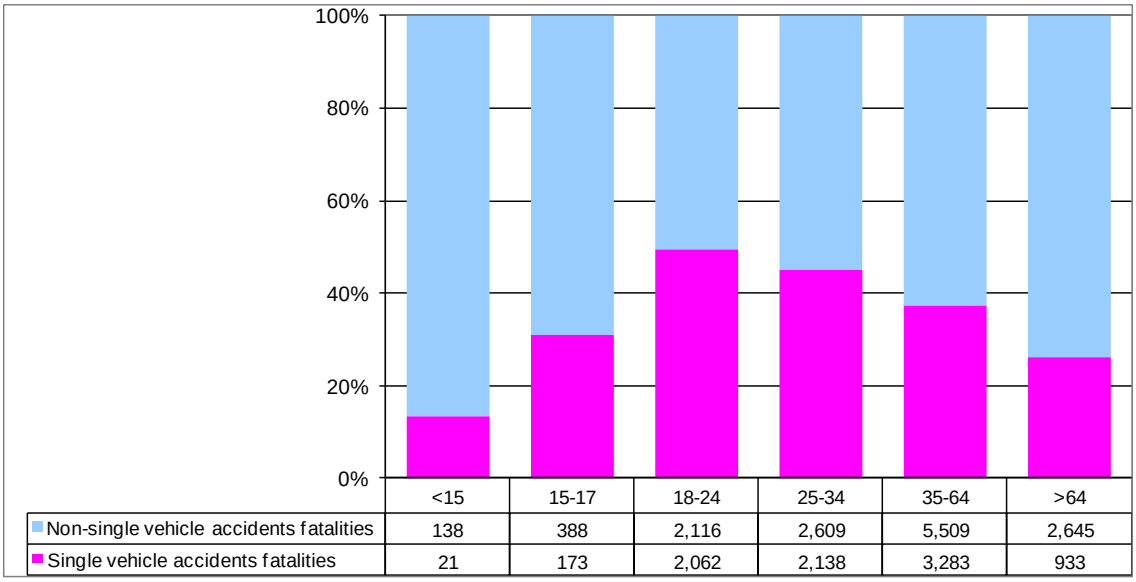
- Bar charts are useful for discrete, grouped data of ordinal or nominal scale.
- Include the zero point on a scale – otherwise the differences are exaggerated.
- Don't overlap bars as the visible area is smaller for overlapped bar.
- Stacked bar charts are good at displaying variation of a variable for all the EU countries. But always include a scale axis rather than displaying the precise percentages as data labels.
- It may be better to display percentages or rates rather than counts
- Make the graph simple enough so that the reader can quickly assimilate the message. It is not clear from Figure 10 what message the author is trying to convey as so much information is presented.

Figure 10: An example of a stacked bar chart with too much information



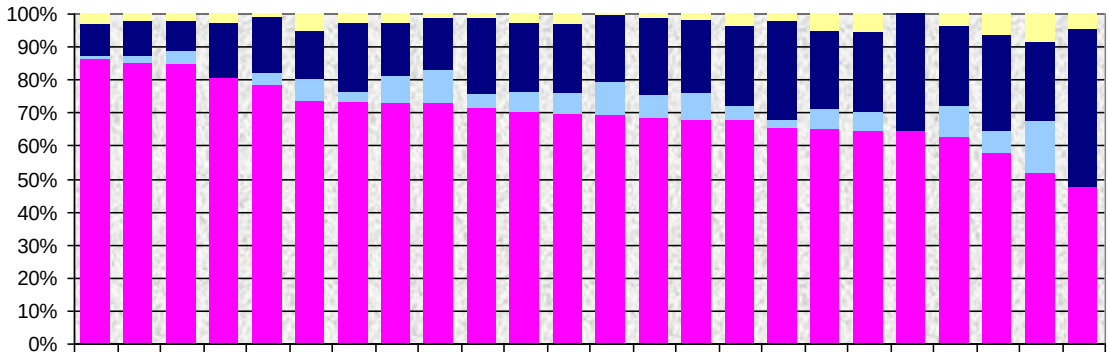
Data labels should not be used in the bar chart to communicate the precise values (especially not to 1 or 2 significant figures) as this overcomplicates the figure and the axis should be sufficiently labelled to read off a percentage. If this level of detail is needed then a table should be used. Some partners favour bar charts which include mini tables. These can sometimes be useful to show sample sizes, as in Figure 11, but a separate table would probably be better in more complicated cases such as Figure 12.

Figure 11: **Example of a stacked bar chart with the sample sizes in a mini table**



	<15	15-17	18-24	25-34	35-64	>64
Non-single vehicle accidents fatalities	138	388	2,116	2,609	5,509	2,645
Single vehicle accidents fatalities	21	173	2,062	2,138	3,283	933

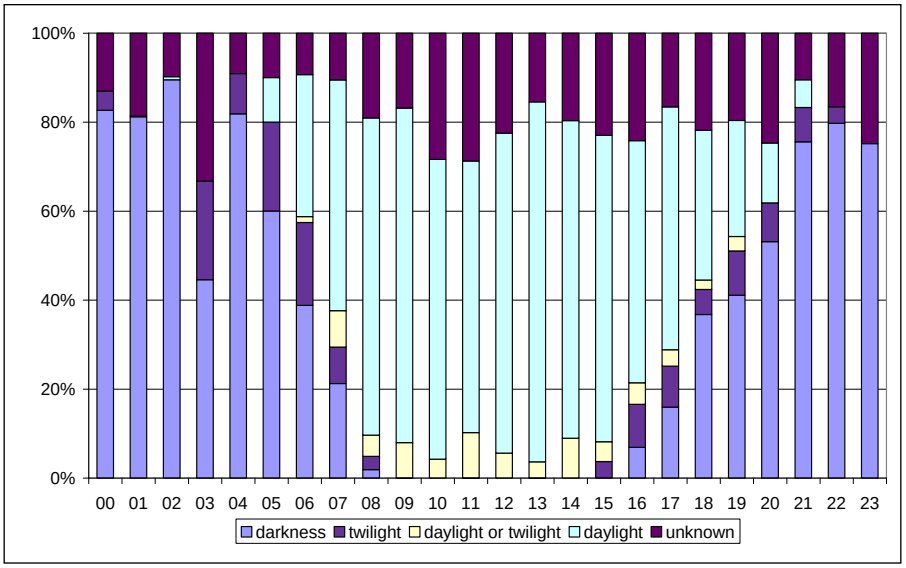
Figure 12: An example of a stacked bar chart with mini table



	LV	SK	PL	EE	CZ	RO	UK	IE*	BE	DE	EU-23	HU	NL	FR	DK	AT	IT	SE	FI	LU	ES	EL	PT	SI
Other	3%	2%	3%	3%	1%	5%	3%	3%	1%	1%	3%	3%	0%	2%	2%	4%	2%	5%	6%	0%	4%	7%	8%	5%
Two-wheelers	9%	10%	9%	17%	17%	15%	21%	16%	15%	23%	21%	21%	20%	23%	22%	24%	30%	24%	24%	36%	24%	29%	24%	48%
Lorries	1%	2%	4%	0%	4%	7%	3%	8%	10%	4%	6%	6%	10%	7%	8%	4%	2%	7%	6%	0%	9%	7%	15%	0%
Car / taxi	86%	85%	85%	81%	78%	73%	73%	73%	73%	72%	70%	70%	69%	68%	68%	68%	65%	65%	65%	64%	62%	58%	52%	48%

Think carefully about the way that categories such as age and time of day are grouped. Ensure the groups are discrete. Figure 13 is confusing because twilight and daylight appear in more than one group, it would make more sense to combine the three groups.

Figure 13: An example of non-discrete groups

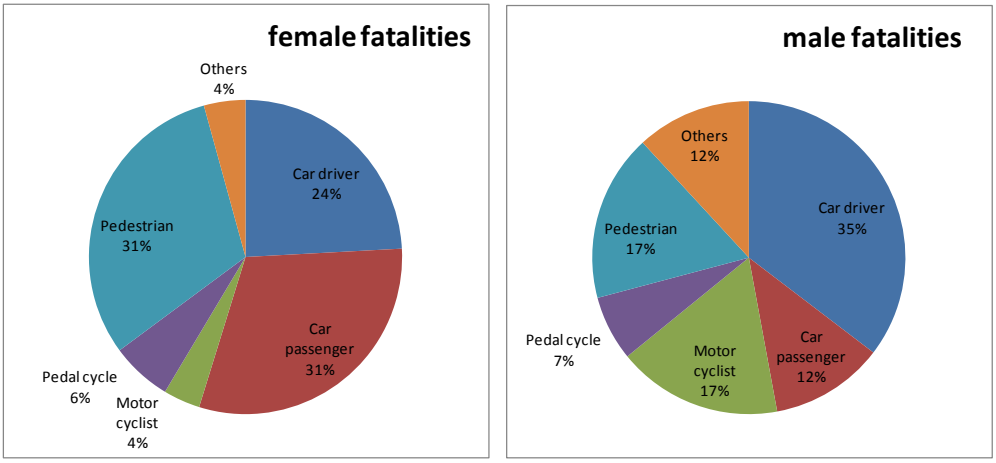


The interpretation of this figure is complicated by the variability of the percentage of unknowns by hour. The average reader will not be interested in such finer points of the CARE database; it would be better to restrict the analysis to those countries where this variable is reported well, then distribute the remaining unknown cases *pro rata*.

4.4.2 General guidance for pie charts:

- Pie charts are effective for presenting a small number of items of data
- Pie charts are useful for different sized segments
- They should only be used when the total sum of all the segments has meaning
- The optimal number of segments is 6 (between 3 and 10 is advised)
- Don't use 3D
- Do include the proportions as labels (rounded to no decimal places)

Figure 14: An example of a pie chart used appropriately



4.4.3 General guidance for line graphs (time series):

- If the line represents point measurements, add points to the lines
- The optimal maximum number of lines on one chart is 3-4, unless they are well separated
- In a set of line graphs use consistent plotting styles and colours and scales
- The selection of the range of the scale on the Y- axis (vertical axis) is very important. Always start the axis at 0 unless there is a very good reason for an alternative origin. For line charts, position the x-axis on the tick marks as shown in Figure 1 (format axis option 'position axis on tick marks').

5 The ASR

The ASR consists mainly of tables of counts of CARE fatality data, illustrated by a series of figures derived from these tables. The purpose of this is to provide a comprehensive reference book of fatality counts. As it complements the BFS, the same principles should apply to the design of Tables and Figures so that the documents are consistent.

It is important that tables in Excel spreadsheets are not imported into the final Word document as graphical images as the formatting is not preserved and can be distorted. Particular design details for the tables are listed below:

- The font size is Ariel Narrow 10pt black.
- The first row and the first column contain the column and row headings in bold; 15% shading should be used.
- All cells should have borders, normally single lines (solid, not broken); significant divisions of the table are marked with double lines (e.g. the EU totals should be separated with double lines at the end of the table.)
- Right justify numbers in columns.
- A hyphen shows that a value is unknown and a '0' is equivalent to a zero count.
- Notes should follow each table, acknowledging the data source and date of when the data was extracted.
- Each table should have a final line showing the EU total; as data will probably be missing for at least one country, this should be labelled "EU-xx". See section 3.3 for advice on calculating this total. Note: any estimated values do not appear in the table, which should only show actual CARE data.
- If data for particular countries are not from the year identified in the table heading, the countries should be marked and the exceptions listed in the table notes. The latest available data should be used for each country.
- Where tables present data by country, the codes shown on page 4 should be used in the order given.
- If there are no data available in CARE for a particular country then these countries should be omitted from the tables and graphs.

Particular design details for the figures are listed below:

- Each Figure should have a border to the chart area and to the plot area.
- The plot area should be white, not shaded.
- When preparing a bar chart, countries should be sorted in order according to the primary variable being displayed. A bar for the EU-xx should be included among the national bars so that a country's values can be compared directly with the EU value.

Annex: Colour palette with RGB definition

Christian Brandstätter has prepared the following colour palette for graphs in the ASR and BFS. On the Color and Fill menus, select 'More colours' then the 'Custom' option; the R, G and B values are then entered according to the following table. This will ensure that colours are used consistently across the ASR and all BFS.

Colours	R	G	B
1	80	130	190
2	190	80	80
3	120	150	60
4	230	110	10
5	140	180	230
6	220	150	150
7	195	215	155
8	250	190	145
9	200	215	240
10	240	190	190