User-friendly presentation of statistics

Guide to creating a dissemination strategy and dissemination guidelines for developing and transition countries

User-friendly presentation of statistics

**Guide to creating a dissemination strategy and
dissemination guidelines for developing and transition countries**

PARIS21 - Statistics Norway

Contents

Preface

In the years to come, the need for a wider dissemination of statistics from national statistical systems (including NSOs, ministries and other producers of statistics) in developing and transition countries is likely to increase substantially. In order to increase not only the quantity, but also the quality and user-friendliness of the published statistics, every NSO should develop an explicit dissemination policy, including a general strategy for future development as well as more detailed plans, actions and guidelines for the various aspects of the dissemination process.

The present publication first discusses briefly some central aspects of and issues relevant for the strategy process. Then, in the following chapters, many of the principal and practical questions and issues of dissemination are discussed in more detail, often using examples from various countries of good – and not so good – practice. These examples are not primarily intended as criticism of the various NSO that made them. They are included to illustrate some general principles and to show how they could easily be improved and because specific examples are more easily understood and remembered than descriptions of general rules and principles.

The discussion does not lead up to any specific recommendations for guidelines for the different aspects of dissemination. These are for the various NSOs to decide, depending on their own evaluation of their present situation, available resources – and their ambitions. The final chapter, however, contains a more systematic listing of the different aspects of dissemination that should be covered by a dissemination guide. Some general suggestions and recommendations are given, but in the end every NSO has to discuss and decide on its own dissemination policy and guidelines.

To formulate and implement a dissemination policy is not something that is done overnight; it is both a lengthy and demanding process. It is our hope, however, that the discussions and examples presented here, will give the NSOs a good basis for starting the process of creating their own dissemination policy.

This guide is the result of the joint efforts of PARIS21 (Partnership in Statistics for Development in the 21st Century) and Statistics Norway. The publication has been prepared in Statistics Norway by senior adviser Jan Erik Kristiansen (with a little help from his friends and colleagues).

Paris and Oslo, October 2009

PARIS21/Statistics Norway

1. Dissemination
Formulating a strategy

This chapter will raise some fundamental, strategic issues related to dissemination. These issues should be considered in every statistical agency, whether a specific dissemination strategy is to be developed or not. At the end of the chapter we outline the main steps in preparing a dissemination strategy.

Fear of dissemination?

During the last decade, there has been an increasing demand for statistics on the various aspects of development. Monitoring social, demographic and economic trends is important in order to evaluate the goals and policies expressed in the Millennium Development Goals (MDGs), Poverty Reduction Strategy Papers (PRSPs) and other policy documents. In order to meet this growing demand, there has recently been a substantial increase in the number of surveys and studies in different areas: Demographic health surveys, labour force surveys, welfare monitoring surveys, income and expenditure surveys, etc. In addition, there is also a growing collection of statistics from censuses and various administrative registers.

However, the dissemination of statistics from these numerous surveys has so far not increased to the same degree. There are many reasons for this limited dissemination of statistical results: In the preceding period there has necessarily been a strong focus on data collection and processing – with emphasis on survey methods and field work. And unlike the various aspects of statistical methods (demography, sampling, questionnaire design, estimation, etc.), dissemination is not a subject taught at universities, from which most statisticians graduate. The various aspects of dissemination are most often learnt by experience ("learning by doing"). Since dissemination is also often decentralized and frequently done by the same people who collect and prepare the data, this experience is still lacking in many countries.

In addition to a general lack of resources and experience, statisticians are by nature very cautious, and they often seem to be suffering from what could be termed "fear of dissemination". This anxiety – which until recently has also been present to a considerable degree in most NSOs – has several sources: By publishing the results, the statisticians make themselves open to criticism or disapproval –from politicians, bureaucrats and colleagues, as well as from journalists and the media. Also, in many countries, there is a long tradition for a more limited dissemination, mostly to governmental organizations. This is why statistics in many countries is "under-communicated" in the sense that its dissemination is too restricted, and often not very user-friendly.

Numerical volcanoes?

On the other hand, the present increase in the number of studies, surveys and censuses, and the growing use of various administrative data for statistical purposes, will probably before long also result in a dramatic increase in the dissemination of statistics from the NSOs in developing and transition countries.

In this situation, the activities of NSOs are often comparable to that of a volcano: After a long period of silence, there will
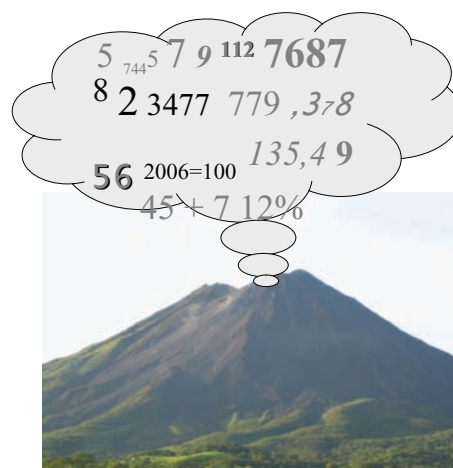
National statistical offices: Numerical volcanoes?

(usually after a major survey or census) be a sudden volcanic eruption, involving a comprehensive output of results – mostly tables and often on a very detailed level. The output is frequently so massive that users are almost drowning in numbers and they will have problems sorting out what are the most interesting and important results. What is often lacking is a more user-friendly presentation, putting the results into a context and making the statistics more understandable, informative, meaningful and thereby also more useful to the reader.

In order to improve their dissemination, it is therefore important that statistical offices have a clear dissemination policy, including a more long-term dissemination strategy. Without a clearly formulated dissemination policy, the resulting output is likely to be very much decided by the different producers, resulting in widely differing products of varying quality.

A strategy is…

To have a *strategy* simply means to look ahead and make some plans for tomorrow – plans for achieving some future goals or visions. A dissemination strategy, therefore, is a plan for reaching some defined, future goal(s) regarding dissemination: It is a *roadmap* to the future, pointing out the goals and the main direction in which we should go and helps in ranking the priorities. The strategy also includes the means to reach these goals. Whereas a dissemination guide describes the day-to-day procedures and routines of dissemination – often on a very concrete and detailed level – a strategy is the broader, forward-looking part of the dissemination policy. If a NSO does not already have a dissemination guide, then making one should be part of the strategy.

But far more important than fancy descriptions of visions and impressive slogans and goals (every NSO cannot be "a world leader") is – however – the will and ability to reach these goals. The strategy, therefore, should be realistic and achievable. And publishing reliable and understandable figures is probably more important than describing visions and missions: Action speaks louder than words!

Strategy: Bridging today and tomorrow
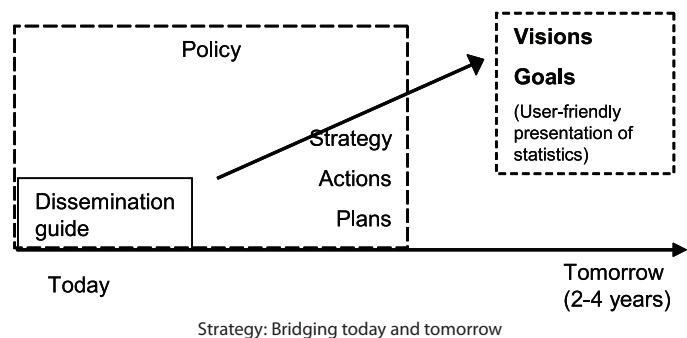
Some strategic issues

When making decisions about future dissemination, there are several strategic issues that must be discussed and decided by every NSO. These issues will be briefly presented here. Some of them will then be discussed in more detail in later chapters, considering their relevance for a dissemination guide.

Independence

In the early years of the nation-building process, statistics is an important and necessary tool for policy-making and planning and there will often be a close relationship between the NSO and the government. In the short run, this close relationship has several advantages: In addition to being financed by the government, the status and authority of the government will also carry over to the NSO.

In the long run, however, too close a relationship between statistics and the government could be damaging to the reputation of a NSO and the public's confidence in its statistics. Statistical organizations should therefore always strive to be independent, especially in their dissemination practice, as this is of vital importance for the users' trust in the statistics.

User-friendliness
The overall focus or goal of the dissemination policy should be user-friendliness, meaning that statistics should be easy to *find*, easy to *use* and easy to *understand*. This means that every NSO must make a selection between all the possible numbers and publish the most relevant, interesting and important figures. User-friendliness also means to make the numbers comparable (or as comparable as possible) and to make their comparison as easy, meaningful and informative as possible. In short, to help the users utilize the statistics in an informative way.

This overall and general goal or objective (user-friendliness) then has to be "translated" or transformed into more specific and operational secondary short-term goals and targets (plans). The specification or clarification of these goals and practices should form the main part of the dissemination guide. At the end of this chapter we try to specify how this can be done.

Role of the users
Sometimes "user-friendliness" is taken to mean something like "let's ask the users". Recently, concepts like user dialogue, user needs, user satisfaction and user surveys have begun to flourish, also in the world of statistics.

But the users of statistics are numerous and the various users may have different needs and they also differ in their capability to make use of the various statistics. And many users do not even know what they want. In addition, we want to reach new user groups. So, "asking the users" will not always bring any clear-cut answers. User-surveys will necessarily have to focus on existing solutions and practices and how to improve these. Therefore, they will seldom result in new and innovative solutions. Of course, contact with users in order to get information about their needs is both useful and necessary, but user surveys should never replace sound judgment based on knowledge of statistics and knowledge of the society. Inspiration for more user-friendly dissemination solutions can also be found by looking at the experience of other NSOs.

Another aspect of today's increasing focus on user-orientation is what is called "audience targeting". Users are classified into various groups (audiences), e.g. media, students and teachers, government organizations, non-governmental organizations (NGOs), the public, media, various experts, etc. The idea is that the NSO should decide on who the audience or target group is, and that the presentation of statistics should be adapted to the needs and qualifications of the various groups.

The problem with this approach is that it is virtually impossible to adjust presentations to the different needs and qualifications of all the various user groups. Of course, some presentations are clearly directed at experts, while other, more popular presentations are mostly for the general public. But the majority of statistics is presented in news releases and reports aimed at all the users. So the best strategy is often not to think too much about who the user is. Instead, we should try to make *all* presentations user-friendly and informative.

Quality and quantity
Considering the recent increase in surveys and studies on the various aspects of development, the coming years are probably going to see a dramatic demand for dissemination of statistics. In many NSOs there will therefore be a conflict between quality and quantity. Output is easily measured in terms of number of press releases, reports, publications, etc., whereas the quality of dissemination is not so easily measured. But when there is a conflict between quality and quantity, the priority should always be given to *quality*.

How to achieve an acceptable level of quality in dissemination is a question that should be part of the more general strategy regarding quality management.

Pricing policy

With the increasing use of the Internet and a growing understanding of statistics as a public good, the context and circumstances for the pricing policy are changing. The main principle of the Internet is that information is free of charge and there are very few – if any – examples of successful websites selling information. This has reduced the users' willingness to pay for statistics (and publications) that can be downloaded for free on the Internet.

Whereas some NSOs in developing countries see the selling of printed publications as a potential source of income, in the age of the Internet, the principle of marginal costs should be employed when pricing printed publications: The price should cover paper, printing and postage (PPP).

For the sake of simplification, the demand for and pricing of statistical products can be classified in the following way:

1. The inner circle represents the basic information requirements of the general public (media, students, libraries) and includes news releases, brochures, statistical publications in libraries, Internet dissemination, replies to minor enquiries, etc. This basic information should be defined as a public good that should be distributed free-of-charge. Also, access to database services are today increasingly given free of charge.
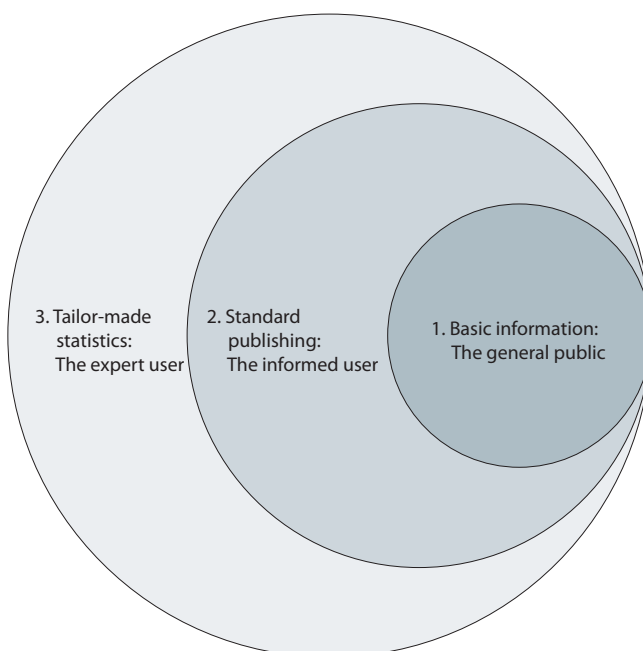
2. The middle circle represents the standard dissemination products of most NSOs: Reference table publications, analytical reports, research reports, etc. directed at the more informed users. These products are part of the information market and as such they have a fixed price. This price can be calculated in different ways. A "market price" would be what the users are willing to pay (but as most statistical bureaus are in a monopoly situation, the market price is difficult to establish). The most common principle is probably pricing based on *marginal costs* – that is the extra cost of producing/printing a publication after the tables have been compiled. Today, these products are usually also available – free of charge – on the Internet.

3. The outer circle represents dissemination to users with highly specialised and detailed needs for compilation or analysis that cannot be met by standard products. Such tailor-made products and services are individually priced, usually on the basis of the number of hours used (usually tasks of less than half an hour are not charged). But today, these user-groups are also increasingly being served by data banks, free of charge.

The role of the media
The media play a crucial role in the dissemination of statistics – not only to the general public, but to other users as well. In addition, media are important in creating confidence and trust in statistics and in national statistical organisations. Having a good relationship with the media is therefore of vital importance to every NSO. Not only to "spread the numbers", but equally important, to increase the visibility and reputation of the NSOs and their products. Using journalists for dissemination and training the media to interpret numbers may give the public better access to more statistical information. (More about media relations in chapter 6 and in appendix II).

Tables vs. analysis
Broadly speaking, printed publications can be grouped into three main categories, which for the sake of presentation could be termed: *Reference publications*, *analytical publications* and *popular presentations*:

Reference publications: Traditionally, most printed publications have been (and still are) typical reference publications aimed at the informed or expert user. These are table publications, including a variable amount of introductory text (preface, summaries, main findings and methodological notes).

Analytical publications: More recently, many NSOs have also started to publish some analytical, thematic publications directed towards the informed users; the most prominent example being publications like *Women and Men in …*, which seem to have been a successful form of dissemination. Analytical publications also include periodical publications: journals, magazines and newsletters.

Popular presentations: Thirdly, there are various popular presentations (brochures, pamphlets, posters), aimed at the general public.

The first business of a NSO is – of course – to produce statistical tables. Therefore it is only natural that the majority of publications should contain mainly tables; especially in the early days of operation: One has to tabulate before one can start analysing. On the other hand: In an increasingly complex society, there is a growing need and demand for more overreaching analyses and commentaries, putting the numbers into a context, explaining their significance: "What do the numbers really mean?" While tables are "product-specific", reporting findings from a specific study or survey, analytical presentations are often thematic; combining and analysing data from different fields of society (good examples are different types of social reports, like *Social Trends* and *Women and Men in …*). The focus of dissemination is thus moving away from the "collection-based" statistical products towards more comprehensive, user-oriented general presentations and interpretations (from distribution of numbers towards dissemination of information).

But timeliness and analysis are conflicting demands that are difficult to combine or integrate into the same product: In the future we will probably to a larger extent have to choose between the two; rapid presentation of numbers *or* a more analytical presentation. At the same time – with the spread of the Internet – we increasingly face a choice between media of dissemination: printed publications *or* electronic publishing.

Printed publications vs. electronic dissemination

Over time, there seems to be three different policies for handling printed vs. electronic dissemination:

- *Yesterday*: Printed publications first, then (maybe) electronic/Internet
- *Today*: Parallel publishing: Printed *and* on the Internet simultaneously
- *Tomorrow*: First on the Internet, then (maybe) a printed version. (In Statistics Norway – as in many other countries – an "Internet first" policy has already been implemented: New statistics are – since 1999 – released daily only on the Internet).

Combining the two preceding alternatives (tables vs. text, electronic vs. paper distribution), we face the following four possibilities, of which two alternatives are the most likely directions for the future. In the years to come, the dissemination of statistics seems to become more *differentiated* and will probably develop in two main directions:

Publishing statistics: Two main directions

	Electronic	Printed
Numbers/ Tables	1	(x)
Text/ Analysis	(x)	2

1. Numbers (in the form of tables) will increasingly be released and published electronically (via the Internet, CD-ROM and on-line databases). The main advantages of electronic media are speed of delivery and the possibility to download tables for further calculations and analysis by the user.

2. Analysis and commentaries – on the other hand – are still unmatched in printed form. And with the growing demand for different types of descriptive and analytical reports, printed publications will still have their place in the future dissemination strategy.

There is, of course, an exception to every rule: Some tables will still be printed in publications. And (short) texts can be read on the Internet (especially news releases, summaries and popular presentations), or printed by the reader. Using the PDF format allows readers to print a copy that is identical to the published pages and this format is well suited for journals and books, where the reader may only be interested in one of several articles/chapters.

So, balancing the dissemination of printed and electronic statistics and analysis will be a major strategic question in the future.

Dissemination: Centralized or decentralized?

How to organize the dissemination activities of a NSO is clearly a strategic issue. Traditionally, dissemination has been decentralized; the publishing of statistics has been the responsibility of the various producing units within the NSOs. Such decentralized dissemination often results in rather different practices and routines.

This situation is not uncommon in statistical offices, but it often implies a lack of co-ordination and standardization that makes it difficult to improve the various aspects of dissemination. Dissemination is a professional activity, and by centralising this activity, people will accumulate knowledge and experience of the different aspects of the dissemination process. A certain centralisation of the production line will also relieve the statisticians of some of the "burdens" of publishing and give them more time to concentrate on their main job; to collect data and produce reliable and valid statistics.

It is therefore strongly recommended that NSOs establish a separate unit responsible for *all* dissemination of statistics and information; in this way the NSO will speak "with one voice". Ideally, the

dissemination unit should be located close to the top management/chief statistician. The responsibilities of such a unit should (depending on resources and staffing) be to…

- contribute to a long-term and integrated dissemination strategy, including printed as well as electronic dissemination plans
- publish and inform about statistical results
- establish and maintain good relations with the media
- give advice in all information and dissemination matters
- organize the production line and co-ordinate and standardize the different aspects of production (tools/programs)
- make rules and guidelines for design and layout for publications, tables and charts and to see that these rules are being followed
- develop and accumulate knowledge about the various aspects of publishing; including design, layout and the making of tables and charts
- have the responsibility for certain "joint" publications, like yearbooks
- initiate, edit and/or co-ordinate (new) thematic publications and presentations involving different subject matter departments
- have the overall editorial responsibility for the web service
- co-ordinate printed and electronic dissemination
- take care of distribution, subscriptions, marketing, contact with users, etc.

Centralizing the overall responsibility for dissemination does *not* imply that all the dissemination activities are carried out by this unit. It means that this unit has the responsibility for developing a dissemination policy as well as tools and rules for dissemination.

Metadata
The question is not whether to publish metadata or not, but rather how and how much. NSOs should therefore develop some kind of minimum standard for metadata for all of their statistical products. The guidelines for this standard should specify what information should be included and how this information should be structured. Examples of metadata are response rates, description of samples and definitions of concepts. This standard should be included in the proposed guidelines for dissemination (see chapter 9.6 for more detailed suggestions about metadata).

The preceding issues are strategic issues in the sense that the top management of the NSO will have to make decisions concerning their place in the future development of the organization. The next step is to include the results of these decisions in the day-to-day dissemination activities by developing a dissemination policy and guidelines. In chapters 3 through 9, we discuss some of the topics and issues that should be included in such a guide.

The strategy process

A dissemination strategy describes the policy for publishing statistics and information as well as media relations. It may be formulated in words and put on paper – or it may not. Either way, the policy exists in the way information is normally disseminated. Formulating an explicit strategy makes it easier to disseminate in a coherent manner and it makes it easier for the users to know what to expect.

A strategy for dissemination will serve as support for the master plan or general strategy of the NSO. Good practice principles regarding formulation of master plans and general strategies can be found in the Guide to Designing a National Strategy for the Development of Statistics (PARIS21 2004) or the Handbook of Statistical Organization, Third Edition: The Operation and Organization of a Statistical Agency (United Nations 2003). The same topics are relevant in formulating sub-strategies such as a dissemination strategy. One difference is, however, that the emphasis on gathering support for the strategy both within the organization and among the users is somewhat weaker for the sub-strategies. The focus of the strategy should be to decide which principles to follow, and to say why they are important. The practical work to be done to meet the goals set out in the strategy ought to be elaborated in an action plan. The action plan should underpin the dissemination strategy. The goals stipulated in the strategy and the action plan must be seen in relation to the present situation of the institution and where it wants to go in the future.

The process starts with an assessment of the current situation in the areas of interest. This includes reviewing relevant documents like the relevant parts of e.g. National Strategies for Development of Statistcs (NSDS), Poverty Reduction Strategy Paper (PRSP) and Master Plan. It also includes discussions with the important stakeholders in the institution, in particular the top management and the units involved in report writing and other parts of the dissemination process. In forming the strategy it is important to consult the involved parties at the NSO to agree on what is a desired and plausible way forward.

The next step is to map out a general dissemination framework or policy, based on the consultations with the stakeholders in the institution. The aim is to put into writing the major decisions on how to handle central issues like quality management, how to disseminate on the web and by means of printed publications, relations to important users like ministries and media – and the organization of the dissemination process.

The strategy should be accepted by central users. The main purpose of involving the users is to let them tell us how we could improve the quality of the statistics. All requests for additional statistics must be viewed in light of the accessible resources. The strategy formulation process can not be expected to be linear. Strategy formulation is a process with regular feedback, monitoring, and adaptation as conditions and needs change.

The practical work of writing the dissemination strategy should be done by someone in charge of dissemination, in close cooperation with the top management. The top management should endorse the strategy and the general direction of the action plan. It is important that the relevant stakeholders feel a sense of ownership to the strategy.

The strategy should not be too lengthy. It could contain titles formulated as mission statements saying something about where the organization wants to go, like "Improving quality", "More publications on the web", "Understandable statistics", "Better media relations" or "Equal treatment of users".

A dissemination strategy suggests a general policy for achieving part of the goals in the Master Plan or the general strategy, whereas an action plan is a way of implementing the strategy. The rest of this document focuses on issues that may be useful to include in an action plan and/or in a dissemination guide. However good the strategy is on paper, it is the implementation that determines its success.

From strategy to guidelines

Below is a simplified, schematic outline of the process of "translating" an overreaching, general strategic goal into more concrete policies and plans and how these are transformed into actions and – finally – guidelines.

From strategy to guidelines: A schematic example
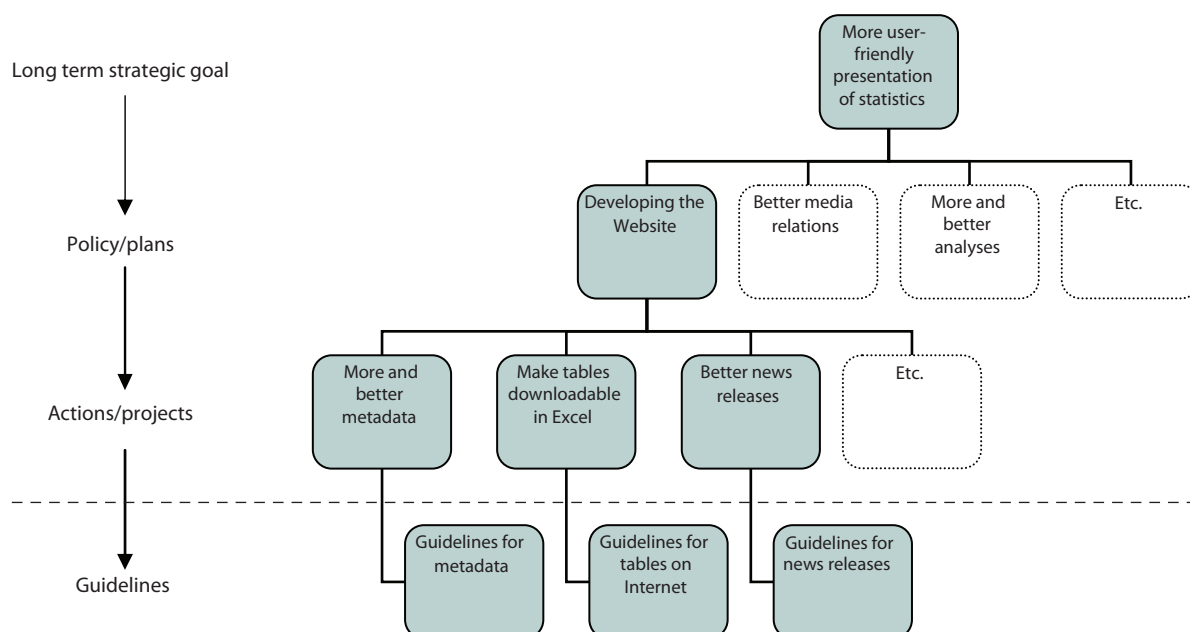
Dissemination guidelines are important for two reasons: They improve the quality of the publications by making them "nicer" and more user-friendly. But perhaps more important, clear rules and guidelines facilitate the internal production process, by making it clear what the publication or website should contain and how this content should be structured and formatted: *"This is how we do it"*. Written rules and guidelines therefore greatly facilitate the production process and the process of approval and quality management, especially as the volume of dissemination increases.

It is therefore strongly recommended that every NSO develops general guidelines for dissemination. Guidelines are a description of the day-to-day routines of dissemination. The guide should contain *rules* (which *must* be followed) as well as *guidelines* (to be seen as recommendations/advice).

The process of preparing the guidelines should involve both statisticians, people engaged in the dissemination process as well as someone with experience and knowledge of layout and formatting.

2. Comparing numbers
Making the numbers talk

Statistics gives a numerical description of society by means of numbers put together in tables or graphs. The purpose of placing numbers together in this way is to compare them in order to uncover differences, correlations and trends. *To compare numbers – after having made them as comparable as possible – is the central element of all statistics.* And user-friendliness means to present the numbers in a way that encourages and enables the users to make comparisons.

To compare is a basic human activity. In everyday life we make comparisons all the time: We compare jobs, prices, wages, cars, newspapers and politicians. But just as we cannot compare the price of two widely different car models without considering their size or quality, we cannot compare the numbers of cars or the number of births in two countries, without taking the size of the populations into consideration.

Misuse or abuse of statistics is rarely the result of deliberate manipulation of the numbers. More often, it is due to various statistical fallacies; the comparison of numbers that are not comparable. Therefore, it is important to know when and how numbers can be compared, and how the results of the comparison should be presented and expressed. This may sound commonplace, but statistics is very much about making numbers comparable – or as comparable as possible – by making sure that they are defined, collected and calculated in the same way, and by presenting them in a manner that makes it easy to understand and use them.

Numbers rarely lie, but they sometimes hide the truth. Numbers speak with two tongues, in the sense that that they seldom give clear and unambiguous answers. Therefore, numbers must be analysed and analysis requires that we understand what the numbers say – and what they don't say.

Compared to what?

The central element of all statistics is comparison: We compare numbers in order to see if there are …

- differences between groups (between men and women, between regions/districts or between various age groups)
- or changes over time
- or both: Are the differences increasing or decreasing or are they stable?

The table on the next page is taken from a selection of tables entitled "Woman statistics" on the website of the Department of Statistics, Jordan. It shows the distribution of women by educational level in urban and rural areas.

This table allows us to compare women's educational level in urban and rural areas and we find that illiteracy is rural and that women in urban areas more often have secondary and higher education.

Percentage Distribution of Jordanian Females Aged 15 Years and above by Education Level and Urban / Rural (2008)

Educational Level	Total	Urban	Rural
Percent	100.0	100.0	100.0
Illitrate	11.4	9.8	19.3
Literate	3.8	3.7	4.1
Primary	9.1	9.1	9.2
Preparatory	16.6	16.9	15.2
Basic	18.6	18.0	21.5
Vocational Apprenticeship	0.0	0.1	0.0
Secondary	19.4	20.2	15.2

From: DOS, Jordan

However, one very important and relevant comparison is missing: The comparison of men and women. Are there any gender differences in education? By not including men in these tables, we lose the possibility of an interesting comparison, the comparison of men and women.

Absolute and relative numbers

Politician A (probably from an opposition party) says:
– *The number of poor people has increased by more than 200,000 over the last five years.*

Politician B (probably from the governing party) replies:
– *Yes, but the proportion of poor people has been reduced by 3 per cent!*

These seemingly conflicting statements reflect two different ways of using and comparing numbers: While A compares the absolute numbers, B – by using percentages – makes a relative comparison.

Another example: In a company, women's wages increase by $580 in one year, while men's wages increase by $620. Men thus get an absolute wage increase of $40 more than women. But as women in this company earn less than men – $22,000 and $35,000 respectively – women's relative increase amounts to 2.6 per cent compared to 1.8 per cent for men. Again, the result is different, depending on whether we use absolute or relative numbers.

Absolute numbers are the numbers we get when we count – the "raw" numbers, so to say: In a certain year there are 65,465 births, 563 persons are killed in road accidents and the number of unemployed was 345,645. If we say that there are 15.8 births per 1,000 inhabitants, that the mortality rate in road accidents is 11.9 or that the unemployment rate was 16.8 per cent, we use relative numbers.

Relative numbers are numbers that have been related to other numbers – to enable and simplify comparison: Percentages, rates and ratios are all calculated to facilitate comparison. But as we will see later, relative numbers can be calculated in different ways. The question then becomes: Which relative numbers should be used – and in what way?

Which relative number?

The table on the right presents the male and female population of Pakistan (From: Compendium on Gender Statistics in Pakistan 2004). In the last column, the proportion of women is given as a percentage of the total. By calculating this relative number, it becomes easier to compare the male and female population over time. (This proportion could – of course – just as well have been calculated as the proportion of men).

Table 1.4 Women - Men Population, 1947-1998

Year	Women (000)	Men (000)	Both Sexes (000)	Propor-tion of women
1947	15,100	17,400	32,500	45.5
1951	15,593	18,147	33,740	46.2
1961	19,920	22,960	42,880	46.5
1972	30,476	34,833	65,309	46.7
1981	40,021	44,233	84,254	47.5
1998	63,478	68,874	132,352	48.0

Source: Population Censuses 1951, 1961, 1972, 1981 and 1998. Estimates for 1947.

In the next table in the same publication, the relative number of men and women is presented in another way – as a sex ratio: women per 100 men. But the "standard" manner of describing the sex ratio in a population seems to be "men per 100 women" (this is why, in some countries, this is called the "index of masculinity").

Table-1.5 Sex Ratio by Area, Pakistan, 1951-1998

Year	Women per Hundred Men		
	Urban	Rural	All Areas
1951	78.6	87.2	85.9
1961	79.6	88.9	86.8
1972	83.6	88.8	87.5
1981	86.7	92.0	90.5
1998	89.2	93.7	92.2

Source: Population Censuses 1951,1961,1972 and 1998

From: Compendium on Gender Statistics in Pakistan 2004

But again, in other contexts, the opposite way of calculating the sex ratio seems to be the standard procedure: When describing

gender inequality in education in MDG-reports (also termed "gender parity index"), this is typically reported as the ratio of girls to boys, as in the table on enrolment in primary education from *The Millennium Development Goals Report.*

(a) Primary education

	Ratio of girls to boys gross enrolment ratios		
	1991	2000	2006
World	0.89	0.92	0.95
Developing Regions	0.87	0.92	0.94
Northern Africa	0.82	0.91	0.93
Sub-Saharan Africa	0.83	0.85	0.89
Latin America and the Caribbean	0.97	0.97	0.97

From: UN, The Millennium Development Goals Report 2008

This is a reminder that most relative numbers can be calculated in many different ways. For instance, when calculating the gender pay gap, this can be done in at least three different ways. Let us assume that in a community men earn $1,000 per month, compared to a monthly salary of $710 for women. This difference can then be described in three different ways:

● Men earn 41% more than women
● Women earn 29% less than men
● Women's salaries are 71% of men's

Other ratios may also be calculated in different ways. In some countries, fuel consumption is calculated as miles per gallon – in other countries the measure is litres per 10 (or 100) kilometres.

A ratio describes the relationship between two numbers, for instance speed = kilometres per hour. Ratios are calculated by dividing one number by the other. Example: In school A, there are 440 pupils and 30 teachers, in school B there are 590 pupils and 51 teachers. If we want to compare the two schools, we could calculate the number of teachers per pupil: A: 30/440 = 0.07. B: 51/590 = 0.09. But by turning the ratio upside down, we get a far more meaningful and user-friendly ratio: 14.7 and 11.6 pupils per teacher, respectively. This ratio is much easier to understand.

So it is not only a question of using absolute *or* relative numbers, but also a question of *which* relative number.

Absolute *or* relative numbers?

But which numbers should we use; absolute or relative? As is often the case, the answer is: It depends on the purpose. But as a general rule we can say that relative numbers are to be preferred: To compare the number of deaths in road accidents in Malawi and Mozambique does not make sense, since the population of Mozambique is almost 50 per cent larger. By computing a relative number – the number of deaths per 100,000 population – we can more easily compare the numbers. Instead of comparing four numbers – both number of deaths and population in the two countries – we compare only two numbers. But again: Another way of calculating a relative number could be number of deaths per 100,000 cars.

Another example: To the right is a table from the 2004 census in Qatar, showing the number of males and females in different age groups that use a computer and the Internet.

السكان القطريون (4 سنوات فأكثر) مستخدمي الحاسب الآلي والإنترنت حسب النوع وفئات العمر
مارس 2004
Qataris Population (4 +) Using Computer And Internet by Gender And Age Groups
March 2004

Table No. (64)   جدول رقم (64)

|	استخدام شبكـة الإنترنت Internet Using			استخدام الحـاسب الآلى Computer Using			فئات العمر
Age Groups	مجموع Total	إناث Females	ذكور Males	مجموع Total	إناث Females	ذكور Males	
4	60	33	27	284	124	160	4
5 - 9	2227	1063	1164	6521	3127	3394	9 - 5
10 - 14	7047	3281	3766	12555	5990	6565	14 - 10
15 - 19	12318	5999	6319	15637	7785	7852	19 - 15
20 - 24	11530	5918	5612	13367	6937	6430	24 - 20
25 - 29	9535	4690	4845	10839	5417	5422	29 - 25
30 - 34	8070	3792	4278	9442	4577	4865	34 - 30
35 - 39	6425	2958	3467	7712	3662	4050	39 - 35

From: 2004 Census, Qatar

But without knowing the total number of persons in the different age groups, this table is not very informative and it is impossible to say anything about the use of PCs or Internet in various groups. Here, the number of men and women in various age groups should have been presented as percentages of the total number of persons in the group.

Generally, relative numbers like percentages, ratios and rates are preferable when comparing. But not always: When the numbers are small, we should use the absolute numbers. Saying that "33 per cent is/ has…" is not very meaningful, if the basis is four out of twelve persons.

Per cent or percentage points?

In the same way, we should be careful when computing relative changes on the basis of small numbers: When something increases from 1 to 3 per cent, this is a 200 per cent increase, but in this case the increase should be presented as 2 percentage points.

And more generally: When commenting changes in percentages, it is important to be clear about the distinction between per cent and percentage points: The graph on the right was commented in this way: "*In comparison with 2006, the number of households who have disposal of a personal computer (incl. laptop or palmtop) increased with 2.0%*". But this increase (from 21.3 to 23.3 per cent) was *not* 2 percent – it was 2 percentage points or 9 per cent.
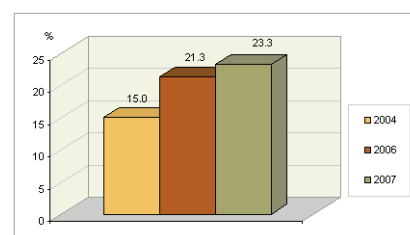
From: National Statistical Institute, Bulgaria

Persons or households?

Let us suppose that in a small community, there are four households, consisting of one, two, three and four persons, respectively. The proportion of one-person *households* is thus one out of four = 25 per cent. But the proportion of "singles" may also be reported as one person out of ten = 10 per cent of all *persons*. This example shows that it is very important to be clear about the basis of the calculation: Are we using households or persons?
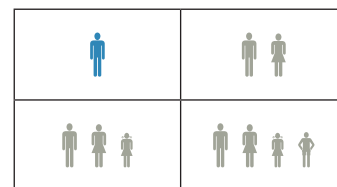
Let us now say that we are conducting a study on the ownership and use of PCs in this community. If the only information we have is that there are ten persons and two PCs in a community, we could say that the PC-diffusion rate is 20 per cent.

But to learn more about the ownership of PCs, we conduct a survey, identifying also the household structure of the population and which of the households that owns a PC: We then find that of the four households, two – the two-person and the four-person household – own a PC. We can now report this finding in two different ways: First, we could say that 50 per cent of all *households* (two out of four) own a PC. But we could also report that 60 per cent of all *persons* are living in a household owning a PC. The difference between the two figures is caused by the fact that large households are more likely to have a PC than small households.

We now have three different estimates of the PC-ownership rate:

20 per cent: 20 PCs per 100 inhabitants

50 per cent: 50 per cent of all households own a PC

60 per cent: 60 per cent of all persons are living in a household with a PC

These seemingly differing results are due to the different ways of calculating the percentages. In this case the best (and most informative) way of reporting the PC-ownership rate, is probably by using persons in households as the basis of the calculation, as everybody in a household most likely have access to the PC.

Decimals and rounding

Decimals often give an impression of a level of accuracy or precision that is unfounded. Percentages are frequently reported with two decimals, as in the following two tables:

National Bureau of Statistics   *The Nigerian Statistical Fact Sheets*

TABLE 55
Employment Gender Ratio by Economic Activities (1999- 2005)

ECONOMIC ACTIVITY	EMPLOYMENT GENDER RATIO (%)															
	1999		2000		2001		2002		2003		2004		2005		NATIONAL	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Agriculture	86.14	13.86	84.30	15.70	82.29	17.71	84.34	15.66	81.04	18.96	81.10	18.90	78.24	21.76	82.49	17.51
Manufacturing & Processing	88.53	11.47	89.50	10.50	89.60	10.40	90.01	9.99	87.66	12.34	87.81	12.19	87.70	12.30	88.69	11.31
Building & Construction	94.01	5.99	94.02	5.98	94.56	5.44	93.73	6.27	94.24	5.76	93.37	6.63	93.41	6.59	93.91	6.09
Hotels, Restaurants & Tourism	67.27	32.73	66.48	33.52	64.48	35.52	64.34	35.66	64.08	35.92	64.55	35.45	65.13	34.87	65.19	34.81
Transport	88.88	11.12	89.18	10.82	90.83	9.17	88.30	11.70	88.42	11.58	87.92	12.08	85.15	14.85	88.38	11.62
Communications	83.57	16.43	84.87	15.13	68.95	31.05	68.71	31.29	72.96	27.04	70.92	29.08	71.31	28.69	74.47	25.53
Education	50.00	50.00	48.43	51.57	58.71	41.29	58.79	41.21	60.01	39.99	65.89	34.11	62.35	37.65	57.74	42.26

From: "The Nigerian Statistical Fact Sheet" (2006), NBS, Nigeria

5.9 Özürlülük oranı, 2002
The proportion of disability, 2002   (%)

	Toplam özürlü nüfus Total disabled population			Ortopedik, görme, işitme, dil ve konuşma, zihinsel özürlü nüfus Orthopedically, seeing, hearing, speaking and mentally disabled population			Süreğen hastalığa sahip olan nüfus Population having chronic illnesses		
	Toplam Total	Erkek Males	Kadın Females	Toplam Total	Erkek Males	Kadın Females	Toplam Total	Erkek Males	Kadın Females
Türkiye - Turkey	12.29	11.10	13.45	2.58	3.05	2.12	9.70	8.05	11.33
Yaş grubu - Age group									
0-9	4.15	4.69	3.56	1.54	1.70	1.37	2.60	2.98	2.20
10-19	4.63	4.96	4.28	1.96	2.26	1.65	2.67	2.72	2.63
20-29	7.30	7.59	7.04	2.50	3.34	1.74	4.80	4.24	5.30
30-39	11.44	10.43	12.42	2.56	3.18	1.95	8.89	7.26	10.46
40-49	18.07	15.15	21.08	2.65	3.29	1.99	15.43	11.86	19.09
50-59	27.67	22.56	32.67	3.23	3.73	2.74	24.44	18.83	29.94

From: Turkey's Statistical Yearbook, 2005

Even if such tables are based on a very large sample or a census, there is hardly ever any reason to use more than one decimal: The second decimal is just numerical "noise" and does not help to improve our understanding of the data. Besides, the table and numbers become much easier to read, if we drop the second decimal.

An exception to this rule is when reporting very small percentages, such as annual population growth: 0.76 per cent.

Bearing in mind that statistics is about comparing numbers; the above table (from the Nigerian Statistical fact Sheet) also illustrates the importance of structure. Here, the most interesting comparison is probably to compare the gender ratio over time. If we restructure the table in the following way, comparison over time becomes much easier.

	Males							Females						
	1999	2000	2001	2002	2003	2004	2005	1999	2000	2001	2002	2003	2004	2005
Agriculture	8.1	84.3	82.3	84.3	81.0	81.1	78.2	13.9	15.7	17.7	15.7	19.0	18.9	21.8
Manufacturing & processing	88.5	89.5	89.6	90.0	87.7	87.8	87.7	11.5	10.5	10.4	10.0	12.3	12.2	12.3
Building & construction	94.0	94.0	94.6	93.7	94.2	93.4	93.4	6.0	6.0	5.4	6.3	5.8	6.6	6.6
Hotels, restaurants & tourism	76.3	66.5	64.5	64.3	64.1	64.6	65.1	32.7	33.5	35.5	35.7	35.9	35.5	34.9

Since this table presents the relative distribution of a variable with only two values, it could also be simplified by leaving out the figures for males or females, reporting only one part of the distribution.

Life expectancy of women and men at birth by regions in 2007

(number of years)

	Total population	Women	Men
Republic of Kazakhstan	66,34	72,58	60,70
Akmolinskaya	63,77	70,23	58,17
Aktyubinskaya	66,17	72,81	59,56
Almatinskaya	66,37	71,94	61,40
Atyrauskaya	66,56	73,11	60,59

From: Women and Men of Kazakhstan, 2008

Another indicator which is often presented with two decimals is life expectancy (see examples from Kazakhstan and Bermuda on the right).

Life expectancy by race and sex continues to improve

Sex and race	1950	1960	1970	1980	1991	2000	Absolute increase 2000—1950
				At birth			
All races							
Total	64.85	67.85	70.32	73.11	74.34	77.67	12.82
Male	62.27	64.80	68.33	69.27	71.06	74.74	12.47
Female	67.48	71.16	75.06	77.14	77.78	80.44	12.96
Black							
Total	63.03	65.25	68.50	72.26	72.43	76.59	13.56
Male	61.28	62.22	65.51	67.98	68.25	73.25	11.97
Female	64.84	68.68	72.70	76.98	76.81	79.69	14.85
White & Other							
Total	68.04	71.78	72.60	74.11	77.25	80.37	12.33

From: The Changing Face of Bermuda›s Seniors, November 2005

This indicator, however, is a prediction, estimated on basis of today's mortality rates and such a degree of accuracy is simply not possible. Even one decimal is probably one too many.

In tables, a special type of rounding is often used: The numbers are rounded to the nearest hundred and presented like this – as thousands with one decimal (table on the right from DOS, Jordan). The idea is that it is easier to read "shortened" numbers. But when

Indicator	2006	2005	2004	2003
Number of Registered Births (000)	163.0	152.3	150.2	148.3
Number of Registered Deaths (000)	20.4	17.9	17.0	16.9
Number of Registered Marriages (000)	59.3	56.4	53.8	48.8
Number of Registered Divorces (000)	11.4	10.2	9.8	9.0

From: Jordan in Figures. 2006

reading this table, we have to do some mental arithmetic to grasp the numbers – we have to remove the decimal sign and add two zeros (00).

For most people, the table on the right is probably easier to read, even if the numbers have more digits.

With larger numbers, however, this type of "shortening" is more acceptable: 3,677,054 may be written as 3.7 millions. Thus, in the graph below, the values on the value axis could be presented as million tons.

Indicator	2006	2005	2004	2003
Births	163,000	152,300	150,200	148,300
Deaths	20,400	17,900	17,000	16,900
Marriages	59,300	56,400	53,800	48,800
Divorces	11,400	10,200	9,800	9,000
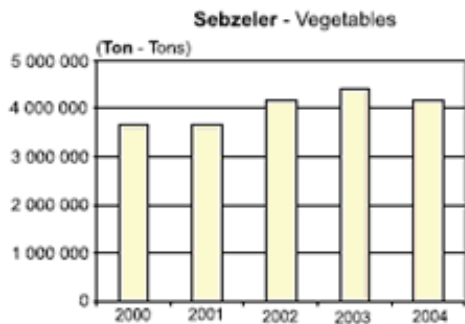
From: Turkeys Statistical Yearbook, 2005
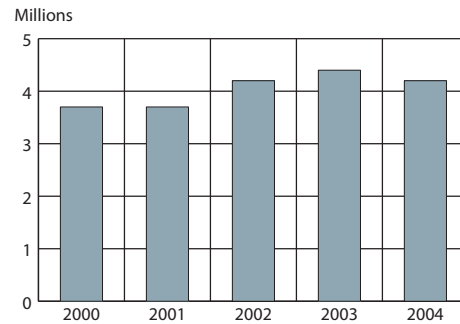
So, according to the general rule of rounding, the digits 1, 2, 3 and 4 are rounded down, while 5, 6, 7, 8 and 9 are rounded up. An important exception to this rule is when people in surveys are asked about their age: A person, who is 43 years and 10 months, will, as a rule, give his/her age as 43 (rounding down). This means that we, when calculating the mean age in sample, should add 0.5 years, assuming that every person is (on average) 6 months older than the reported age.

3. In columns and rows
Constructing tables

A table is simply a way to organise the presentation of (several) numbers in columns and rows and its main function is to make it easier to compare the numbers. And despite the increasing use of graphs, maps and other means of visualization, tables will also in the future continue to be an important tool for presenting and disseminating statistics. Therefore, their construction and design is very important.

Two types of tables

There are two main types of tables:

1. Reference tables (also called "library tables"/"source tables" or "documentation tables") and

2. Presentation ("summary"/"demonstration") tables.

Reference tables
- These are for future reference or documentation purposes
- They are often big (typically: One page or more)
- They are detailed (many indicators/classifications).
- They often present exact, absolute numbers
- They often have a standardized structure
- They are now increasingly being replaced by data warehouses and data banks, where users can construct their own tables

Reference publications belong to standard dissemination procedures in every NSO and they will probably constitute the main publishing channel for many years to come. Because of their importance, the content and structure of these publications should be constantly scrutinised and evaluated.

The number and size of these publications have a tendency to grow year by year, and this increase in "output" is often used as a measure of "productivity". But instead of producing more and more tables, one should ask: Is this table relevant, interesting, meaningful and understandable? If the answer is no, the table should probably be improved, replaced or dropped.

Typical reference table

4.1 İlkokul ve İlköğretim okullarında; okul ve cinsiyetlerine göre öğretmen, öğrenci ve diploma alanlar sayısı
Number of schools; teachers, enrollment and graduates by sex in primary schools

Öğretim yılı	Okul	Öğretmen Teachers			Öğrenci Enrollment			Diploma alanlar Graduates		
School year	Schools	Toplam Total	Erkek Male	Kadın Female	Toplam Total	Erkek Male	Kadın Female	Toplam Total	Erkek Male	Kadın Female
1923/24	4 894	10 238	9 021	1 217	341 941 (1)	273 107	62 954	- (3)	-	-
1924/25	5 987	13 822	10 480	3 342	390 368	301 381	88 987	- (3)	-	-
1925/26	5 975	14 309	10 445	3 864	406 788	313 893	92 895	- (3)	-	-
1926/27	5 995	14 231	10 864	3 367	435 563	348 978	86 585	- (3)	-	-
1927/28	6 043	15 194	10 948	4 246	461 985 (2)	325 695	133 969	- (3)	-	-
1928/29	6 600	15 718	11 153	4 565	477 569	323 260	154 309	26 275	19 031	7 244
1929/30	6 562	15 308	10 898	4 410	469 071	308 028	161 043	19 823	14 047	5 776
1930/31	6 598	16 318	11 504	4 814	489 299	315 072	174 227	21 179	15 175	6 004
1931/32	6 713	16 973	11 810	5 163	523 611	335 921	187 690	21 928	15 041	6 887
1932/33	6 778	15 064	10 440	4 624	567 361	366 125	201 236	24 740	16 864	7 876
1933/34	6 383	15 123	10 320	4 803	591 169	385 247	205 922	26 680	18 417	8 263
1934/35	6 402	15 102	10 216	4 886	647 360	426 798	220 562	30 079	20 934	9 145
1935/36	6 275	14 949	9 996	4 953	688 102	454 128	233 974	37 700	26 451	11 249
1936/37	6 202	14 777	9 805	4 972	714 178	473 720	240 458	42 208	29 435	12 773
1937/38	6 700	15 775	10 483	5 292	764 691	509 949	254 742	48 638	33 950	14 688
1938/39	7 862	17 120	11 607	5 513	813 636	547 180	266 456	55 953	39 375	16 578
1939/40	9 418	19 063	13 306	5 757	905 139	619 246	285 893	57 054	40 842	16 212
1940/41	10 596	20 564	14 583	5 981	955 957	661 279	294 678	71 854	53 710	18 144
1941/42	10 948	22 042	15 858	6 184	939 829	650 455	289 374	68 889	51 060	17 829
1942/43	11 404	21 613	15 202	6 411	940 411	649 471	290 940	71 107	51 670	19 437
1943/44	12 182	22 387	15 865	6 522	995 999	680 384	315 615	76 663	56 157	20 506
1944/45	12 903	25 687	18 545	7 142	1 246 818	804 1				3 950
1945/46	14 010	27 317	19 736	7 581	1 357 740	865 8				3 224
1946/47	15 131	30 206	22 208	7 998	1 413 983	899 2				9 373
1947/48	15 925	32 081	23 743	8 338	1 448 093	926 4				1 218
1948/49	16 119	33 185	24 502	8 683	1 468 382	938 5				3 387
1949/50	17 106	34 822	25 922	8 900	1 591 039	1 003 4				1 633
1950/51	17 428	35 871	26 714	9 157	1 616 626	1 016 915	599 711	145 899	102 888	43 011
1951/52	17 417	36 510	27 022	9 488	1 643 034	1 035 328	607 706	122 592	107 718	14 874
1952/53	17 763	37 561	27 612	9 949	1 692 135	1 066 739	625 396	148 843	105 989	42 854
1953/54	17 948	37 932	27 542	10 390	1 762 351	1 104 477	657 874	170 033	119 227	50 806
1954/55	18 393	40 035	29 230	10 805	1 866 666	1 164 616	702 050	175 712	121 891	53 821
1955/56	18 724	42 169	31 011	11 158	1 983 668	1 238 327	745 341	186 980	128 351	58 629
1956/57	19 390	44 700	33 124	11 576	2 140 347	1 335 704	804 643	212 512	143 680	68 832
1957/58	20 030	47 322	35 017	12 305	2 260 451	1 412 228	848 223	230 358	154 137	76 221

> Most tables in a reference publication are structured in the same, standardised way; very often indicators at the top and classification variables in the first column.

From: Statistical Indicators 1923-2004. Turkstat

Presentation tables
- Are smaller and simpler (extract from or summary of a reference table)
- Their main function is to present the numbers in a user-friendly way
- Presentation of "indicators" (percentages, rates, indices, averages) rather than absolute numbers
- Numbers are often rounded
- They are focused (few variables; often only two or three). In this respect, presentation tables can be compared to graphs
- Used in news releases, presentations/analyses/reports to illustrate some specific point

Typical presentation table

Life expectancy at birth in selected countries. 2006

	Women	Men
Japan	85.8	79.0
Spain	84.4	77.7
France	84.4	77.3
Switzerland	84.2	79.2
Finland	83.1	75.9
Sweden	83.1	78.8
Norway	**82.9**	**78.2**
Iceland	82.9	79.5
Austria	82.8	77.2
Germany	82.4	77.2
Belgium	82.3	76.6
Portugal	82.3	75.5
Ireland	82.1	77.3
Netherlands	82.0	77.7
Greece	81.9	77.2
Denmark	80.7	76.1
Poland	79.7	70.9

Source: Eurostat and Statistics Bureau of Japan.

Formatting tables: Some basic principles

Depending on the tools of production and the media of dissemination, the formatting of tables may be done in different ways. Nevertheless, some basic principles should be followed.

The table below shows the total population in some African countries in 1995 and 2005 and is the result of "standard formatting" in Excel.

Total population in selected African countries

	1995	2005
Angola	12279700	15941400
Eritrea	3097300	4401400
Kenya	27225900	34255700
Madagascar	13945500	18605900
Malawi	10110500	12883000
Mozambique	15853700	19792300
South Africa	41894000	47431800
Uganda	20893300	28816200
Zambia	9559400	11668500

Source: Gender, Poverty and Environmental Indicators on African Countries. Volume VII. African Development Bank 2006

This is a relatively small and simple table. But still it is rather difficult to read, because all columns are centred. To increase readability, the text column should be justified to the left, whereas the numbers should be right-justified. Also, there is no need for columns to be this wide. And all the vertical lines and most of the horizontal ones should be dropped. The resulting version is better, but still not very user-friendly.

Total population in selected African countries

	1995	2005
Angola	12279700	15941400
Eritrea	3097300	4401400
Kenya	27225900	34255700
Madagascar	13945500	18605900
Malawi	10110500	12883000
Mozambique	15853700	19792300
South Africa	41894000	47431800
Uganda	20893300	28816200
Zambia	9559400	11668500

Since population figures are very large, grouping the digits will increase their readability. We therefore insert a comma before every third digit. (In some countries, the grouping is done by using a blank space).

Total population in selected African countries

	1995	2005
Angola	12,279,700	15,941,400
Eritrea	3,097,300	4,401,400
Kenya	27,225,900	34,255,700
Madagascar	13,945,500	18,605,900
Malawi	10,110,500	12,883,000
Mozambique	15,853,700	19,792,300
South Africa	41,894,000	47,431,800
Uganda	20,893,300	28,816,200
Zambia	9,559,400	11,668,500

If this table is used for presentation purposes rather than as a reference table, we may further simplify the table by rounding to millions, using one decimal.

Total population in selected African countries. Millions

	1995	2005
Angola	12.3	15.9
Eritrea	3.1	4.4
Kenya	27.2	34.3
Madagascar	13.9	18.6
Malawi	10.1	12.9
Mozambique	15.9	19.8
South Africa	41.9	47.4
Uganda	20.9	28.8
Zambia	9.6	11.7

Furthermore, since the table presents population figures for two years, the main purpose is probably to compare the population increase in the period. In order to make this comparison easier, we could calculate the relative (percentage) growth in a separate column.

Finally, the table could also be sorted by size of population, in descending order. This way we can more easily see the relative size of the populations. Alternatively, the table could have been sorted by growth rate.

Total population in selected African countries. Millions

	1995	2005	% growth
South Africa	41.9	47.4	13.2
Kenya	27.2	34.3	25.8
Uganda	20.9	28.8	37.9
Mozambique	15.9	19.8	24.8
Madagascar	13.9	18.6	33.4
Angola	12.3	15.9	29.8
Malawi	10.1	12.9	27.4
Zambia	9.6	11.7	22.1
Eritrea	3.1	4.4	42.1

Columns and rows

A table consists of columns and rows, but there are no clear rules for deciding which variables go where. In the table to the right (from NBS, Tanzania), years have been put in the head of the table and food crops in the first column.

Production of Food Crops in Tanzania Mainland 1994-2002 (Quantities in 000 tonnes)

Food crops	1994	1995	1996	1997	1998	1999	2000	2001	2002
Maize	1,458	2,875	2,822	2,386	2,073	2,848	2,870	3,348	3,495
Paddy	192	517	495	413	847	439	443	1,010	1,054
Wheat	44	47	49	51	53	68	61	65	68
Millet	295	222	269	195	50	76	72	74	77
Sorghum	258	443	360	449	249	363	365	364	380
Cassava	1,697	1,812	1,873	1,936	2,048	2,187	2,118	2,007	2,095

Source: Ministry of Agriculture/National Bureau of Statistics

If the main purpose here is to compare the different quantities of food crops, this table is ok. But if the main purpose is to study the changes over time for the different crops, the table should be "turned around", because it is generally much easier to compare numbers vertically. (We can now more easily see the increase in paddy or the decrease in millet).

	Maize	Paddy	Wheat	Millet	Sorghum	Cassava
1994	1,458	192	44	295	258	1,697
1995	2,875	517	47	222	443	1,812
1996	2,822	495	49	269	360	1,873
1997	2,386	413	51	195	449	1,936
1998	2,073	847	53	50	249	2,048
1999	2,848	439	68	76	363	2,187
2000	2,870	443	61	72	365	2,118
2001	3,348	1,010	65	74	364	2,007
2002	3,495	1,054	68	77	380	2,095

Tables with absolute *and* relative numbers

Sometimes we want to present both absolute and relative numbers in the same table. In the table below (from the Statistical Yearbook of Turkey, 2005), we find a mixture of numbers and percentages (with two decimals) of illiterate and literate males and females; making it difficult to compare both numbers and percentages.

3.11 Okuryazarlık ve cinsiyete göre nüfus
      Population by literacy and sex
      [6 ≥ yaş - age]

Okuryazarlık - Literacy		1975	1980	1985	1990	2000
Erkek - Males						
Okuma yazma bilmeyen - Illiterate		4 096 110	3 802 455	2 932 964	2 779 172	1 857 132
	(%)	23.79	20.02	13.48	11.19	6.14
Okuma yazma bilen - Literate		13 118 658	15 188 078	18 824 697	22 066 860	28 384 266
	(%)	76.21	79.98	86.52	88.81	93.86
Bilinmeyen - Unknown		41 645	8 568	43 193	10 496	4 047
Kadın - Females						
Okuma yazma bilmeyen - Illiterate		8 048 078	8 394 868	6 770 698	6 808 809	5 732 525
	(%)	49.49	45.33	31.84	28.02	19.36
Okuma yazma bilen - Literate		8 212 708	10 123 133	14 497 065	17 488 623	23 875 115
	(%)	50.51	54.67	68.16	71.98	80.64
Bilinmeyen - Unknown		13 406	6 521	43 720	9 150	6 158

Not. Oranlar hesaplanırken bilinmeyen kapsanmamıştır.   Note. Proportions are calculated by excluding unknown.

From: Turkey's Statistical Yearbook, 2005

A more user-friendly way of structuring this table, would be to present all the absolute numbers first, followed by percentages (leaving out the second decimal place):

	1975	1980	1985	1990	2000
Males			*Number*		
Illiterate	4 096 110	3 802 455	2 932 964	2 779 172	1 857 132
Literate	13 118 658	15 188 076	18 824 697	22 066 860	28 384 266
Unknown	41 645	8 568	43 193	10 496	4 047
Females					
Illiterate	8 048 078	8 394 868	6 770 698	6 808 809	5 732 525
Literate	8 212 708	10 123 133	14 497 065	17 488 623	23 875 115
Unknown	13 406	6 521	43 720	9 150	6 158
Males			*Per cent*		
Illiterate	23.8	20.0	13.5	11.2	6.1
Literate	76.2	80.0	86.5	88.8	93.9
Females					
Illiterate	49.5	45.3	31.8	28.0	19.4
Literate	50.5	54.7	68.2	72.0	80.6

Simplify tables …

In the table above, the main purpose is probably not to show the development in population trends, but rather to compare literacy trends among men and women. For this purpose, the absolute numbers could have been left out and the table could have been reduced to the bottom part – the per cent distribution. And because the dependent variable has only two values, we could also drop one of the values, presenting only the percentage literate (or illiterate) among men and women, like this:

	1975	1980	1985	1990	2000
Males	76.2	80.0	86.5	88.8	93.9
Females	50.5	54.7	68.2	72.0	80.6

The following table is taken from INE, Mosambique and shows the total number of crimes, the solved crimes and the unsolved crimes, both shown as absolute numbers and as percentages and distributed by province.

Quadro 4.7 Operatividade dos Crimes em Geral, Segundo Província. N

Província	Crimes Conhecidos	Crimes Esclarecidos	Crimes Não Esclarecidos	% Esclarecidos	% Não Esclarecidos
País (Total)	36313	26975	9338	74	26
Niassa	1492	1153	339	77	23
Cabo Delgado	1405	1135	270	81	19
Nampula	3154	2664	490	84	16
Zambézia	2980	2734	246	92	8
Tete	2308	1953	355	85	15
Manica	1216	1087	129	89	11
Sofala	4070	3396	674	83	17
Inhambane	2645	2004	641	76	24
Gaza	1979	1596	383	81	19
Maputo Província	3743	2809	934	75	25
Maputo Cidade	11321	6444	4877	57	43

Fonte: Comando Geral da Polícia (PRM)

Since we here have a variable with only two values (solved/ unsolved crimes), we can drop one of the values without losing any information. The following, simpler table contains almost the same information as the first one, but it is much more user-friendly.

Here we present only the total number of crimes and the detection rate. The other figures in the first table can – if required - be calculated on basis of these two indicators.

Below is another table (from Turkstat), showing "literacy rates" by sex since 1935. The table is in a way simple, but still difficult to read, because the categories "Total, Male and Female" have been replaced by "codes": A, B and C. In addition, the table reports both literacy *and* illiteracy rates (Total 1935: 18.7 + 81.3 = 100.0 per cent).

Quadro 4.7. Operatividade dos Crimes em Geral, Segundo Província. 2000

Província	Crimes Conhecidos	% Esclarecidos
País (Total)	**36 313**	**74**
Niassa	1 492	77
Cabo Delgado	1 405	81
Nampula	3 154	84
Zambézia	2 980	92
Tete	2 308	85
Manica	1 216	89
Sofala	4 070	83
Inhambane	2 645	76
Gaza	1 979	81
Maputo Província	3 743	75
Maputo Cidade	11 321	57

Fonte: Comando Geral da Polícia (PRM)

1.12 Yetişkin nüfusun cinsiyete göre okuryazarlık oranı(%)
  Adult population literacy rate by sex

  [15 ve daha yukarı yaştaki nüfus -Population 15 years of age and over]

  A. Toplam - Total   B. Erkek - Male   C. Kadın - Female

%

Sayım yılı Census year		Okuma yazma bilen oranı Literate	Okuma yazma bilmeyen oranı Illiterate
1935	A	18.7	81.3
	B	30.8	69.2
	C	8.0	92.0
1945	A	28.5	71.5
	B	44.3	55.7
	C	13.5	86.5
1950	A	31.9	68.1
	B	47.7	52.3
	C	16.7	83.3
1955	A	38.8	61.2
	B	56.3	43.7
	C	21.3	78.7
1960	A	38.1	61.9
	B	54.8	45.2
	C	21.1	78.9
1965	A	46.2	53.8
	B	64.7	35.3
	C	27.6	72.4

From: Statistical Indicators 1923-2004, Turkstat

By rearranging and simplifying the table (dropping the "Illiterate" and "Total"), we can more easily follow the trends in male and female literacy.

... and titles

The titles of tables and graphs often become very long and complicated. But for most presentations, the titles can be simplified.

Literacy rate. Adult men and women

	Men	Women
1935	30.8	8.0
1945	44.3	13.5
1950	47.7	16.7
1955	56.3	21.3
1960	54.8	21.1
1965	64.7	27.6
1970	71.0	38.2
etc.	:	:

A very common – and "statistical" – way of formulating a
title of a table or graph is to specify the classification variables by using "by …":

Unemployed by sector and gender.
But most readers would more easily understand the title:
Unemployed men and women in private and public sector.

In the same way:
Instead of: *Percentage distribution of households by type of household*,
write: *Household types. Per cent.*

27

Instead of: *Movimento Geral de Crimes Segundo Tipos de Crime*,
write: *Tipos de Crime*.

Not: *Pupil to teacher ratio*,
but: *Pupils per teacher*.

Percentages in tables: Two ways of comparing

Percentages are among the statistician's best friends. Because absolute numbers are often difficult – or impossible – to compare, we therefore compute percentages.

Let us start with a simple frequency distribution, showing the number of teachers working in primary and secondary schools. This is a so-called one-way table.

Number of teachers in primary and secondary schools

Primary schools	65 376
Secondary schools	26 618
Total	**91 994**

This table gives the *absolute* number of teachers. We easily see that the majority of teachers work in primary schools, but we get no precise impression of the relative size of the two groups and it would be difficult to compare the distribution with earlier years or other countries.

To compare the results more easily, we calculate percentages, thereby getting a *relative* distribution of the teachers. We can now give a much more precise description of the size of the two groups: We see without difficulty that seven out of ten teachers work in primary schools.

Teachers in primary and secondary schools. Per cent

Primary schools	71.1
Secondary schools	28.9
Total	**100.0**

Since we assume that women are more likely to work in primary schools than men, we construct a new table, introducing gender as a second variable. This cross tabulation gives us the following table. It is not important which of the two variables is put in the rows and which is put in the columns.

Number of teachers in primary and secondary schools. Men and women

	Men	Women	Total
Primary schools	19 334	46 042	65 376
Secondary schools	14 491	12 127	26 618
Total	**33 825**	**58 169**	**91 994**

This table is often called a two-way table, since it shows the combination of two frequency distributions and this is a very common way of presenting statistics. Here too, the absolute numbers give a rough impression of where men and women work. Men more often work in secondary schools. But by calculating percentages, it will be much easier to describe the gender differences. But the question is, which way to calculate the percentages?

As a general rule, the percentages should be calculated on basis of the so-called independent or classification variable – in this case on the basis of gender, i.e. men and women. We therefore calculate the percentage of men and women that work in primary and secondary schools, in the same way that

Teachers in primary and secondary schools. Men and women. Per cent

	Men	Women	Total	
Primary schools	57.2	79.2	71.1	Calculate
Secondary schools	42.8	20.8	28.9	percentage
Total	**100.0**	**100.0**	**100.0**	

Compare

we calculated the total relative distribution. (This is what is often called a gender-specific distribution). We then get the following table.

We can now conclude that while 79 per cent of the women work in primary schools, this applies to only 57 per cent of the men.

Because the dependent variable (type of school) here only has two values – primary and secondary – we do not lose any information if we drop one of the values. We also leave out the "totals" and get the following, simple table. While the original table contained nine numbers (or cells), this has now been reduced to two – and the table has become easier to read and understand.

Male and female teachers working in primary schools. Per cent

Men	57.2
Women	79.2

What happens if we compute the percentages the other way – on the basis of the variable "type of school"? We then get the following table (the gender distribution), which can be read like this: Of those working in primary schools, women make up 70 per cent. Since their share of the total number of employees is only 43 per cent, women are clearly overrepresented in primary schools. We often refer to this as the proportion of females among teachers; totally, in primary and secondary schools. This way of calculating percentages is often used when analysing differences between men and women. This is due to two things: 1) We usually know the distribution of men and women (ca 50:50). 2) In many cases, it is also a political goal that the proportion of men and women should be almost equal.

Teachers in primary and secondary schools. Men and women. Per cent

	Men	Women	Total
Primary schools	29.6	70.4	100.0
Secondary schools	54.4	45.6	100.0
Total	**36.8**	**63.2**	**100.0**

We often add a third variable in our analysis. For instance, we might want to see if the proportion of female teachers at different levels is different in urban and rural areas. This could easily be done by constructing separate tables for urban and rural areas. But remember that many (if not most) users have problems understanding tables with three or more variables.

A third way?

A third way of calculating percentages is to use the grand total as a basis for the calculation, as in the table below. The interpretation here is that of the total, adult population, females using a computer within the last three months amounts to 5.77 per cent, compared to 11.88 per cent males, etc. But as this way of calculating percentages is very unusual, most people will have trouble understanding a table like this.

25.10 Cinsiyete göre bilgisayar ve İnternet kullanım oranları, 2005
Proportion of computer and Internet use by gender, 2005 (%)

	Bilgisayar / Computer			İnternet / Internet		
	Toplam Total	Kadın Females	Erkek Males	Toplam Total	Kadın Females	Erkek Males
Son üç ay içerisinde / Within the last three months	8 503 659	5.77	11.88	6 712 495	4.33	9.60
Üç ay ile bir yıl arasında / Between 3 months and a year ago	903 812	0.71	1.17	733 525	0.54	0.99
Bir yıldan çok oldu / More than one year	1 646 859	1.53	1.89	1 011 389	0.74	1.36
Hiç kullanmadım - Never used	37 124 331	42.28	34.78	39 721 252	44.68	37.76

From: Turkey's Statistical Yearbook, 2005

Here, it would be better to calculate gender-specific distributions, showing that 11.4 per cent of all females used a computer during the last three months, compared to 23.9 per cent of males. Dropping the absolute numbers and rounding to one decimal, we get a much more user-friendly and comprehensible table.

Computer and Internet use by gender, 2005. %

	Computer		Internet	
	Females	Males	Females	Males
Last three months	11.4	23.9	8.6	19.4
Three months - one year	1.3	2.4	1.0	2.2
More than one year	3.0	3.9	1.4	2.7
Never used	84.5	69.7	89.3	75.7
Total	**100.0**	**100.0**	**100.0**	**100.0**

And again, without loosing too much information, this table too could be simplified, by focusing on use during the last three months.

Computer and Internet use last three months, by gender, 2005. %

	Females	Males
Computer	11.4	23.9
Internet	8.6	19.4

Summing it all up

To recapitulate some of the main points about tables, let us use an example from Office National de Statistique, Algeria. On the right is a table showing the number of unemployed persons in Algeria, by sex and urban/rural dimension.

First, we notice that the numbers are centered. Second, the table contains both absolute *and* relative numbers (percentages) – but because they are mixed together, it is very difficult to compare the numbers. Third, the makers of the tables have not been able to decide which way the percentages should be computed – so they have calculated them both horizontally (en ligne) *and* vertically (en colonne). The resulting table is therefore – despite its small size – difficult to read and not very user-friendly.

**Tableau N° 09 :
Répartition des chômeurs par Sexe et Strate**

	URBAIN	RURAL	ENSEMBLE
MASCULIN	588 196	400 093	988 288
% en ligne	59,5	40,5	100
% en colonne	75,7	86,3	79,6
FÉMININ	189 170	63 383	252 553
% en ligne	74,9	25,1	100
% en colonne	24,3	13,7	20,4
TOTAL	777 366	463 475	1 240 841
% en ligne	62,6	37,4	100
% en colonne	100	100	100

From: www.ons.dz

To improve this table, we could do several things. We should:

- align all numbers to the right
- separate the absolute numbers and percentages
- decide which way to compute the percentages
- simplify the title – most people will not understand the concept "répartition" (distribution)
- remove the various background colours/shades

The "new" table could be something like the one to the right – with a clear distinction between the absolute and relative numbers (percentages). This table is simpler and is also better structured: It is easier to read – and it is easier to compare and understand the numbers.

9. Chômeurs par sexe et strate

	Urbain	Rural	Ensemble
Masculin	588 196	400 093	988 288
Féminin	189 170	63 383	252 553
Total	**777 366**	**463 475**	**1 240 841**
	Pour cent		
Masculin	75.7	86.3	79.6
Féminin	24.3	13.7	20.4
Total	**100.0**	**100.0**	**100.0**

4. From table to graph
Why and how?

In addition to tables and texts, statistical information is today increasingly being presented by using various types of graphs and maps – both in printed publications as well as on the Internet.

Why graphs?

But why use graphs, when tables usually give far more detailed information? The answer is that in large tables, the key information often disappears in numerical noise: There are simply too many numbers.

Graphs compress data (they have high data density) and they are – when properly designed – effective means of dissemination of statistical information. Graphs help us compare the numbers. Graphs – in short – give a quick, general, intuitive, visual impression of:

- distributions
- trends and developments
- differences between groups
- relationships between variables

Example

The following table shows the percentage of girls (in Norway) given the first name «*Anne*" every year since 1875 and the table contains 132 observations. From the table it is almost impossible to get any impression of the "rise and fall" of "*Anne*", even if you study it closely.

Year:	%	Year:	%	Year:	%	Year:	%	Year:	%	Year:	%
1875	3.1	1897	1.4	1919	1.1	1941	3.5	1963	5.0	1985	1.4
1876	2.7	1898	1.1	1920	1.2	1942	3.5	1964	4.6	1986	1.4
1877	2.5	1899	1.1	1921	1.2	1943	4.1	1965	4.7	1987	1.2
1878	2.2	1900	1.1	1922	1.3	1944	4.1	1966	4.4	1988	1.1
1879	2.5	1901	1.0	1923	1.3	1945	4.2	1967	4.2	1989	1.0
1880	2.1	1902	1.1	1924	1.4	1946	4.3	1968	3.8	1990	1.1
1881	2.1	1903	1.1	1925	1.4	1947	4.1	1969	3.5	1991	1.0
1882	1.8	1904	1.0	1926	1.4	1948	4.4	1970	3.1	1992	0.9
1883	2.2	1905	0.9	1927	1.7	1949	4.5	1971	2.8	1993	0.9
1884	2.0	1906	0.9	1928	1.6	1950	4.7	1972	2.5	1994	0.8
1885	2.1	1907	1.0	1929	1.6	1951	4.7	1973	2.5	1995	0.8
1886	1.8	1908	0.9	1930	1.8	1952	5.0	1974	2.3	1996	0.8
1887	1.7	1909	1.0	1931	1.9	1953	5.0	1975	2.1	1997	0.7
1888	1.8	1910	1.0	1932	2.0	1954	4.9	1976	2.0	1998	0.6
1889	1.7	1911	1.0	1933	2.2	1955	5.1	1977	1.9	1999	0.5
1890	1.9	1912	0.9	1934	2.1	1956	4.9	1978	2.0	2000	0.5
1891	1.5	1913	1.0	1935	2.4	1957	5.3	1979	1.9	2001	0.5
1892	1.4	1914	1.0	1936	2.5	1958	5.4	1980	1.8	2002	0.5
1893	1.4	1915	1.1	1937	2.9	1959	5.3	1981	2.1	2003	0.4
1894	1.4	1916	1.0	1938	2.8	1960	5.2	1982	1.7	2004	0.4
1895	1.4	1917	1.1	1939	2.9	1961	5.0	1983	1.7	2005	0.3
1896	1.4	1918	1.1	1940	3.4	1962	5.0	1984	1.5	2006	0.4

From: Statistics Norway

In a graph (below) – on the other hand – the trend stands out very clearly. This is also the case even if you add a second data series, increasing the number of data points (observations) to 264. Comparing the two tables, however, would be almost impossible.
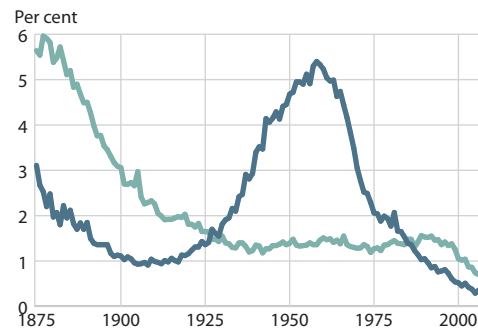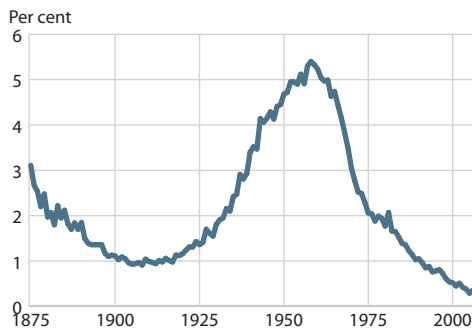
- Graphs are pedagogical, because they are easier to remember
- Graphs are "eye catchers"
- Graphs encourage comparison of trends and analysis of differences and relationships
- Graphs speak directly to the eye and do not require any detailed understanding of statistics or percentages!

> *"It is often the case that a picture can be worth a thousand words, and equally worth a thousand numbers presented in tabulations! The visual representation of something is far easier to grasp, and is less intimidating, than a sheet of figures for policy makers and the public at large"*
>
> Tim Holt, former Director of the Office for National Statistics, UK

On the other hand, statistical publications and websites are full of meaningless graphs. If not properly designed, graphs can be misleading and confusing. It is therefore very important that great care is taken when designing the graphs.

How (not) to lie with graphs

In the following graph (left below), the value axis starts at 280 – thereby giving an impression of a substantial increase in the number of hospitals. This kind of dramatization is very common, especially in the media.

In the graph to the right, the increase is much more moderate. A general rule is that when graphing only one time series, the value axis should *always* start at zero.

From: Sudan in Figures 2006
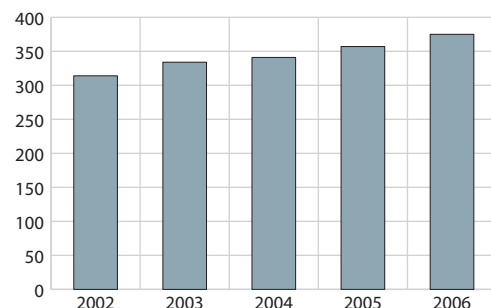
But when the purpose is to compare two or more time series (for instance life expectancy of men and women), it may sometimes be acceptable to "shorten" the value axis, as in the example from Mauritius on the right.

From: Mauritius in Figures 2007

Another example of misleading graphing is given below. Be careful when the data points in a time series are not evenly spaced; like in the following example from IBGE, Brazil: In the graph below (left), there seems to have been a sharp decrease the infant mortality rate in the 1990-ies, but that the downward trend has later levelled off. But when we adjust the distance between various years (right), we see a relatively smooth, downward trend. (Here, since we do not have data for every year, it is appropriate to use symbols to indicate the years for which we have data).

Taxas de Mortalidade Infantil - 1990-2008
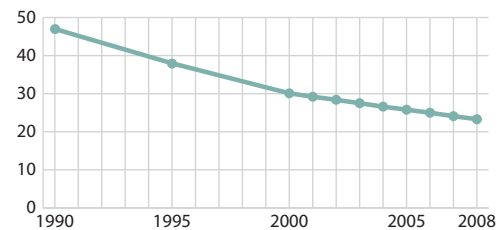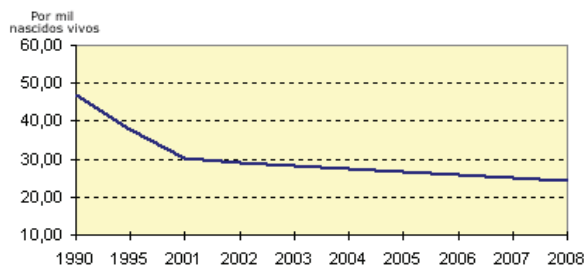
From: Instituto Brasileiro de Geografia e Estatística, www.ibge.gov.br

Elements of a graph: Issues and solutions

Size

Very often, graphs are made too big: There is no reason why a graph should take up a full page (or even half a page). For most purposes, the size used in this manual (approximately 6.0 x 4.0 cm) is big enough. (For good examples of small, but informative graphs, see *The Economist*).

Patterns

The selection of different patterns and shades will depend on the software used. The best strategy is to use colours or percentages of black, for example 1, 15, 50 and 100. With more than four values, use other patterns in addition, as it is difficult to differentiate between more than four shades of grey. But "fancy" patterns like those on the right should be avoided.

Grid lines

Grids are used to help the reader to more easily compare different values (horizontal grids) or identify specific years in line graphs (vertical grids). The question is: How many grids should you use? The number of grids is, of course, very much a matter of taste, but in our opinion the one on the bottom right is best.

Legend

By default, the legend (key) is often placed outside the plot area, usually below or on the right-hand side, like this:
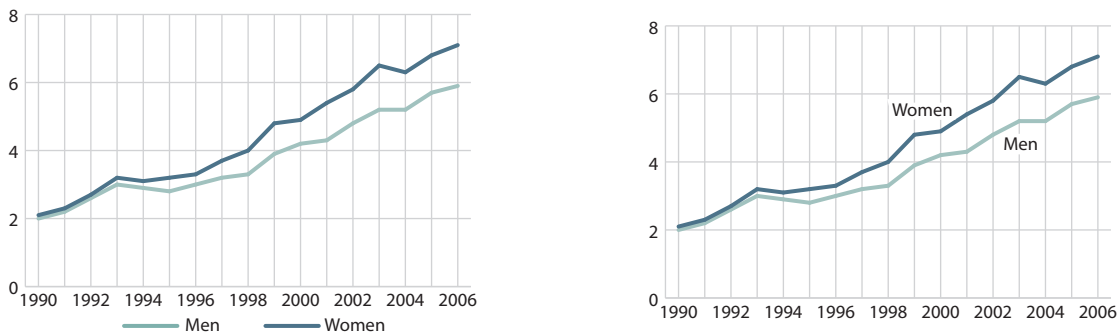
A better solution is often to bring the legend into the plot area, thereby increasing the readability. The best solution is usually to write the labels in the plot area, close to the series.

Numbers in the graph?

Should graphs also give the numbers? And how? Generally, graphs are designed to show only the main trends and differences, not to give the details. So our recommendation is *not* to report numbers in graphs. But if you do, put the number inside the bar. In a line chart it is sometimes useful to give the first and the latest value and minimum/maximum.
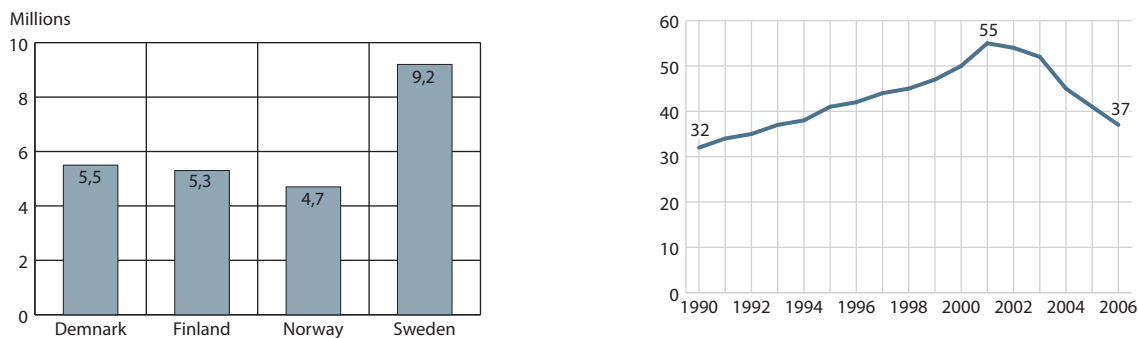
Different types of graphs and their use

The table on the right shows the number of adopted children in Norway, by age and gender for a period of eight years. Below is a selection of all the possible charts that can be made on basis of this (relatively small) table. One important conclusion from this example is always to carefully consider which chart is best suited or most functional; from every table there are usually more than one chart that can be drawn.

Number of adopted children, by age and gender

	Age			Gender	
	Under 3 yrs.	3-11 yrs.	12 yrs. +	Boys	Girls
2000	531	170	91	368	424
2001	590	182	136	379	529
2002	660	189	140	376	613
2003	594	127	149	307	563
2004	592	106	93	298	493
2005	645	103	142	320	570
2006	406	108	143	286	371
2007	377	166	160	329	374

Some charts from the above table:

Number of adopted children
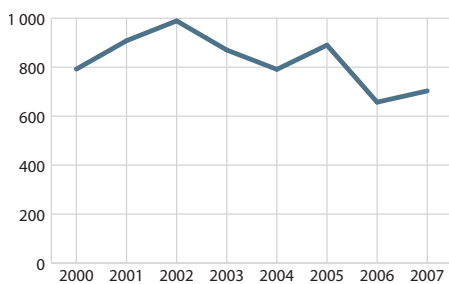
Adopted children, bye age. Per cent
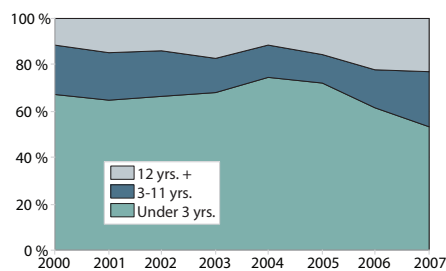
Number of adopted children
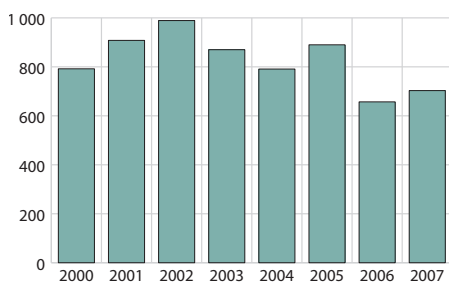
Number of adopted children. Boys and girls
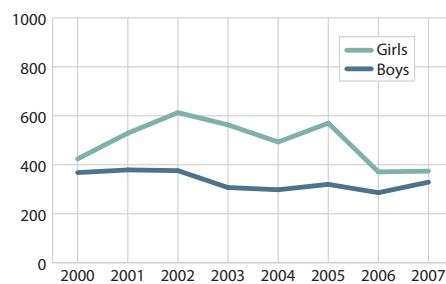
Number of adopted children, by age
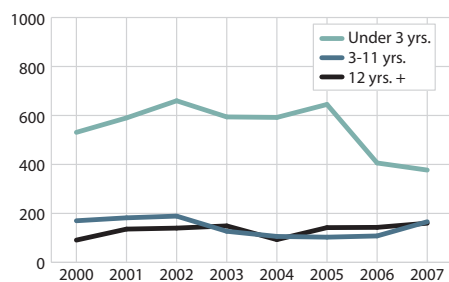
Number of adopted children. Boys and girls
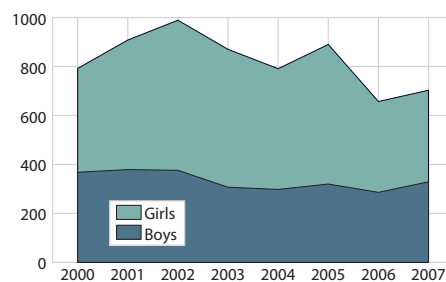
Number of adopted children, by age
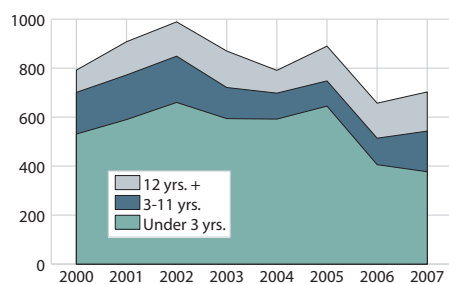
Adopted children. Percentage girls
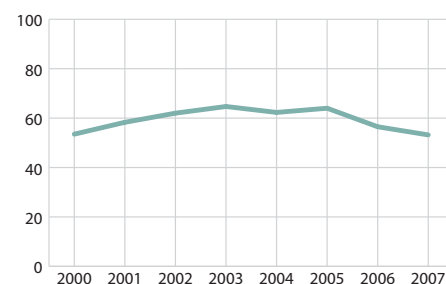
Bar charts
Also called column charts, this type is the simplest and most basic chart type. It is used to compare groups, like countries, districts, men and women, age groups, etc. With only one classification variable it is best to use only one colour/pattern.
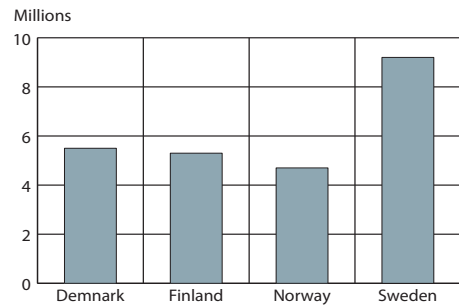
If you want to focus on one class/group, this can be done by using a different colour/pattern:
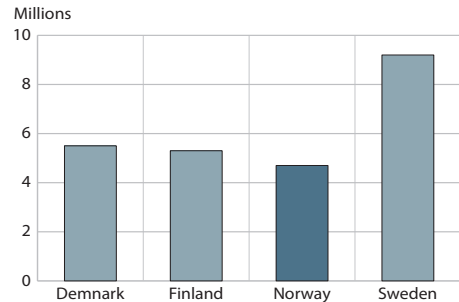
Grouped bar charts
Used when there are two (or more) categories/classifications. When grouping categories together, there is usually no space between the bars (here: between males and females):

With two variables, we have two possible ways of grouping, inviting different types of comparison: The first grouping invites a comparison of males and females in different age groups: "The difference between males and females is largest among the young".
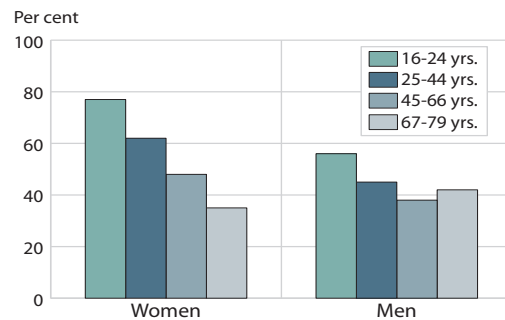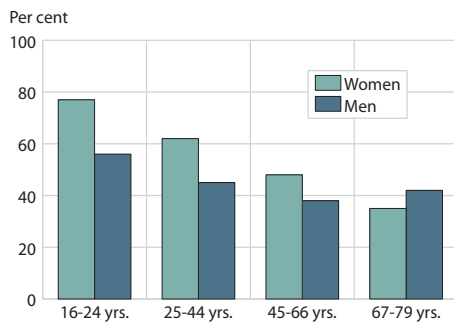
The second grouping invites a comparison of the effect of age, separately for the two sexes: "The percentage using a library decreases with age, but more so for females than males".

"3D" bar charts
So-called "3-dimensional" graphs have become very popular lately. But because of the added perspective, the real values are difficult to see: In the graph on the right, the number of births seems to be approximately 13 500 in 2006, but the number is actually 14 120.
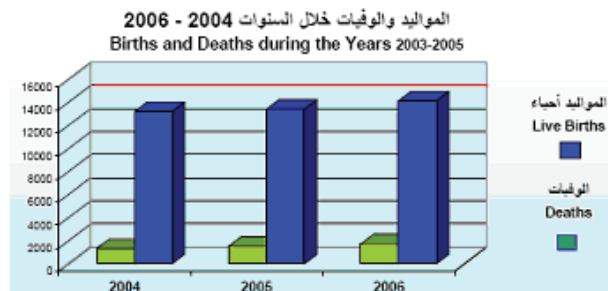
From: Qatar in Figures, 2008

Stacked bar charts
Also called "accumulated" bar charts. Show total frequency *and* how the total is divided into different components (here: daily and occasional smokers = total).

Daily and occasional smokers. Per cent

36
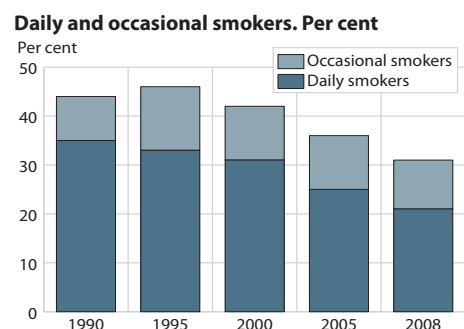
Horizontal bar charts

Often used when labels are long and/or when there are many groups. Here, the text should be justified to the right. By sorting by the value of the indicator (dependent variable), it is easier to see which group has the highest/lowest values.
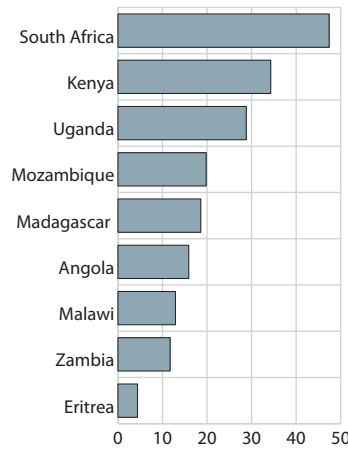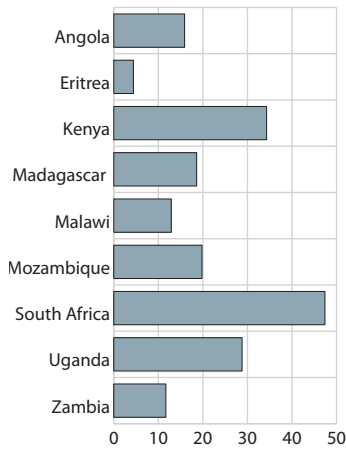
Also horizontal bar charts can be stacked. But take care not to stack or accumulate values that should not be accumulated. Male and female transition rates from primary to secondary school should not be accumulated: A transition rate of 167 per cent has no meaning.

Instead, the bars for men and women should be placed side by side, which also makes comparison easier.
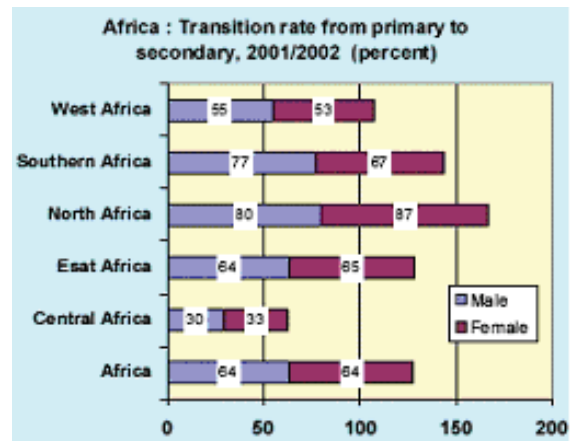
From: Gender, Poverty and Environmental Indicators on African Countries, African Development Bank Group 2006.

A special type of horizontal bar chart is the population pyramid (also called two-way bar chart). This is basically two horizontal bar charts, joined together; "back to back".

Line charts

Line charts are used for showing *time series*. Time = years, quarters, months, weeks, days, hours and minutes. Symbols (■ ∇) are often used to differentiate between series, but they often overload the chart, especially with many or crossing time series. Instead of symbols, it is better to use different colours and/or line styles.
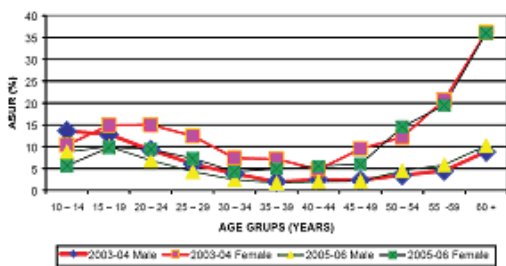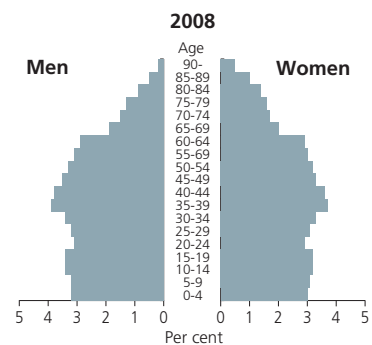
From: Pakistan Labour Force Survey 2005-2006. Statpak
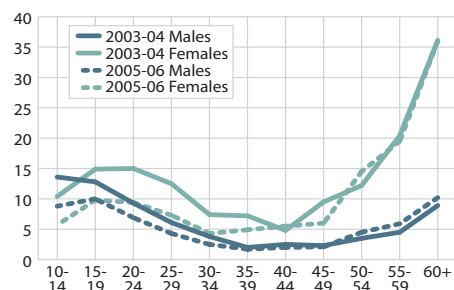
When designing line charts, take care not to include too many lines, especially when they are very close and/or cross each other. If not, the result will be a "spaghetti"-chart, like in the example to the right (from Eurostat), where it is impossible to identify and compare the various countries.

Line charts are for time series. Consequently they should *never* be used to illustrate differences between groups, for instance countries (left below). Here, a more suitable chart would be a horizontal bar chart.

2.4.2.6. Percentage of male population severely overweight, by age (BMI >= 30) (ECHP, 1998)
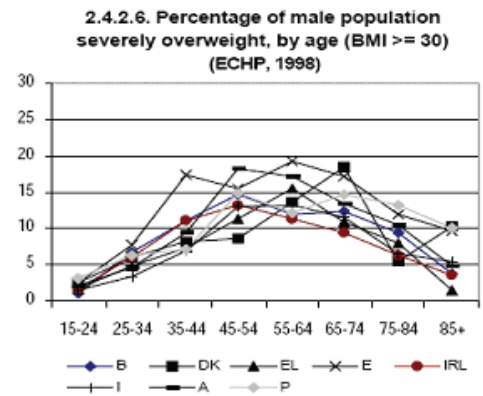
From: Eurostat. Health in Europe. 2003
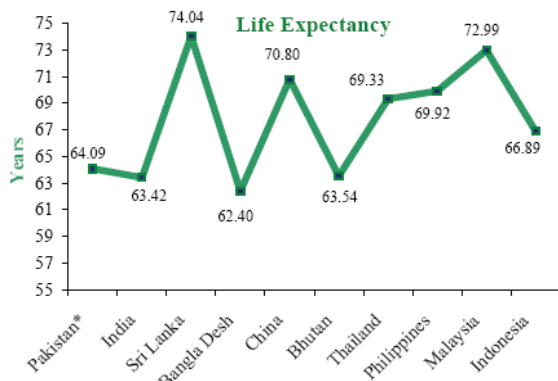
From: Pakistan Statistical Pocket Book 2006
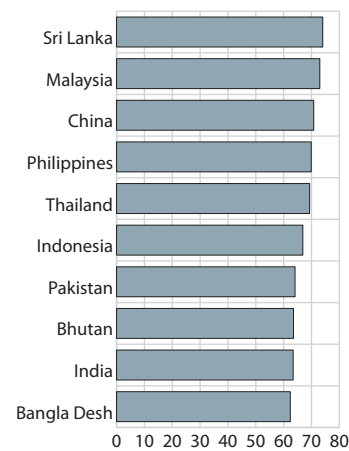
Cumulative line charts

Show the cumulative frequencies (number or per cent) for certain years. For instance, estimated divorce frequencies at different durations for various cohorts (left below) or cumulative number of deaths in road accidents by month (right).
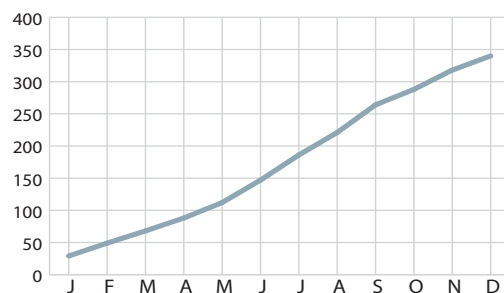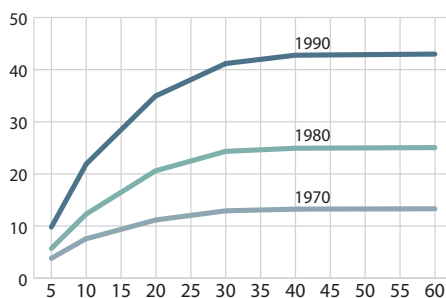
Curve smoothing

Line charts often present a picture of "ups and downs" (left below). Sometimes series like this are "smoothed", using a special function in Excel. This function should *not* be used, as it tends to distort and hide the trend.
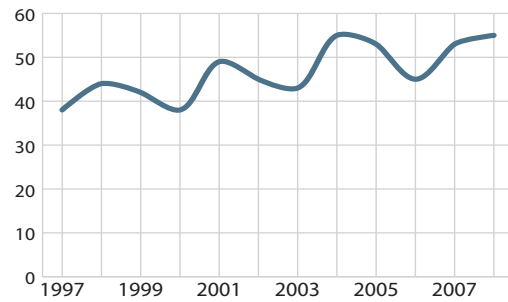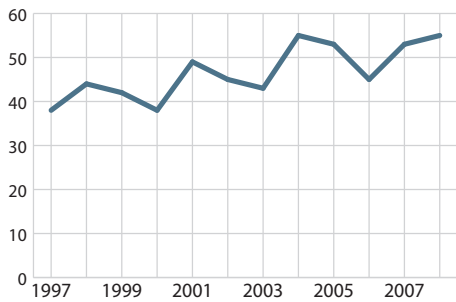
Area charts

Area charts are accumulated line charts; like stacked bars. They show both total *and* parts. Area charts should not contain too many groups/variables.
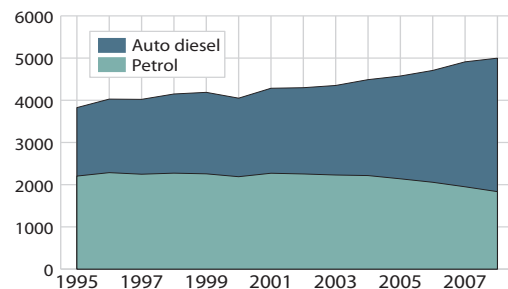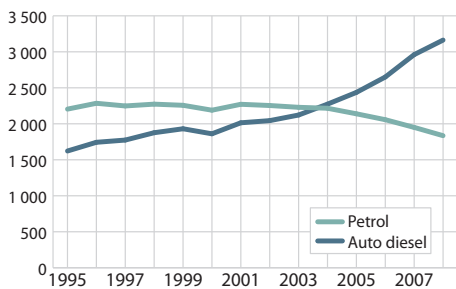
But remember: Take care not to stack or accumulate values that should not be accumulated. Deaths and births should not be stacked, the accumulated rates (left below from Trinidad and Tobago) has no meaning. Instead, the two rates should be shown as separate lines.
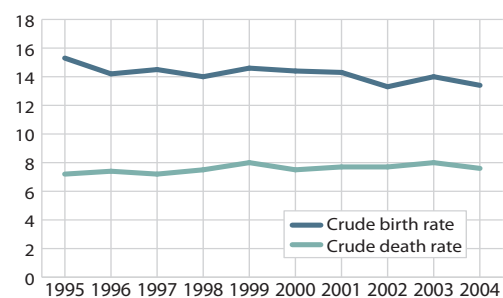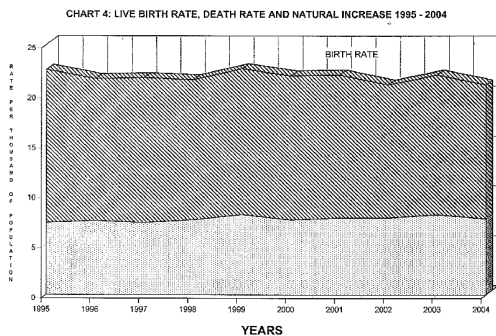
From: Annual Statistical Digest 2006, Trinidad and Tobago

Pie charts

Pie charts show the distribution of (qualitative) variables. They have no value axis, but the total area = 100%
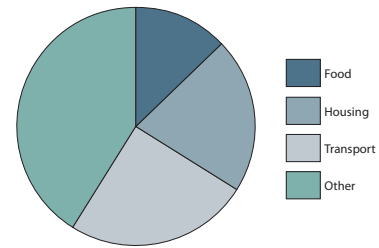
Instead of legend, use labels (easier to read). Since pie charts have no value axis, the percentages should be shown.
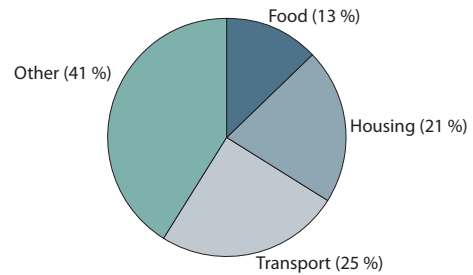
Use a maximum of 5 or 6 values/groups: With too many values, the reader loses interest.

If we want to draw attention to a specific subgroup/ sector, this can be done by "exploding" the segment and/or bolding the label.
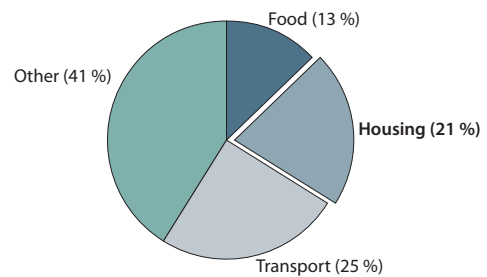
When we want to take a closer look at or focus on a certain segment or subgroup, we could add a separate bar, showing the distribution.
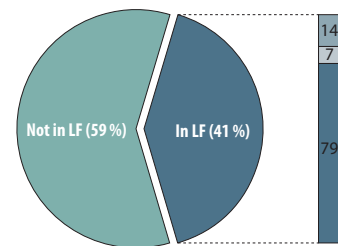
Alternatively, we could use two pie charts, where the second pie is smaller than the first (in this case approximately 40 per cent).
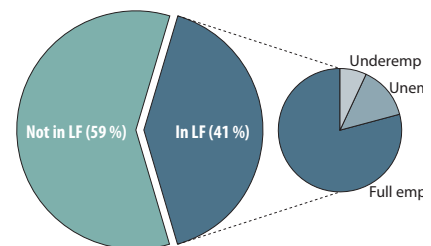
Pie charts should *never* be used to illustrate time series. The data in the chart to the right – from the Statistical Yearbook of Sudan 2006 – should instead have been shown as a bar chart or as a table.

From: The Statistical Yearbook of Sudan 2006

40
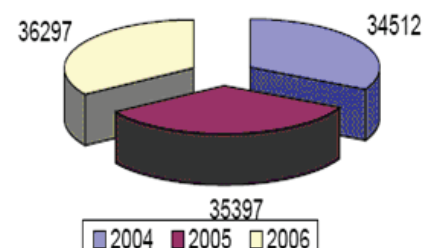
Other types

Combination graphs (also called "overlay"): Bar and
line. Here showing the proportion of males and females
in different age groups that are employed.
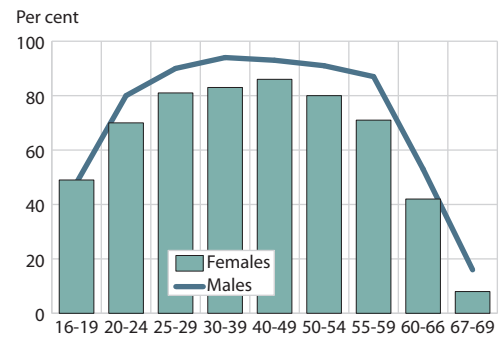
When presenting monthly data, a *spider* (or "radar")
graph (circular line graph) is sometimes a useful alternative. Here,
showing monthly births as deviations from the average (=100).
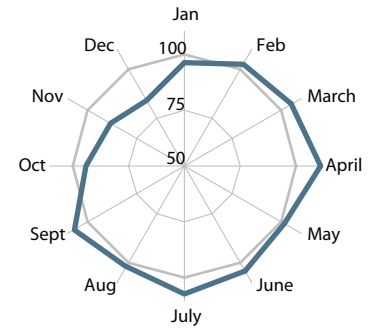
Minimum - maximum charts (also called "floating column"). Below is an example of minimum and maximum temperature. In addition to minimum and maximum, the monthly average may be added.
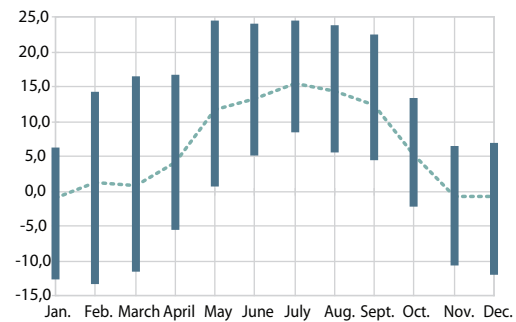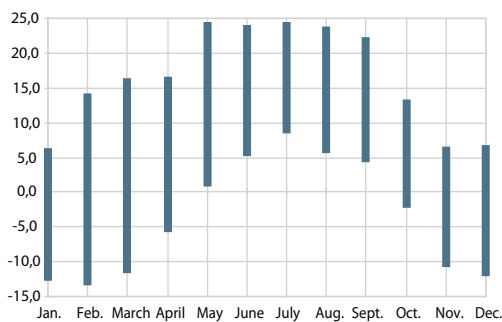
A *graph matrix* is often used when graphing three or more variables. It enables the reader to study many graphs simultaneously, comparing the effect of several variables. Below, we can easily see the trend in the use of various media in different age groups: While newspaper-reading is decreasing among the youngest (16-24 yrs.), newspapers still seem to attract the attention of middle ages and elderly people. Radio listening is decreasing in all age groups and the use of PCs is increasing at all ages, but especially among the youngest.

Use of various media on an average day, by age. Per cent. 1991-2008
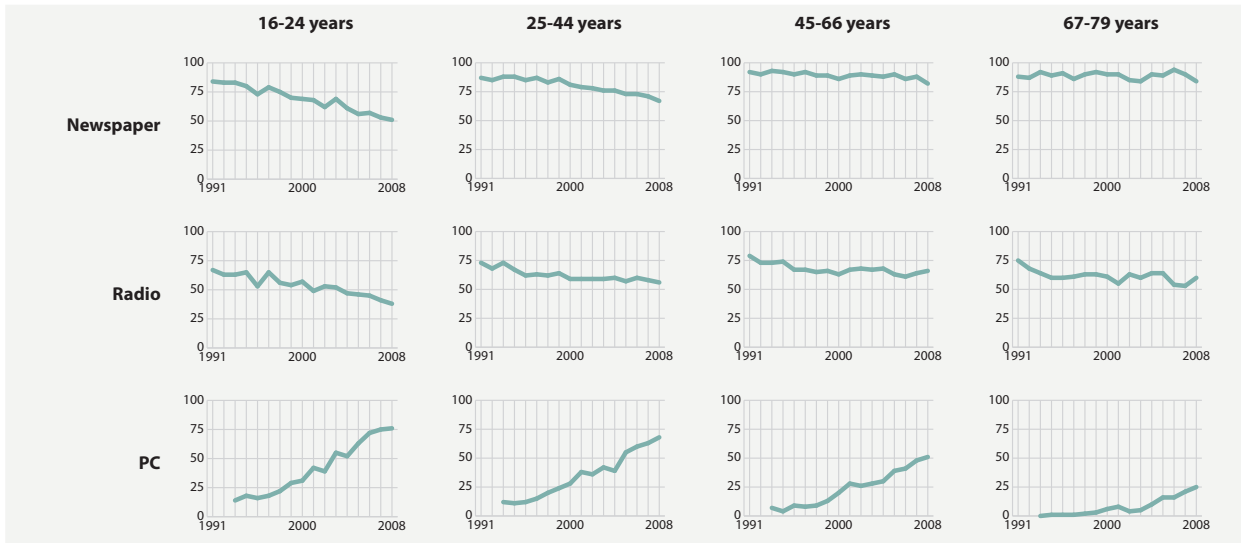
Correlation (scatter) charts show the relationship between two variables and have two value axes. They are often used when presenting regional data (municipalities, regions, countries). Here: Percentage married women using contraception *and* total fertility rate in some countries. When the number of units is not too large, the various units may be identified by labels, increasing the informative value of the chart.
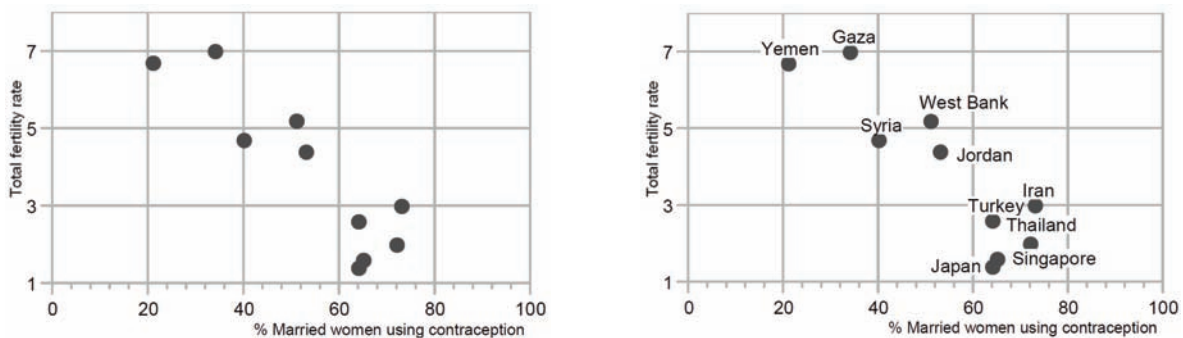
Geography and statistics: Thematic maps

Statistics about regional units like municipalities, districts, regions or countries can of course, like other statistics, be presented in tables and charts. But because these data contains a geographical or spatial dimension, it is sometimes more appropriate to present the information as thematic maps.

Unlike a reference map, which shows geographic features like roads, rivers, cities, etc., a thematic map shows the spatial variations or patterns in one (or a few) social, demographic, economic or other statistical phenomena. Thematic maps are used to display size, density, distributions or movements, by means of colours or symbols.

There are three main types of thematic maps: shaded maps (also called choropleth maps), dot density maps and maps with proportional symbols.

Shaded maps

Shaded (or coloured) maps are usually used for presenting relative numbers, like ratios, rates or percentages, rather than absolute numbers. The data are grouped into a limited number of classes (usually 4-6), with each class representing a range of data values. Each class is then given a different colour, shade or pattern. Usually, dark colours or shades represent high values, while lighter colours indicate low values.

The map to the right shows the total fertility rate in various European countries, grouped into five classes.

Shaded maps, however, tend to cover up differences between some of the regional units. Therefore, graphs are sometimes a better way to present this kind of data.

Sometimes, especially when comparing countries, information may be missing for some countries in a region. In this case, it is better to use a table or graph, indicating in the title that it only includes "selected countries".

Total fertility rate 2006

Total fertility rate 2006
- No data
- 1.20 - 1.29
- 1.30 - 1.49
- 1.50 - 1.79
- 1.80 - 1.99
- 2.00 - 2.20

Source: Unicef and Eurostat
Copyright: Norwegian Mapping Authority
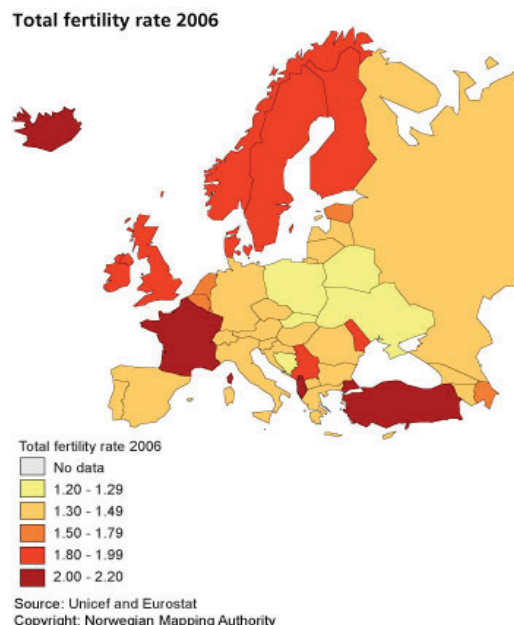
Dot density maps

This is a map where dots are used to represent the absolute number or amount or density of various phenomena. Each dot or symbol used on the map may represent a single unit (one dot = one person) or a group (one dot = 1000 people).

The map on the right shows population density in Norway per km^2, and here the dots have been given different colours to signal the different values that they symbolize. From the map we can easily see the concentration of the population in the south-east.

Inhabitants per 1 km^2 grid square
- 1 - 4
- 5 - 29
- 30 - 199
- 200 - 499
- 500 - 16 300
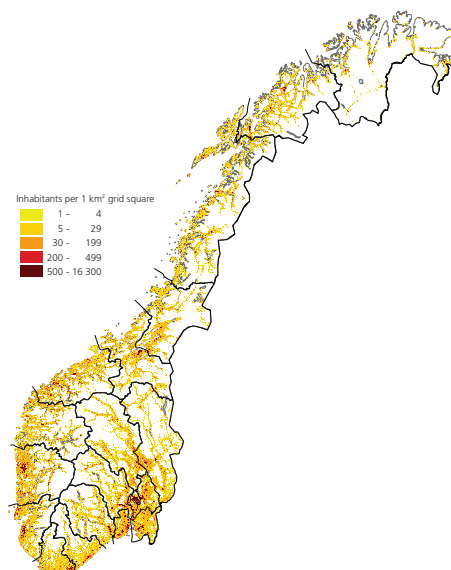
Proportional symbol maps

Proportional symbol maps are also used for showing raw data (absolute numbers), for example population size. But here, the data is displayed with proportionally sized symbols, usually circles or bars.

The map to the right shows the total immigrant population in various counties in Norway, using circles to represent the size of the population.

Number of immigrants in Norway. 2008

Smallest symbol: 4 562
Largest symbol: 139 978
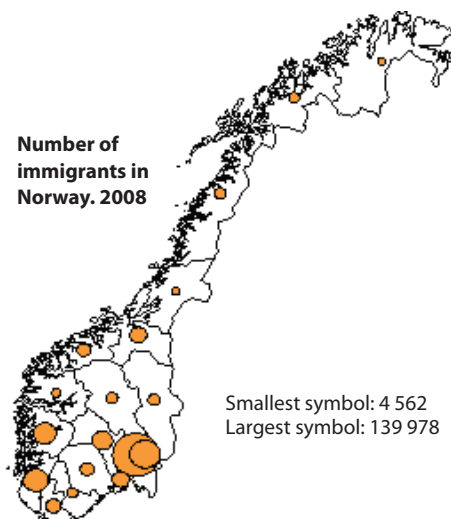
Combining maps *and* graphs

Depending on the map-making tool used, shaded maps may also show a second variable, by adding for example a pie chart or one or several bar charts. To the right is a map of Europe, showing *both* total fertility rate (shaded) *and* the proportion of children born inside/outside marriage (pie chart).

This map presents a rough picture of the regional pattern of both variables: Fertility is highest in the north and in the west, lowest in the south and in the east. The proportion of children born outside marriage shows a similar pattern.

The map, therefore, indicates a correlation between the two variables: In countries with high fertility rate, a large proportion of children are born outside marriage.

But again: Always consider if a chart would be better. In this case perhaps a correlation chart would be more suitable for the purpose.

In the chart to the right, the correlation between the two variables becomes much clearer. At the same time the country's values on the two variables are also seen more clearly. Identifying the various countries adds information to the graph.

A more comprehensive discussion of the use of maps can be found in: *Making Data Meaningful. Part 2: Style guide on the presentation of statistics* (UNECE 2009).
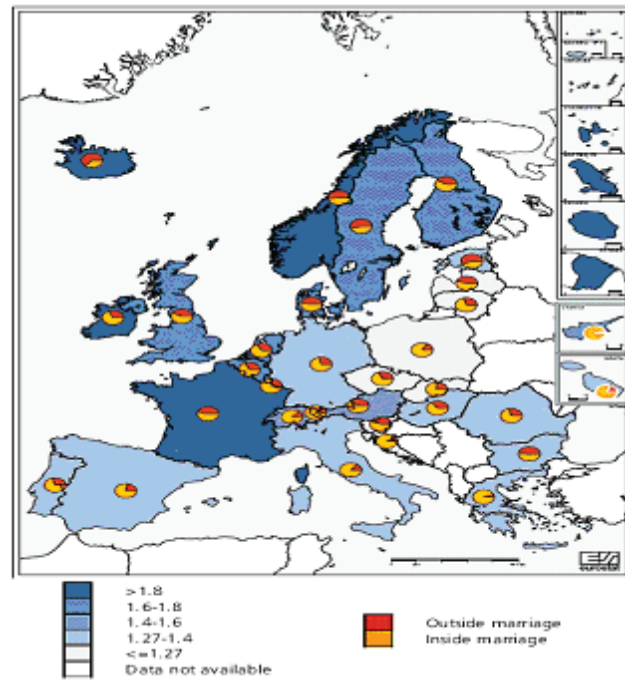
From: Eurostat
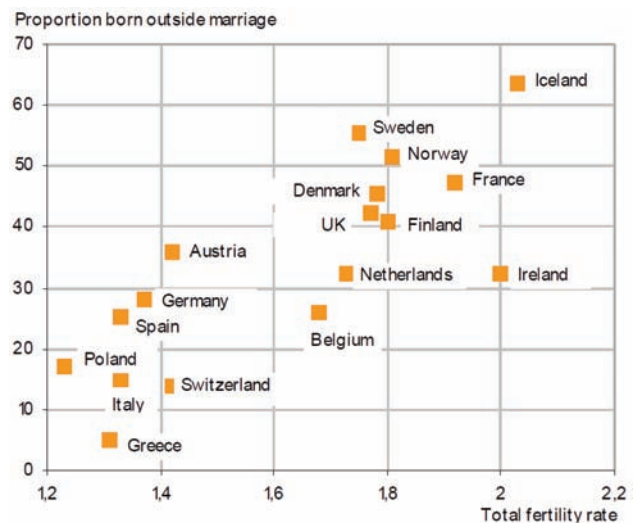
5. Writing about numbers
Adding meaning (and value) to the statistics

In an increasingly complex and sectorized society, where social, economic, demographic and cultural phenomena interact, readers need to be guided through the numerical jungle: What do the reported numbers really mean? The media, the informed users and the general public – even the experts – need and want explanations, interpretations and comments.

Analysis also forces the statistician to look closer at the concepts, definitions, measurements, sampling, etc. Analysis therefore provides a necessary feedback into the statistical production process and helps raise the quality of statistics, by uncovering errors and deficiencies in the statistics.

What is analysis?

Analysis is basically to compare numbers with other numbers and to describe this comparison in words: After having made the numbers comparable, we compare and describe the differences between men and women, different age groups, districts, etc. Or we analyze how one or several indicators are developing over time.

> *"Analysis is simply finding a story in the data and communicating that story to our readers".*
>
> From: Guidelines on writing for *The Daily* (Statistics Canada, 1995)

Constructing tables is the first step of analysis. When constructing a table, we implicitly start analysing, by asking: What is the dependent variable (indicator)? And what are the (most important) background (classification) variables (and why)? A table is always (or should always be) constructed on basis of certain ideas about *relations between variables*, which is also the basis of analysis.

Depending on the type of statistics, the various indicators can be broken down/disaggregated in a lot of ways; by sex/gender, age, urban/rural, region, education, income, etc. and by combinations of these. After having run all the basic tables, we start analysing by selecting among all the possible numbers and focusing on a few main points or aspects of the study: When analyzing we should not try to verbalize all figures in the tables, but to focus on what is important? Relevant? New?

The simplest form of analysis is to give the readers some descriptive guidance: What does this table/chart tell us? Are the differences/changes significant? What do the changes/differences mean? Are the changes part of a more general pattern? In short: Analysis is to make the statistics informative and meaningful to the reader.

Writing about numbers

There are many types of "analysis" – news releases, popular presentations, articles, thematic reports – but regardless of the type, there are some general guiding principles that should be observed.

KISS

Keep It Short and Simple: If a news release, an article or a report is too lengthy, people will be likely to drop out. "Simple" means that you should use a clear and straightforward language, avoiding statistical jargon. Also use short sentences and paragraphs.

Limit numbers in the text
Don't overload the text with too many numbers. But when numbers are relevant or necessary, they can often be rounded. Read and compare the following two sentences:

> *"From 1997 to 2007 the number of female students increased from 32,765 to 65,756, while the number of male students increased from 28,435 to 43,567".*

> *"From 1997 to 2007 the number of female students increased from 32,800 to 65,800, while the number of male students increased from 28,400 to 43,600".*

By rounding the numbers, the text becomes more user-friendly. Most people will find it much easier to read the second version. And more important: They also find it easier to compare the numbers.

Include tables and graphs
Analysis should consist of a good mixture of text, tables and graphs (+ metadata). Good tables and graphs should support and complement the text.

Bottoms up!: The inverted pyramid
Traditionally, statisticians tend to write in an "academic" way: They have a tendency to start with an introductory description of the study (or survey or census), describing the methods used, sample, sources of error, etc. before giving a presentation of the results and – finally – the conclusions.

This structure resembles that of a detective story, where the culprit is not revealed until the end of the novel. The writer deliberately holds back the climax to build suspense. This structure can be described as a pyramid – starting with the foundations (data) and finishing with the top (conclusion).

A problem with this way of writing is that only the most interested readers will stay on until the end – the rest will probably give up before reaching the conclusions.

This way of writing is very different from the journalistic style – known as the inverted pyramid –, where you start with the most important results/conclusions and go on to expand and explain the results, for instance by looking more in detail at different groups and subgroups. The idea is that if an article is too lengthy, it could easily be cut at the end, without any important information being lost.

Conclusion(s)
Results/analysis
Data/methods
Introduction

Most important
information
Important
information
Less important
information
Details

The inverted pyramid

News releases

News releases are probably the most widespread type of analysis published by NSOs. When writing news releases, there are a few important points to remember:

1. KISS: **K**eep **I**t **S**hort and **S**imple. Journalists have a deadline and they don't have time to read long and complicated articles. A news release should be between one and two pages.

2. Short title – maximum one line

3. Use substantial titles: Not "Results from the Labour Force Survey", but: "More women are working"

4. Use subheadings

5. User-friendly tables and/or graphs should support and complement the text

6. Give the date of release

7. Include (or give a link to) some basic metadata (see 9.6)

8. Include contact information: telephone/e-mail address

Example: Structure of a news release (adapted from Statistics Norway)

For other examples of short and concise news releases, see the websites of Office for National Statistics (ONS), United Kingdom (http://www.statistics.gov.uk/) and Statistics Netherlands (CBS) (http://www.cbs.nl/en-GB/). One example from ONS is shown to the right. (However, a more comprehensive presentation of this survey was also available on the ONS website).

Internet Access
70% of households had access in 2009

Per cent with Internet

Households with access to the Internet, UK
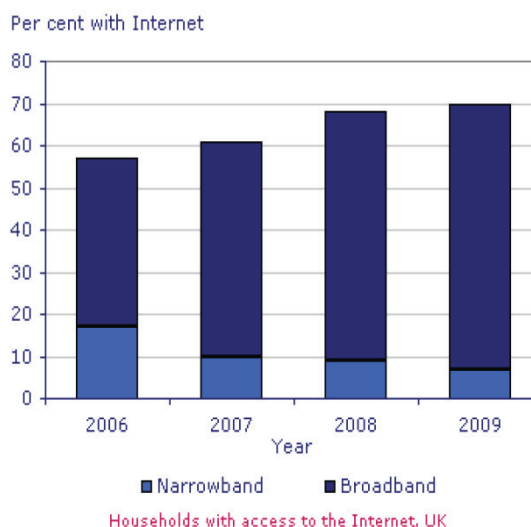
18.3 million households in the UK (70 per cent) had Internet access in 2009. This is an increase of just under 2 million households (11 per cent) over the last year and 4 million households (28 per cent) since 2006. UK estimates are not available prior to 2006.

The region with the highest level of access was London, with 80 per cent. The region with the lowest access level was Scotland, with 62 per cent.

Sixty-three per cent of all UK households had a broadband connection in 2009, up from 56 per cent in 2008. Of those households with Internet access, 90 per cent had a broadband connection in 2009, an increase from 69 per cent in 2006.

In 2009, 37.4 million adults (76 per cent of the UK adult population) accessed the Internet in the three months prior to interview. The number of adults who had never accessed the Internet fell to 10.2 million (21 per cent) in 2009.

Sixty four per cent of all adults who were recent Internet users (having accessed the Internet in the three months prior to interview) had ever purchased goods or services over the Internet in 2009. Of these, 83 per cent (26 million) had purchased within the last three months.

Sources: National Statistics Omnibus Survey

Published on 28 August 2009 at 9:30 am

6. The media
Our best friends!

The relationship between statisticians and journalists is – at times – strained or even somewhat hostile: While many journalists find statistics both incomprehensible and boring, statisticians do not always trust the journalist to be able to communicate their results in an accurate and acceptable manner. Lately, however, both statisticians and journalists have come to realize that they can benefit from working together: While statisticians have a story to be told, journalists need a story to tell. Together they are dynamite!

Why the media are important

Having a good relationship with the media and journalists is vital for every statistical agency in order to succeed. The media represents an important channel to the outside world. Without this channel, reaching out to the public at large and fulfilling our role as a provider of objective and relevant statistical information would become increasingly difficult – if not impossible.

In addition, media are important in increasing the visibility and improving the reputation of the NSO and thereby creating confidence and trust in the statistics, which in turn will increase the public support for the national statistical organisations.

Traditional dissemination model:
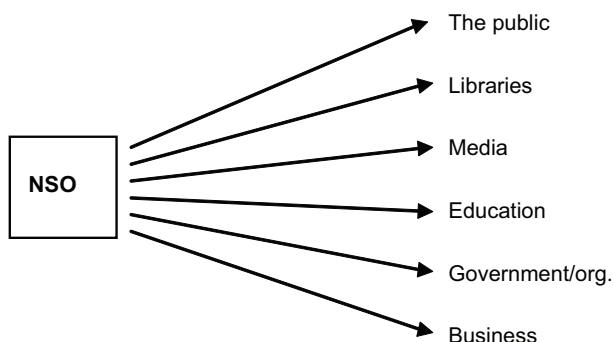
Traditionally, the media have been looked upon as just one among the many different types of users.

But experience has taught us that the media play a very important role in dissemination and is a vital and effective channel of communication for the promotion and use of statistics among all the users, not only the general public.
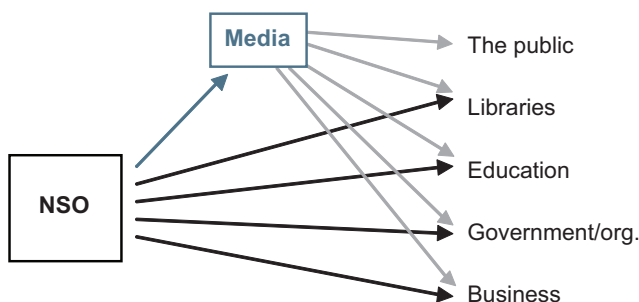
When statistics are reported or used by the media, it will also more easily come to the attention of various other users in government and the business sector. Every statistical office should therefore have an explicit policy for communication with/through the media.

"New" dissemination model:

Communicating with/through the media

When communicating with the media, there are two basic rules that should be followed:

Equal treatment

The first rule when communicating with the media is equal treatment; that all media should have access to releases of new statistics *at the same time*. In this way there will be no suspicion of favouring certain media. It is, of course, possible to pre-release the statistics to (certain) media under embargo. But this is often risky, since such rules are not always respected.

In the same way, pre-release access to government authorities should also be avoided, as this may undermine the public confidence in statistics by creating an impression that the releases are approved or censored by the authorities. Government and the media (as well as the public) should generally be given access to the releases at the same time. Sometimes, however, it may be necessary to give government authorities pre-release access. If this is the case, these exceptions to the general rule should be described and published by the statistical organisation.

The principle of equal treatment should also apply to the media. New statistics or analyses should not be given exclusively to certain media. Such a system of exclusivity may easily raise suspicion both among media and the public about our independence. The public may question our motives for using this practice and ask whether we are compromising our independence by developing close ties to selected media.

Release calendar

Tell the media about the coming releases in advance by establishing a *release calendar*. In this way, journalists have some time for planning/research. The earlier the release is placed in the calendar, the better. In some countries, the schedule is for four months, but even a shorter period – like one or two weeks – will be helpful. Announcing coming statistics in advance also confirms the NSO's independence of political authorities and other stakeholders, which in turn increases the trust and confidence in the NSO.

Example (extract) of release calendar (from Statistics Norway)

If possible, statistics should be released at a certain hour every time, e.g. at 10:00 am. In this way journalists will know when to look for news on your website and will have plenty of time before deadline.

Talking to the media

The media are not passive consumers of our statistics. When using statistics and analyses, they will often contact the NSO to get some background information, comments or explanations.

Enquiries from the media should always be given high priority. But who should answer these enquiries? In general, the persons who are most familiar with the statistics should be the spokesperson. If the enquiry is more complicated or controversial, it should be sent further up the line to be handled at the senior level. Where necessary, the enquiry shall be forwarded to the director general.

Press conferences

Press conferences should be reserved for major releases of statistics (for example results from a census or a key survey) or other important events. Journalists are very busy and will usually not attend press conferences unless they are very important.

A more comprehensive discussion of relations with the media can be found in *Communicating with the Media: A Guide for Statistical Organizations* (UNECE 2004).

7. The Internet
Something for everybody

While printed publications are often aimed at specific user groups, dissemination via the Internet is different: the website must have something for everybody. The Internet is used by both the general public, surfing and searching for information, as well as by the expert user, looking for his/her detailed CPI figures or foreign trade statistics. So, while these user groups probably use different printed publications, they will all use the same website. This, of course, presents web editors with a challenging task; to structure and present information to a very heterogeneous user group.

The Internet is a medium offering several advantages as a tool of dissemination:

- It's fast: This means that it is very well suited for dissemination of the latest statistics (news releases)
- It attracts people who otherwise do not use statistics: The first time they may be surfers ("tourists"), but if they find something of interest, they may become regular users. This means that even "serious" institutions like statistical offices should offer something of interest to this group
- The Internet offers the possibility of disseminating loads of tables/files of data in a user-friendly and inexpensive way

In many countries, like the Scandinavian countries, the Internet has become the most important channel for disseminating statistics. And in many other countries, the Internet will in the next few years increasingly dominate the dissemination process.

The development of a user-friendly website has many different aspects. We will not go into the technical details here, but concentrate on some of the content and dissemination aspects, especially from the users' point of view. The growing importance of the Internet as a tool for dissemination makes it imperative not to treat it solely as a technical instrument in line with other IT equipment. It is necessary to involve staff with both subject matter and dissemination skills as well as technical IT staff when planning the dissemination of information on the Internet.

Give priority to statistics

Many websites are dominated by information about the organization and its activities: Words of welcome, statements about visions and missions, organizational charts, statistics acts, master plans, annual plans, contact information and the like.

But most users come to the website of a NSO in search of statistical information. Therefore, the home page should primarily focus on the presentation of statistics, news releases and publications, while most of the organizational matters can be moved to a separate page ("About us"). One example: When you visit the Central Administration for Statistics, Lebanon, their

From: Central Administration for Statistics, Lebanon

(English) homepage is dominated by information about the General Director and a presentation of the statistical act. But for most users, the statistical act is *not* of great importance or interest, even if it is very important to the NSO and its role in society. This kind of information should therefore be available in the background, for instance under a link called "About us".

Another example: On the home-page of The Ministry of Statistics and Analysis of the Republic of Belarus, press releases and information from the Press Service of the President of the Republic of Belarus have a very prominent place. But this is not what users are looking for at the homepage of a NSO and such a practice will definitely not increase the users' confidence in the ability of an organization to produce independent and trustworthy statistics.

So content should be king: We should give priority to presentation of statistics:

- *News releases* – presenting new statistics – should be given a prominent place on the website. Tables presenting the results should be accessible from the news release.

- *Tables* will – for most NSOs – represent the bulk of the output. It is therefore important that the tables are formatted and presented in a user-friendly way. For instance, try to limit or eliminate extensive scrolling – especially horizontal scrolling. To view the following tables from Agence Nationale de la Statistique et de la Demographie, Senegal, you have to scroll both vertically *and* horizontally. As scrolling is particularly problematic when reading tables, you should try to avoid it. Or: Try to "freeze" the first column and the top row(s) (with the labels). When presenting text, some vertical scrolling may be both necessary and acceptable, but never horizontal.

From: ANSD, Senegal

Another important aspect is to format tables in a way that enables users to download them (for instance in Excel format), as many users want to use the figures for their own calculations. Examples of such downloading mechanisms can be found for example on the website of Statistics Norway and Institut National de la Statistique, Tunisia. Also, graphs and maps should have a format that makes it possible for the users to copy them for their own use.

- Pre-defined and pre-formatted tables are today gradually being replaced by on-line *statistical databases*, where the users may access, select, retrieve, format and download their own tables. The databases often offer the possibility of visualization of data and also include some metadata management. The development and daily operation of such a database is, however, a resource-demanding activity. For a start, many NSOs would probably want to make the figures available via downloading features, like the one described above.

- Publications are also an important part of a website. The most common format for presenting publications is the PDF-format, allowing the user both to read the publication on screen and to print it. To avoid very large PDF-files, which take long to download, it is advisable to divide the publications into chapters.

   Publications should be accessible from the main page, and presented both chronologically and by topic.

- *Metadata* should also be available on the Internet. Chapter 9 presents more detailed guidelines concerning the content of the metadata. On the Internet, metadata should be available "in the background", with links from the various statistics.

Keep it simple!

Another element that will contribute to sending users away, is asking them to register or log on with a password.

Example: When you go to "Population & Estimates" on the homepage of the Central Agency for Mobilization and Statistics, Egypt – you get the message "Log in required", asking you to subscribe to this service.

This requirement is sometimes used to get some information about the users of the website. But even when the service is free of charge, it is *not* user-friendly and only very highly motivated users will accept this time-consuming registration procedure, the rest will go away.

Also: Avoid unnecessary features like fancy and flashing "intros" with animated illustrations and/or text (some even have music), flying flags, etc. Users are looking for statistics, not Flash-applications. By keeping it simple, the web-page will load faster. Our business is communication, not animation.

Try also to avoid designing links or points of entry that look like advertisements, as most users tend to ignore them because they think they are commercial products.

Update as often as possible

If an Internet site has not been updated for several months, the users will probably stop visiting it. Especially if some item is given a label saying "NEW!" or something similar. After some time, this will no longer be news, and the label will be more misleading than informative. Instead, everything that is published on the website should be dated, so that the users can see whether this is old or new.

But in many NSOs in developing and transition countries, new statistics are still few and far between. A possible strategy to encounter this lack of "news" is to portion out statistics in several stages. For example by publishing a news release with some main results from a census or large survey as soon as possible and then, at a later stage, present more detailed results, for example on gender, regional aspect, etc.
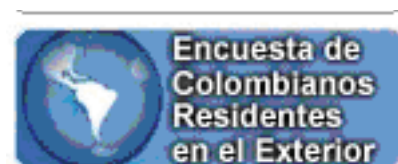
This second-stage presentation could be a news release, an article on the web or a (tabular or analytical) report. In some cases, such a procedure could also contribute to an increasing timeliness of the results, by allowing the major findings to be published earlier than would otherwise be possible.

Design and layout

Content is king. But design and layout play an important role in attracting and guiding the users, helping them to find what they are looking for. When designing a website, there are some basic rules or conventions that should be followed:

- Menus should be placed horizontally on the top and/or vertically on the left. Today, the "trend" seems to be horizontal menus. The menus should be text-based. Graphical menus tend to confuse the users

- The most important content should be placed to the left at the top of the page – this is where users will start to look for information. The less important content should be placed to the right at the bottom of the page. (Webpages in Arabic often have the most important content to the right)

- All elements on a page should be "on line", both vertically and horizontally (use "grids")

- For all text, use the same font: preferably a standard sans-serif font, like Arial, Helvetica, Tahoma or Verdana
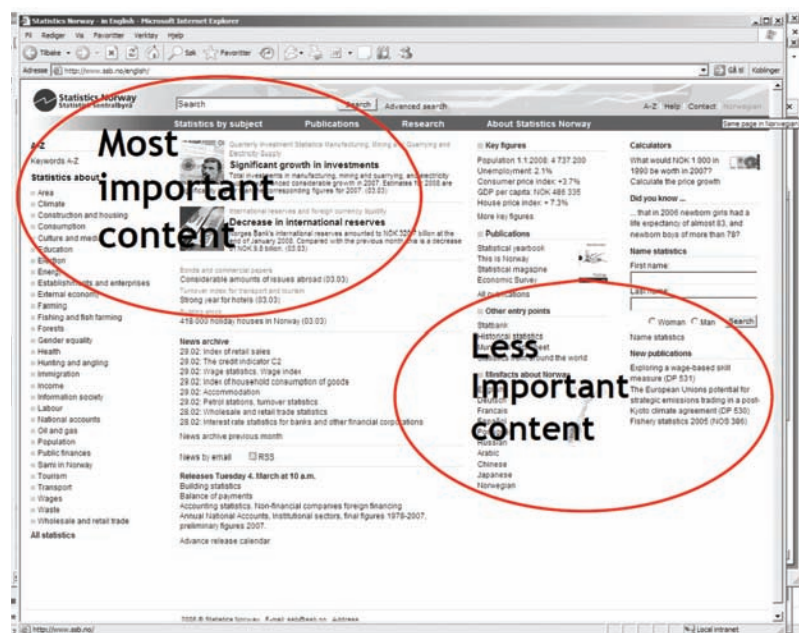
- Use only a limited number of font sizes, for instance:
  - **12 point bold for titles**
  - **10 point bold for subtitles**
  - 10 point normal for main text

- Use plenty of white space between the different elements of a page

Navigation

Hyperlinks are the most common means of navigation between web-pages. Therefore …

- Make it clear what is clickable: All hyperlinks should be clearly marked (underlined or in colour –usually blue – or both). A link should change colour after clicking on it

- Check hyperlinks regularly: Non-functioning hyperlinks must be fixed or removed as soon as possible

Use terms and concepts that are familiar also to non-expert users. An example of statistical jargon or slang is when statistics in many French speaking countries are classified as "Données conjoncturelles" and "Données structurelles" (see: CNSEE, Republique de Congo, INSD, Burkina Faso and INSAE, Benin). Even some experts will probably have problems with this classification.

Copyright?

The copyright symbol (©) is frequently found on web-pages of NSOs (and in their publications), but its meaning or implication is often unclear. On some of its web-pages, the Department of Statistics, Jordan writes: "*Copyright 2006 by Department of Statistics, Amman (Jordan). All rights reserved*". In a similar way, the Statistics Authority of Qatar states: "*All rights reserved © Statistics Authority 2008*". In "Albania in Figures 2008", the wording is: "*No part of this publication can be reproduced or transmitted in any form or by any means without the prior written permission of the copyright holder. Copyright © INSTAT 2007*".

What such copyright statements really mean is unclear. But they will definitely not encourage the readers to use the statistics in their work. If anything, such a restrictive copyright statement will probably make the user uncertain and doubtful about his/her rights.

Since we want encourage the the users to utilize our statistics, the website (and publications) should perhaps have a more inviting and user-friendly copyright statement, saying something like: "When using our statistics, please refer to Statistics Xxxxxxxx (name of NSO) as the source".

Lost in translation?

Statistics is an international language. But in order to reach the international community, most web-sites will have to translate statistics and information into a second language – which is very often English. But it seems that in many NSOs the process of translation is not taken seriously, in the sense that not enough care is taken to make sure that the translations are correct and understandable.

On many websites some of the English text is confusing, sometimes even misleading: Mistakes, ambiguities, missing information, etc. could easily be damaging to the NSO's image of reliability and high quality statistics. This problem probably reflects that statistics is a difficult subject and indicates a need for professional help when translating into English. Also, it is always a good idea to double-check the work of translators. Using the English spelling control in *Word* might also be of some help.

Since most NSOs consider dissemination to English-speaking users to be very important, it is recommended that much more attention is paid to control the quality of the final products. This applies to all texts and tables on the website, as well as in printed publications.

8. In the shadow of the Internet
Printed publications

Despite today's increasing use of the Internet as the main dissemination channel, printed publications will for many years to come continue to be an important (the most important?) means for distribution of statistics, especially in developing and transition countries.

Publication series?

Printed publications often differ in their structure and content – and it may be functional to group them into different series. Series are important, because they help to structure the (printed) dissemination of statistics and analyses. Series create expectations with the users/readers, because they signal something about the content and type of publication. At the same time, series are helpful to the writer/producer in defining user groups and thus they help him/her to direct the dissemination towards this specific group(s).

The number of different series will of course depend on the total dissemination output of the NSO, but in most NSOs, the main series would be something like:

- Tabular publications
- Reports
- Yearbooks
- Periodicals
- Popular presentations
- Internal series

Tabular publications

These have previously been characterized as "reference" publications and their main function is to serve as a historical archive, publishing relatively comprehensive results from the various surveys, studies or censuses. These publications are typically published annually or whenever the survey is done. They should contain the main results of the survey or study in a tabular form. In many countries (for instance the Nordic countries) these kinds of series are called "Official statistics of ….", or something similar.

In addition to a comprehensive presentation of the results, this kind of publication should also present a broad range of metadata in a systematic way (see chapter 9).

Reports

Statistical reports usually contain a good mixture of text, tables and graphs. Reports contain analysis and comments – but there are many types of reporting:

1) Some reports present and analyze the results of a specific survey or study – like "*Labour force survey 2006*", "*Health survey 2007*" or "*Household expenditure survey 2008*". These are mainly descriptive reports and give a comprehensive presentation of the results.

As a rule, typical reports from a survey have the following structure: After an introduction and/or summary, most reports first give a very comprehensive description of the methods used (sampling, response

rate, sampling errors, uncertainty, definitions and classifications used, etc.) before presenting the results from the survey.

But as most users are much more interested in learning about the results than about the various methodological aspects, the order should be reversed and the results should be presented first. Readers, who want to have a closer look at the methods and calculations behind the figures or to learn more about the different aspects of various indicators, should be referred to an appendix at the end of the report or to other publications.

2) Other reports may give more thematic presentations, i.e. "*Women and Men in ......", "Social Trends"* or "*Poverty*", bringing together statistics from different fields and sources. These kinds of publications are often termed "social reports", and their purpose is to provide a broad description or overview of various social indicators, to monitor their development over time and to compare different groups in the population in key areas such as poverty, health, education, crime and housing.

Social reports are primarily intended for the informed users and the general public and should contribute to a more informed public debate. Consequently it should be written in a straightforward and plain language, relying on simple tables, graphs and maps for presenting the results.

Yearbooks

For the users, a yearbook serves two basic functions: 1) It presents a concise, but comprehensive summary of the most relevant, interesting and important statistics. 2) Secondly, a yearbook represents an introduction to other national (and sometimes international) statistics, by guiding the reader on to other, more detailed, statistics.

The special role of yearbooks

Therefore, yearbooks are an important channel of dissemination in most NSOs, especially in developing and transition countries. But in addition to giving a summary of – and an introduction to – national statistics, a yearbook also has a central secondary function: Yearbooks play an important role in the sense that they are part of the nation-building strategy or process: In the same way as national symbols like flags, national days, national monuments and national airlines, yearbooks symbolize and help shape national unity, integration and identity. In many ways a statistical yearbook represents the tip of the dissemination pyramid, and one of the functions of this kind of flagship is also to present (and represent) the nation state and the publishing institution and to increase their visibility and public recognition. Therefore, it rarely takes long after a NSO is established, before a yearbook is published.

Yearbooks come in many forms and in different shapes and sizes: In some countries, the yearbook has developed in the direction of a "coffee-table book", with lots of background information, photos and a nice design (Canada and New Zealand), other yearbooks are more analytical (Switzerland, Australia), whereas some consists of mainly tables (USA, Germany).

On the other hand, in some countries yearbooks are now being phased out or downsized: In France, the yearbook ceased to exist in 2008; after 110 years the 2007-issue was the last one. And in the Netherlands, the volume and format of the yearbook has been significantly condensed. In many countries the sale of yearbooks has been dramatically reduced in recent years, due to the increasing use of the Internet as the main dissemination channel.

Statistical yearbooks represent a type of publication that does not easily fit into the categories of "reference" or "analytical" publications. But statistical yearbooks should be regarded as (analytical) presentations, and treated accordingly. If and when a NSO decides to publish a yearbook, every effort should be made to ensure that the presentation aspects are central: The selection of indicators (tables/charts) should be done with great care, considering carefully the quality of the various indictors/tables as well as the needs of the users, and putting some effort into the production (layout and design). And of course, being a flagship, a yearbook should also be available on the website.

Periodicals

Publishing reports and analytical publications is relatively demanding. Since the motivation varies and the ability to do this kind of analysis to a large degree is a result of experience and training, maybe another strategy is feasible – to publish some kind of periodical journal, magazine or newsletter. Here, the NSO could report on social, demographic and economic states and trends, in the form of relatively short articles on various subjects (marriage/divorce, fertility, age structure, education, labour force participation, tourism, traffic accidents, crime, men and women, etc., etc.). To get more and fresh ideas for analysis, look at publications from other NSOs: In this business, "stealing" is not only common, it is also acceptable and should be encouraged.

There are several advantages of a magazine or newsletter of this kind: The individual articles are short and it's therefore easier to get someone to write them. At the same time, such writing will function as a sort of training for writing other and (longer) analytical commentaries. Such a journal/newsletter can include articles of various formats – everything from short "Notes" and information about special events to more in-depth articles on the various subject matters covered by the statistics.

If and when a NSO decides to start such a journal, here are some points to bear in mind:

- A title with the word "journal" in it, signals something rather ambitious. A title like "*News from XXX*" or "*XXX Newsletter*" is perhaps more fitting
- Start on a small scale, for instance 2 issues per year; one in spring and one in autumn
- Plan for 4, 8 or 12 pages
- Use A4-format
- The newsletter should be open to people from outside the NSO using NSO data, in fact you should invite contributions from outside
- All "articles" (except short notes) should be signed or (at least) have a "contact person"

Example of statistical newsletter: NIGER STAT, Bulletin trimestriel de liaison du Système Statistique Nigérien

- In addition to articles and statistical presentations, the newsletter could also bring short notes and news from and about the NSO and its various activities (new statistics, new publications, planned surveys, etc., etc.).
- The newsletter should be aimed at the educational sector, governmental offices, NGOs, etc.
- Use the newsletter to advertise new publications (such advertisements are also nice for filling up empty space)
- The newsletter should be published on the Internet (PDF-format), but should also be available in a printed version

Popular presentations

More popular presentations are mainly intended for the informed user and the general public. Other important target groups are pupils and students. Brochures and leaflets presenting key figures about a country or a specific topic (gender, population, education, housing, etc.) is a good way to "market" the NSO and make it more visible.

Producing such popular presentations is also a useful exercise for the NSO, as it will force us to make a selection of the most interesting and important figures and to present them in a straightforward and user-friendly way. (For an example of such a user-friendly publication, see appendix I).

Internal series

In the production of statistics, many issues are involved, such as methodological issues, issues of classification and standardization, etc. These questions are discussed from time to time, decisions are made, standards and guidelines are adopted. Most of these discussions are done by circulating papers, memos and documents.

Such discussions and decisions are of course of vital importance to the daily operation of a NSO. Therefore, when, for instance, a new classification or a new method has been adopted; it should be published in an internal series, and thereby made available to all staff. Even if such a series is primarily internal, the publications should also be available to interested external users.

Other examples of documents that may be published in such a series are result of pilot studies, handbooks/guides, strategy documents, plans, annual reports, etc.

9. "This is how we do it"
Making guidelines for dissemination

Dissemination guidelines and standards are important for two reasons: They improve the quality of the publications and news releases by making them "nicer", more uniform and more user-friendly. But just as important, clear rules and guidelines ease the internal production process, by making it clear what the publications should contain and how this content should be structured and formatted. *"This is how we do it".* Written rules and guidelines therefore greatly facilitate the production process *and* the process of approval and quality management.

It is therefore strongly recommended that every NSO should develop general guidelines for printed publications (including tables and graphs), news releases and media relations. The guide should contain *rules* (which *must* be followed) as well as *guidelines* (which are more to be seen as recommendations or advice).

The process of preparing the guidelines should involve different people, including statisticians, people engaged in the dissemination process as well as someone with experience and knowledge about the different aspects of layout and formatting.

More specifically, these guidelines should cover the following aspects of dissemination:

1. News releases and media relations

1.1.   When are news releases to be issued? Only when publishing new statistics or also when launching a new publication? Other occasions?

1.2.   Who writes them? Statisticians or …?

1.3.   Distribution of news releases: Printed and/or Internet/e-mail?

1.4.   Should there be a press conference? Who decides?

1.5.   Content and structure of news releases (text/tables/graphs/metadata/appendix tables).

1.6.   Should anybody (ministries, media?) get access to the statistics before it is officially released? And if so, under what conditions?

1.7.   If possible, an *advance release calendar* should be established, giving the date of forthcoming releases. What time span should be covered by the calendar and how should it be updated? What kind of distribution?

1.8.   Contact person/who should speak to the press?

2. Publication series

When there are several publication series, the guidelines should give a normative description of the different series: Which publication belongs to which series? Who decides?

2.1.   For the different series, *content and structure* of the publications should be described (elements that should/could be included, order of elements (preface/introduction, summary/abstract, contents, list of tables, list of graphs, main findings, etc.).The guidelines should also include a standard for the reporting of methodological issues/metadata (see 6. below).

2.2. Standard for *reference and citation*: How to refer to books, reports, articles in journals/in books, etc.

2.3. Rules for dating (of statistics and publication)

2.4. Rules for quality management and approval of the publications in different series

2.5. Rules for assigning serial numbers, pricing, number to be printed, distribution, etc.

2.6. *Design of series*: The different publication series should have their own "identity" in the sense that their design and layout should be distinctive and individual, clearly distinguishing them from the other series. But at the same time – of course – all the series should have some common features, identifying the publication as a product of the NSO

2.7. Rules for design of/information on: Cover, title page (page 1), etc. More or less detailed rules for design and layout: fonts, font sizes, one or two columns, alignment, pagination, titles/subtitles (levels, numbering), footnotes *or* end notes, placement of page number, use of headers and footers, appendices, etc. All these elements may seem like trivial details, but they should all be described and examples given

3. The Internet

Since this area is subject to very fast technological changes, rules and guidelines will probably have to be updated very often. Nevertheless, some *basic* rules and routine descriptions should be written down and distributed to everyone who contributes to the Internet. Such as:

3.1. Release calendar: How and when to update it?

3.2. Guidelines for publishing/editing/formatting news releases

3.3. Format of published tables: Excel, PDF or html? Should the tables be downloadable?

3.4. Publications: How should they be presented on the website? Summary or full text?

4. Tables

There should be a separate standard for the design of tables, including – among other things:

4.1. Fonts (type and size, normal/bold/italic): Use a sans serif font like Arial or Helvetica for the whole table. The font size should be 7 or 8 points normal

4.2. Use of horizontal and vertical lines/rulers: Minimalize the use of lines (especially vertical lines)

4.3. Number of decimals in percentage tables/tables giving rates, averages

4.4. Use of standard classifications (education, occupation, regional classification, age grouping, etc.)

4.5. Special symbols are often used in tables. These should be explained

4.6. The source of the table should be placed at the bottom. In a publication where all tables are from the same survey, it is not necessary to give the source for every single table. In publications presenting statistics from different sources – for example in a yearbook – the source of every table should be given: Health Survey 2006, Statistics Norway

Example of symbols used in tables	
Category not applicable	.
Data not available	..
Data not yet available	...
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0,0
Provisional or preliminary figure	*
Break in the homogeneity of vertical series	--
Break in the homogeneity of horizontal series		

From: Statistics Norway

5. Graphs
Guidelines for graphs should also be included, covering – more or less – the following elements:
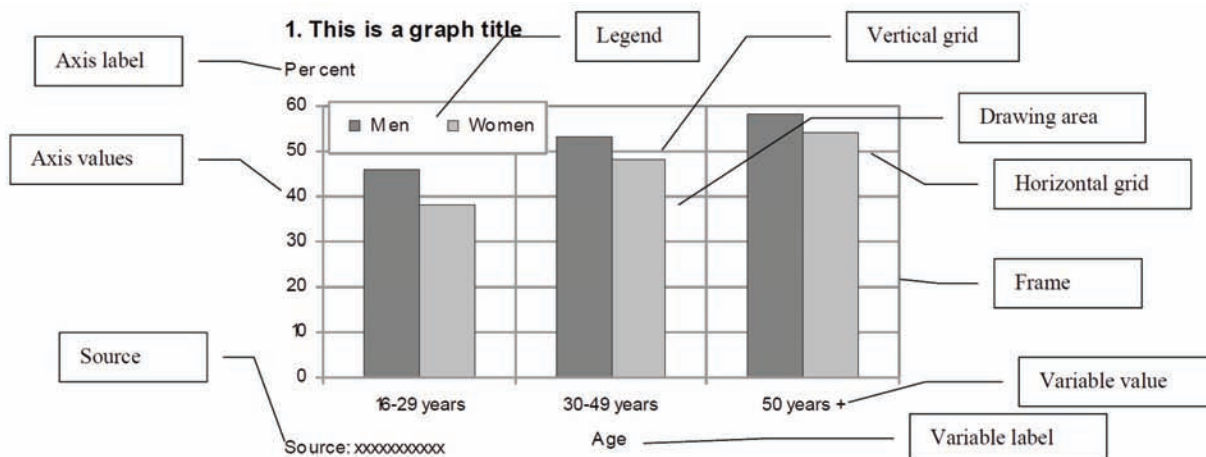
5.1.   Fonts (type and size, normal/bold): Choose a sans serif font like Arial or Helvetica for the whole graph. The general font size could be 7 points normal, but for the title use 8 points bold. Avoid italic

5.2.   Size and format: For most purposes, the size of the graph above (6.0 x 4.0 cm) should be suitable. Some graphs, like horizontal bar charts, require a different format and size. Different for printed publications and the Internet?

5.3.   Legend: If possible, place the legend inside the drawing area

5.4.   Axis label: Always specify the unit of measurement: Number, per cent, per 1,000, tonnes, etc.

5.5.   Use of frames, grids, patterns/colours: Decide on a standard – and stick to it!

5.6.   Titles of tables and charts (formulation, numbering, placement); listing of classification variables ("by..."?), units, dating

5.7.   Sources: See 4.6 above

When you have decided on the design/formatting of the graph (fonts, lines, colours/patterns, etc.), instead of changing all these elements every time you make a new graph, you can create your own templates or models in Excel. This is done in the following way:
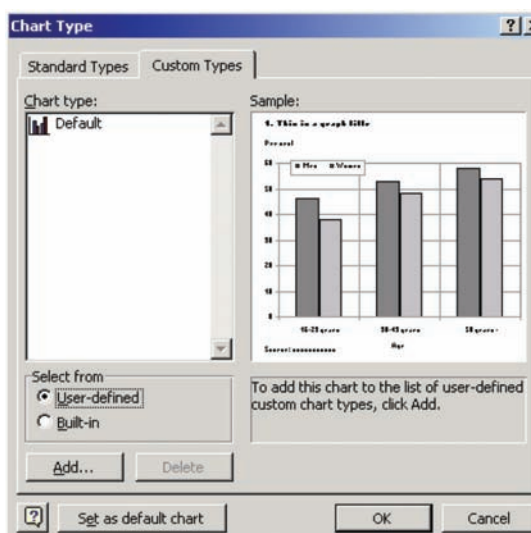
Click on the new, formatted graph and open "Chart Type". Choose "Custom Types" and then "User-defined". Your new graph will then be shown as a "Sample". Then select "Add".

You can now enter a name for the new template, for example "My column chart" (you may also add a more comprehensive description of the template, for example: "To be used in Reports and news releases"). Click OK – and the template is saved. Repeat this procedure for the
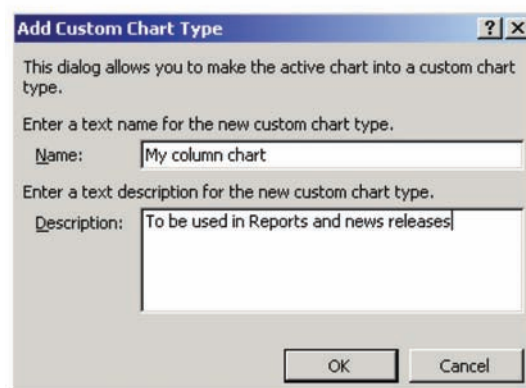
different types of graphs. Next time you make a graph, choose one of these templates instead of a standard type.

Note that this Excel template does *not* include the graph size; the graph will have to be resized every time, unless you use the default size.

6. Metadata

Metadata is basically "data about data". The concept seems, however, to be used in two ways: In the more restricted sense, metadata is a documentation of the data and gives information that is necessary to understand and interpret statistics correctly ("interpretation metadata"). Traditionally, metadata includes documentation about the population of a survey, observation unit, sampling, concepts and definitions, information about methods, calculation and estimation, etc.

More recently – with the spread of the Internet – metadata has taken on a more extended meaning: "Navigation metadata" includes all information that will assist the user in navigating and searching the website. In this sense, metadata includes almost everything and such an all-embracing concept becomes almost meaningless. In the present context we shall therefore use the concept in the more restricted, traditional sense of the word. And we shall also assume that the metadata requirements for Internet dissemination are not essentially different from that of printed publications (although the presentation may differ between the two media).

How much metadata?

The general answer to a question of how much metadata should be published, is: Enough. But what is enough for the layman is, of course, not always enough for the expert. In publications directed towards the general public, the reference to metadata is often simply given by listing the source and – if necessary – one or several notes. In reference publications, of course, the metadata presentation should be more detailed.

Every NSO should develop some kind of minimum standard for metadata ("data about data") for all their statistical products – to be published in printed publications as well as on the Internet. The guidelines should specify what information should be included and how this information should be structured. Whatever they are called ("About the statistics", "Declaration of content", "Guide to statistics", "Documentation") this kind of metadata usually describes several aspects of the statistics, like:

6.1. Content of the statistics: Population, data source(s), data collection, questionnaire

7.2. If relevant: Type of sample, sample size, response rate, etc.

6.3. Definitions of concepts and variables, classifications

6.4. Methods used: Estimation, index construction, seasonal adjustment, computation of rates and ratios

6.5. Frequency of publication, reference period. Are the results final or preliminary? If preliminary, when will final results be available?

6.6. Sources of error and uncertainty: Non-response errors, sampling errors, other sources of error. If relevant, give estimates of variance

6.7. Comparability: Over time and with other relevant statistics

Finally, this information could be organized in some kind of internal database, enabling easy production of both printed and electronic versions of the metadata. In this way, NSOs could gradually build a "Guide

to Statistics", giving information to users about content, structure and dissemination of all statistical products.

In addition to these rules and guidelines, there should be a descriptive guide of the production line: Who should be doing what? How should the manuscript be delivered from the statistician to the person responsible for the layout; what kind of formatting should be done by whom? Very much time today is spent by layout people, de-formatting and "cleaning" the manuscripts from writers. People enjoy experimenting and formatting in *Word* and *Excel*. But remember: Layout and formatting is a job for professionals; it is *not* the job of statisticians or writers. They have far better things to do.

When these rules and guidelines have been finalised and adopted by the organisation, they should be printed in an appropriate internal series and introduced to the organisation through a series of meetings with all the subject-matter departments, explaining and giving the reasons for adopting the standard (better, more user-friendly publications/more efficient production lines, etc.). Telling people *why* they should do it this way is far better than just telling them that they should do it.

The basic principles of the dissemination policy should also be published externally, so that central users are informed about the main aspects of the policy, such as equal treatment of all users, etc.

Read more?
Some suggestions

PARIS21: *Guide to Designing a National Strategy for the Development of Statistics*, 2004. http://www.paris21.org/documents/1401.pdf

Rafalowska, Helena: "Building the reputation of a statistical office through effective communication". *Statistical Journal of the United Nations ECE* 22 (2005) 147-156.

United Nations Economic Commission for Europe*: Communicating with the Media. A Guide for Statistical Organizations*. Geneva, 2004. http://www.unece.org/stats/documents/media/guide/Welcome.html

United Nations Economic Commission for Europe: *Making Data Meaningful Part 1: A guide to writing stories about numbers*, Geneva, 2009. http://www.unece.org/stats/documents/writing/

United Nations Economic Commission for Europe: *Making Data Meaningful Part 2: A guide to presenting statistics*, Geneva, 2009. http://www.unece.org/stats/documents/writing/

United Nations: *Handbook of Statistical Organization*, Third Edition: The Operation and Organization of a Statistical Agency, 2003. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf

Appendix I:
This is Norway — What the figures say
An example of a user-friendly publication

In 2003 Statistics Norway published for the first time a small, summary presentation of statistics about Norwegian society, aimed at a wider circle of readers: *This is Norway – What the figures say*. The main idea was that a serious publication about statistics does not necessarily have to be boring – or complicated and difficult to understand! Feedback from various users has been very positive, indicating that this publication has been relatively successful in presenting statistics in an informative and user-friendly way to a wide variety of users. The booklet has been updated and revised twice and a similar booklet on *Woman and Men in Norway* has also been published.

In the following we shall outline some of the basic principles guiding the production of this booklet. If you want to make your own judgement about the publication, it is available in PDF-format on the website of Statistics Norway: http://www.ssb.no/norge_en/. Here, printed copies may also be ordered, free of charge

Structure of publication

This is Norway is based on the principle of **KISS**: **K**eep **I**t **S**hort and **S**imple. Altogether 22 different aspects of Norwegian society are covered in 44 pages. As can be seen from the table of contents, the publication covers most aspects of Norwegian society; presenting a wide selection of demographic, social as well as economic and business statistics.

To keep the publication as simple as possible, there is no summary, detailed table of contents, list of tables or graphs, list of abbreviations, appendix or subject index.

To draw the attention of the readers, the various chapters were given striking titles. At the same time, the titles should be substantial – in the sense that they should give a summary of or an indication of the content of the chapters.

The structure of the different chapters varies, of course, depending on the subject and the available indicators in various areas. But in all chapters, a major focus is on long-

This is Norway: Table of contents

term changes, described by time series (up to 50-60 years). Another central aspect is international comparison; how does Norway compare with other countries? In the presentations of population and social statistics, gender plays a central role as a classification or background variable, comparing women and

Example of presentation: Fertility

men to see if and how gender differences are changing. Also regional variations are presented; in graphs and maps.

As in the example shown above, all topics or subjects are presented in two pages, consisting of a suitable mixture of text, tables, graphs and/or maps. Using exactly two pages on every subject is – of course – a constraint, but it gives the publication a clear structure. It also forces you to concentrate on the main aspects, without going into the details of the various phenomena.

According to the KISS-principle the text is written in a clear and simple language. It is mainly descriptive, but with some interpretations and explanations. The use of short paragraphs and subheadings makes the text more inviting and easier to read. Also, most of the subheadings are substantial; instead of "Age of mothers" it says "Older mothers".
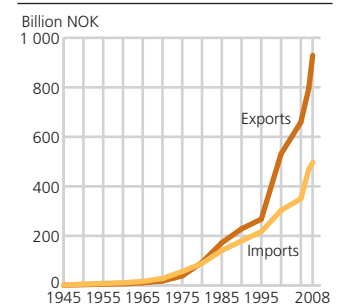
The size of tables and graphs are kept at a minimum. And to keep them small, it is important that they are clear and simple, not trying to compare too many groups.

Exports and inports of goods (including oil and gas)

Exampel of graph: Full scale

Content: Something for everybody

Since the publication is aimed at a wide circle of readers, short-term economic and business statistics are somewhat underrepresented. And although most of the statistics belong to the standard range of statistical products, some statistics have been included as a curiosity; like lists of most popular first names, most common makes of car and also shopping, drinking and sleeping habits. Motto: "Something for everybody".

The data sources are mainly statistics published by Statistics Norway; but – when relevant – data from other sources are also included. International comparisons are based on data from international organization, like Eurostat, OECD and IEA.

Appendix II:
Training for more user-friendly dissemination
Some possible topics

Since dissemination of statistics – unlike other, more methodological aspects of statistics – is not a subject that is thought at universities, the most common way to learn dissemination is by experience ("learning by doing"). But because dissemination is often decentralized and is done by the same people that collect and process the data, this experience is seldom cumulative and therefore does not always become part of the institutional memory.

But another way of increasing and improving the knowledge and understanding of user-friendly dissemination is by organizing in-house training for employees, arranging workshops and giving courses on the various aspects of dissemination. Every NSO should therefore make an assessment of its needs for training in the various aspects of dissemination – as a first step towards making a more detailed program or plan for organizing capacity building in this area. Here are some aspects of dissemination which are suitable for training:

1) Writing better reports and news releases

While many journalists have problems understanding and using numbers, many statisticians have problems using words; writing about numbers does not come easily to every statistician. Statisticians tend to write in an "academic" way, starting with the various aspects of data collection and methods, before presenting the detailed results and – finally – the conclusions. But when writing for the general public and the media, the structure should be reversed: First conclusions, then more detailed results and (maybe) something about methods.

By hiring journalists to give a practical course in "popular" writing, news releases and reports can be greatly improved. A very simple, but practical and effective way of teaching better writing is to give participants some tables and ask them to write a short "news release": What do the figures say? What are the main results? In a process of writing, feedback/discussions and rewriting the participants will gradually improve their writing skills and their ability to focus on a few, central results.

2) Making better graphs

Statistics is today increasingly presented by means of graphs. But many statistical publications and websites are filled with meaningless and ambiguous charts that are often even misleading. Making good, clear and informative graphs requires some experience and/or training.

The production of user-friendly graphs can be divided into two parts: Content and design. The content of a graph is often decided by the statistician: What kind of graph is suitable, what data should be included, etc.? The statistician will then typically make a draft or a sketch, which will be discussed with and finalized by somebody taking care of the final design and layout. Both these groups – statisticians and designers – could probably benefit from a basic training course in the construction of graphs.

For both these courses, the ideal number of participants is 6-10 (maximum 15). The courses should start with a general introduction to the topic, followed by more practical training sessions. The courses should last for at least 2 days and the participants should have access to PCs and a printer.

The trainer(s) could be national as well as international experts. To transfer the skills and knowledge generated during the course, it is often a good idea to appoint some kind of local "assistant trainer", who can later arrange similar courses.

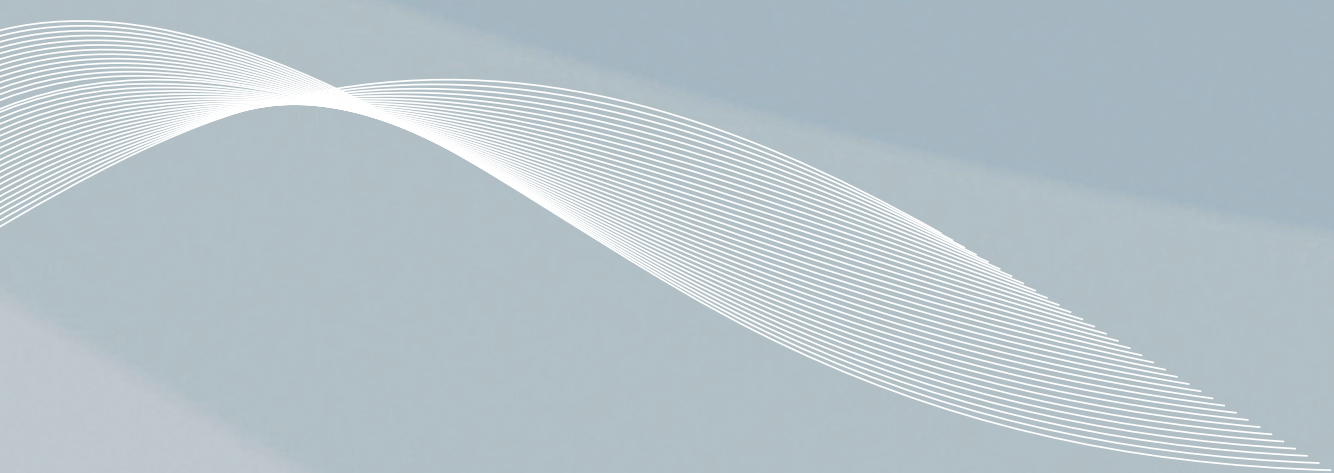## 3) Media relations and training of journalists

Since the media are important in so many ways (reaching the public, increasing visibility and confidence in the statistics), inviting the media into the NSO to learn more about statistics, is an investment in the future. At the same time, this will give the statisticians a better understanding of how media think and work.

- On the one hand, giving journalists some basic knowledge of statistics will increase their understanding of and interest in statistics, thereby making them more motivated and better prepared for using statistics in the future.
- To increase the statistical literacy of journalists, they should be given a course in some basic statistical concepts: Absolute and relative numbers, percentages and percentage points, ratios, rates, index numbers, etc. In addition, they should be taught something about sampling, errors and uncertainty. It is also a good idea to take them on a "guided tour" on the NSO's website, so that they can more easily find what they are looking for.
- On the other hand, journalist can give statisticians a better understanding of the media and how to handle them. What are the needs of the media? Do different media – radio, newspapers and television – have different needs? What is considered to be a good "story"? In addition, journalists could give some advice about being interviewed, to give statisticians more self-confidence when talking about statistics.

For more information on media relations, see *Communicating with the Media: A Guide for Statistical Organizations* (UNECE 2004), which also includes a detailed section on the training of statisticians by journalists and vice versa, with many useful hints about techniques and approaches.