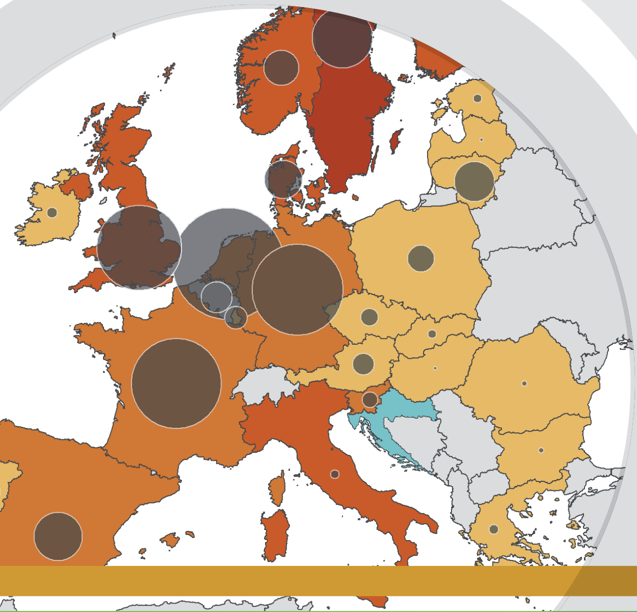
ecdc

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL



TECHNICAL DOCUMENT

Guidelines for presentation of surveillance data

Tables, graphs, maps

www.ecdc.europa.eu

ECDC TECHNICAL DOCUMENT

Guidelines for presentation of surveillance data

Tables, graphs and maps

This publication of the European Centre for Disease Prevention and Control (ECDC) was coordinated by Joana Gomes Dias and Frantiska Hruba (Epidemiological Methods, ECDC).

Contributing authors

Joana Gomes Dias (lead author), Denis Coulombier, Silviu Ionescu, Fabrice Donguy, Virginia Estevez and Yvan Hutin.

The publication was reviewed by Bruno Ciancio, Frantiska Hruba, Julien Beauté and Philip Zucs.

Suggested citation: European Centre for Disease Prevention and Control. Guidelines for presentation of surveillance data. Stockholm: ECDC; 2018.

Stockholm, March 2018

ISBN 978-92-9498-092-2

doi: 10.2900/452488

TQ-04-17-831-EN-N

© European Centre for Disease Prevention and Control, 2017

Reproduction is authorised, provided the source is acknowledged.

Disclaimer
Please note that all tables, graphs and maps showing distribution of diseases (in particular, Figures 30, 34, 35, 36 and 37) are designed to serve as fictive illustrations and do not in any way reflect the actual situation in the countries.

Contents

Figures

Tables

1 Introduction

Data presentation is the basis for describing data and generating hypotheses for further testing. It consists of a number of methods for summarising data in order to support the identification of patterns. Data can be summarised in a number of forms, including tables or graphical representations such as graphs or maps.

Effective data presentation requires an understanding of the principles introduced in this guide.

1.1 Aim of the guidelines

This document provides guidance to support epidemiologists and surveillance experts in producing tables, graphs and maps to show the results of their data analyses following harmonised principles and practices. It aims to maximise the potential of presentation tools to depict relevant patterns in data, where possible using ECDC examples.

Experts are encouraged to apply the principles and practices provided by these guidelines, taking into account the most appropriate data representation, which ultimately depends upon the characteristics of the data. Therefore, these guidelines are not meant to impose one specific method of data representation, but rather to support epidemiologists and surveillance experts in making informed decisions when selecting a presentation type. The decision to apply a particular presentation type remains the prerogative of epidemiologists and surveillance experts. Moreover, this decision will depend on the formatting guidelines stipulated by the scientific journal targeted for publication.

Key areas addressed in these guidelines are:

- How to frame the message;
- How to choose the appropriate visual presentation (table, graph or map) for the data.

Throughout the document, there are tips for creating good tables, graphs and maps for scientific publication and descriptions of common errors in order to facilitate informed choices.

When choosing between tables, graphs and maps, some general principles should be considered:

- Tables show all the details of numeric data and, for example, allow for the results and conclusions of the analyses to be verified by the reader;
- Graphs enable information to be summarised and displayed visually, often centred on one key phenomenon/message;
- Maps are the preferred method for displaying information by geographical area where relevant.

1.2 Topics covered

These guidelines are intended to provide epidemiologists and surveillance experts with effective ways for creating outputs such as tables, graphs and maps with a consistent presentation style to ensure better representation of data.

The document is organised as follows:

Chapter 1 - An introduction describing the document’s rationale, objectives and structure;

Chapter 2 - An overview of main data characteristics;

Chapter 3 - An overview of aspects to be considered when preparing graphs, tables and maps;

Chapter 4 to Chapter 6 - Specifics for each type of output and effective methods for the most commonly used disease indicators at ECDC;

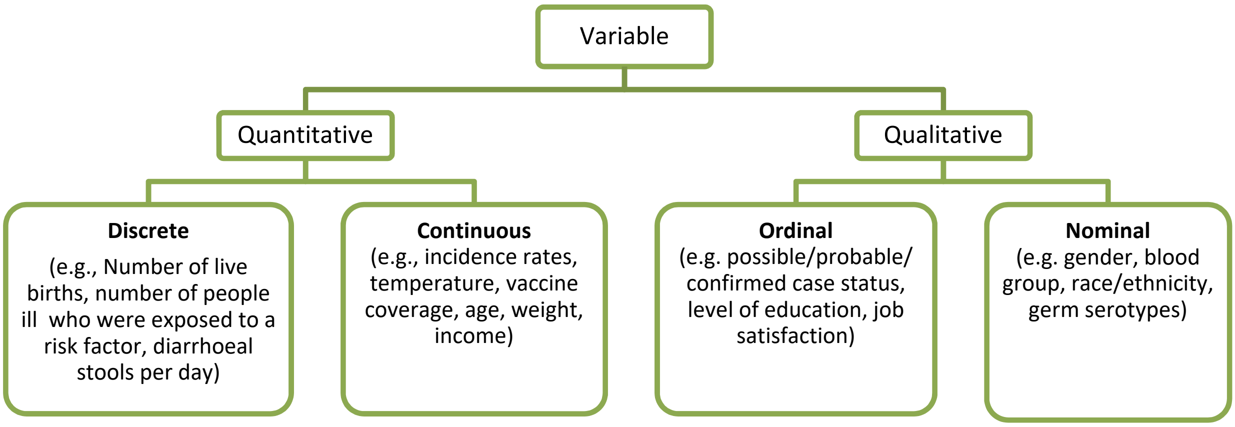
Chapter 7 - Bibliography for further reading.

2. Data considerations

2.1 Type of variables

Data to be presented consist of quantitative or qualitative variables (Figure 1). The nature of the variable is important as it will result in different outcomes and determine the most suitable type of presentation.

Figure 1. Types of variable flowchart



2.1.1 Quantitative variables

Quantitative variables have mathematical properties and can be discrete or continuous. A discrete variable results from an enumeration of events or cases (e.g. number of cases per week). It is expressed as an integer. A continuous variable derives from a calculation (e.g. rate, risk ratio, vaccine coverage, etc.) or a measurement (e.g. age, weight, temperature or blood pressure). It is expressed as a real number, because observations can take any value between a certain set of real numbers.

2.1.2 Qualitative variables

Qualitative variables are used for classification. The classification can be either ordinal or nominal. An ordinal variable is logically ordered or ranked — for example to describe the spread of influenza in the EU (low, medium, high, very high). A nominal variable, however, has no intrinsic order — for example gender or country.

Note that a quantitative variable, such as age, can be reclassified as age groups through recoding to become a qualitative ordinal variable when age categories are not at equal intervals.

2.2 Epidemiological indicators

The initial step in epidemiological studies is to measure the frequency of health events in order to make comparisons between groups, either within a population or across populations.

Epidemiological information is usually expressed as counts, proportions, ratios or rates.

A **count** is an enumeration of events of interest, such as the number of affected individuals. Counts take the form of integers. They are used to quantify the burden of a condition for allocating public health resources in a specific community setting. Since they lack information on the size of the underlying population, they are of limited use in analytical epidemiology (e.g. comparing risks across populations.) For example, we can count three cases of tick-borne encephalitis on the west coast of Sweden in a given week, but we are not able to compare the risk they represent with another area of Sweden or with the number of cases at another time of the year. Nevertheless, counts are commonly used in stable populations to describe trends over time (e.g. in an outbreak situation using an epidemic curve).

A **ratio** results from a fraction whereby the numerator is not included in the denominator. Ratios are real numbers that can take values from 0 to infinity. In a ratio, anything goes – for example, the ratio of children to trees in a forest. The sex ratio or incidence rate ratios are commonly use in epidemiology.

A **proportion** results from a fraction whereby the numerator is included in the denominator. Proportions are real numbers that can take values from 0 to 1 or from 1 and 100 when expressed as a percentage. For example, of the 47 children who went on a school trip to the archipelago, 34 (72%) had received tick-borne encephalitis vaccine.

A **rate** results from a fraction whereby the numerator is the number of health events that occurred over a time-period and the denominator is the average number of subjects under observation at risk during the same time-period. These are then multiplied by a constant, also called the multiplication factor. There are also rates where a denominator has person-time (e.g. 5 cases/100 000 person-days), used in cohort studies. Rates are real numbers that can take values from 0 to infinity. In general, we should use a multiple of ten as the multiplication factor so that the smallest rate calculated is a small whole number, for example 5.1/1000, instead of 0.51/10 000. When comparing two rates, be sure that the same multiplication factor was used to calculate both rates. Choose a multiplication factor that is compatible with commonly published rates for the topic (e.g. rates per 100 000 population for deaths and rates per 1 000 population for live births).

Table 1. Some epidemiological indicators and measures

Epidemiological indicator	Measure
Number of cases	Count
Male-to-female ratio	Ratio
Standardised mortality ratio (SMR)	Ratio
Odds ratio	Ratio
Case-fatality	Proportion
Vaccine coverage	Proportion
Incidence/prevalence rate	Rate
Notification rate	Rate

3. Presentation considerations

Data presentation involves summarising data in tabular or graphical format to explore patterns in the data.

An effective data presentation is defined by the following quality factors:

- Clarity – the reader can easily understand what is represented and the main characteristics of the data displayed;
- Veracity – the presentation is a true expression of the data pattern;
- Objectivity – the presentation does not bias the perception of the data pattern;
- Conciseness – the representation summarises the information concisely, avoiding additional non-essential information.

3.1 Digital versus analogue

When considering how to represent epidemiological data it is first necessary to decide on a digital or analogue presentation (Figure 2). Digital communication is more precise, takes a numerical format and provides a detailed description of the information to be communicated. Analogue communication is less precise, takes more of a graphic format and provides a good overall understanding.

Figure 2. Digital and analogue representation of time

Digital clock **Analogue clock**

3.1.1 Digital presentations

Digital presentations use tables that list numeric values according to one or more classification variables. The data presented are extremely accurate, precise and unambiguous. All values presented can be checked individually. However, it can be difficult to grasp the overall pattern in the data from a digital presentation. The example in Table 2 of digital information on an outbreak of monkeypox shows the exact numbers of primary and secondary cases.

Table 2. Cases of monkeypox by month of onset, Katako-Kombe, Democratic Republic of Congo, 1996–1997

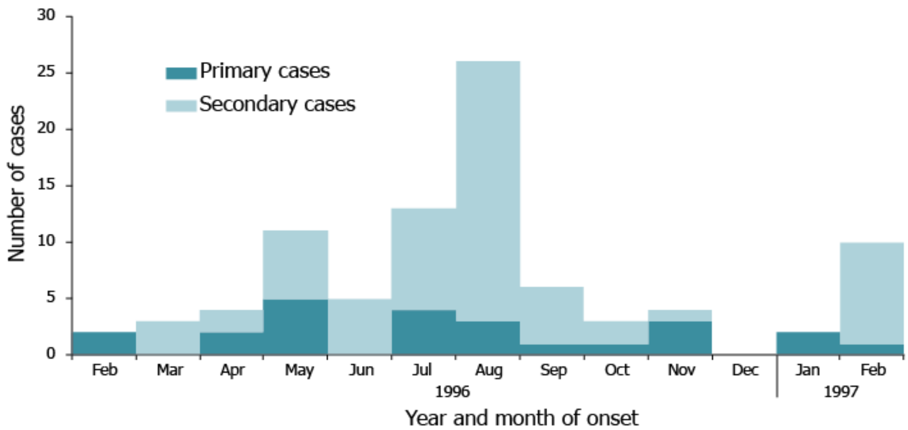
Month	Primary cases	Secondary cases
February 96	2	0
March 96	0	3
April 96	2	2
May 96	5	6
June 96	0	5
July 96	4	9
August 96	3	23
September 96	1	5
October 96	1	2
November 96	3	1
December 96	0	0
January 97	2	0
February 97	1	9
Total	**24**	**65**

Source: Hutin YJF, Williams RJ, Malfait P, Pebody R, Loparev VN, Ropp SL, et al. Outbreak of Human Monkeypox, Democratic Republic of Congo, 1996 to 1997. Emerg Infect Dis [abbreviated publication name]. 2001, Jun [date cited]. http://dx.doi.org/10.3201/eid0703.017311

3.1.2 Analogue presentations

Analogue presentation uses an analogy to depict the phenomenon of interest. In the analogue clock represented in Figure 2, the rotation of the needles is an analogy with the apparent rotation of the sun during a day. On an analogue clock, it is impossible to know the exact time at a glance. However, it is possible to immediately estimate the time left until a meeting without the need for mental calculation. Analogue presentations are effective for enabling data exploration and the identification of patterns and relations in data. Figure 3 is an example of analogue information regarding the outbreak of monkeypox. It shows how adding the secondary cases to a baseline of primary cases provides a better depiction of the persisting nature of this outbreak.

Figure 3. Cases of monkeypox by month of onset, Katako-Kombe, Democratic Republic of Congo, 1996–1997



Source: Hutin YJF, Williams RJ, Malfait P, Pebody R, Loparev VN, Ropp SL, et al. Outbreak of Human Monkeypox, Democratic Republic of Congo, 1996 to 1997. Emerg Infect Dis [abbreviated publication name]. 2001, Jun [date cited]. http://dx.doi.org/10.3201/eid0703.017311

3.1.3 Choosing between digital and analogue presentations

An analysis of epidemiological data usually includes both representations of the distribution of the variables used. Digital representations are used to display unambiguous values for indicators, required when using thresholds or target values. Analogue representations allow a very large amount of data to be grasped easily and also enable visual exploration and extrapolation.

Representation of certain types of data imposes a mode of presentation. For example, to describe the dynamic of an outbreak, an epidemic curve (Figure 18) is the most appropriate mode of presentation. On the other hand, relative risks or odds ratios associated with specific exposures in the context of an outbreak investigation are better represented in a tabular format, sorting the variables by magnitude of the association. At times, both representations can be used, conveying different, yet complementary information. Digital and analogue information should not be presented together because they compete with one another and make it difficult to understand the data. For example, it is not advisable to indicate the numerical value associated with a data point on a graph.

3.2 Data presentation: common attributes

Common principles apply when defining the titles of tables, graphs and maps.

Titles should include the indicator(s) presented (specifying if the results are adjusted or not), the population group from which the indicator was calculated (e.g. prevalence of MRSA carriers among healthcare workers), the period covered by the data, and the place where the data were collected. In some cases, there could be a disclaimer or note(s) to provide further information or explain a special situation.

A title should be a self-sufficient description of the representation. Titles should be written using lower case font, except for the first letter and proper nouns. The use of acronyms should be restricted to recognised ones (e.g. ECDC, WHO, TB, etc.) and acronyms must be explained, either in a list of abbreviations, in parenthesis the first time they appear in the text and/or in footnotes.

Footnotes help make tables self-explanatory. They clarify points of potential ambiguity. While some abbreviations are difficult to avoid (e.g. OR, CI), they must still be explained in the footnotes. Footnotes may also include symbols, codes (e.g. N.A. for not applicable), exclusions, or details of the data source, if applicable.

An example of an appropriate title is ‘Distribution of confirmed cases of *Salmonella* Mikawasima by week of reporting, EU/EEA countries, week 30 to 52, 2013’. Titles should be kept concise and informative. The number of cases in the representation is often shown in parenthesis at the end of the title (n=1 253), but this is optional. The title does not include the source of the data that can appear in a note below the representation/figure.

The date of data collection or data extraction for ongoing collections may be included if it is important for the understanding of the graph, table or map. If this is the case, it should be expressed in the DD/MM/YYYY format and be located in the bottom left-hand corner.

Figures, maps, publications and other sources not produced by ECDC must always be acknowledged and permission to reproduce them must be obtained from the copyright holder. This is also the case for Eurostat products and for Nomenclature of Territorial Units for Statistics NUTS.

The ECDC standard for tables, figures and their captions is that these are left-justified (i.e. they are not centred on the page).

3.3 Choice of colour palette

The choice of a colour palette is important as it has a strong impact on the perception of your data by the reader. The perception will be different depending on whether an area is depicted in red or in green, and this in turn will result in an initial positive or negative perception of your data. It seems obvious but common mistakes are often made. Therefore, the palette used should be chosen carefully depending on the message to be conveyed.

The colour palette in Table 3 below will help you to create clean, readable and consistent graphs and maps. A palette usually consists of a maximum of seven steps as the human eye is not able to distinguish between many shades of the same colour. Limiting the number of steps is also a good way to make your graphs and maps clear for the reader. To ensure optimal readability, use the different colours according to the number of steps in your scale.

Colours are defined by three characteristics:

- their hue, corresponding to the wave length of the light emitted, as in a rainbow;
- their luminance, corresponding to the amount of light emitted;
- their saturation, corresponding to the purity of the colour.

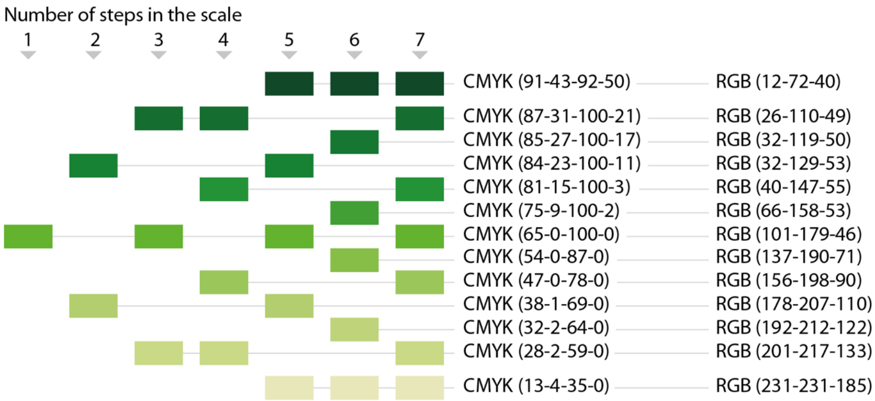
When choosing a colour to represent data, some of the basic rules to consider are set out in Table 3 below.

The colours chosen are the result of a combination of different requirements in order of importance: adherence to the basic rules and standard practices for data visualisation, being easily understandable, offering maximum accessibility to those who are colour-blind, being compatible with the ECDC corporate identity and being pleasing to the eye.

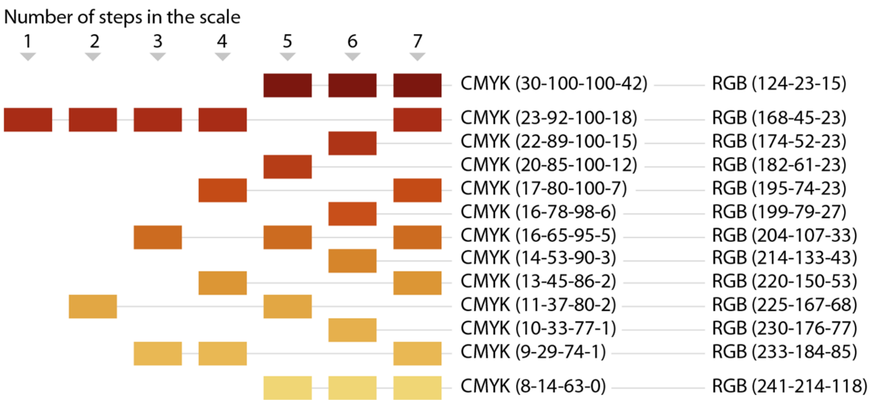
The colours below fit ECDC's requirements but are also valid for any other organisation and can be used widely. It is also possible to adapt these colours to suit different needs.

Table 3. **Colour palette**

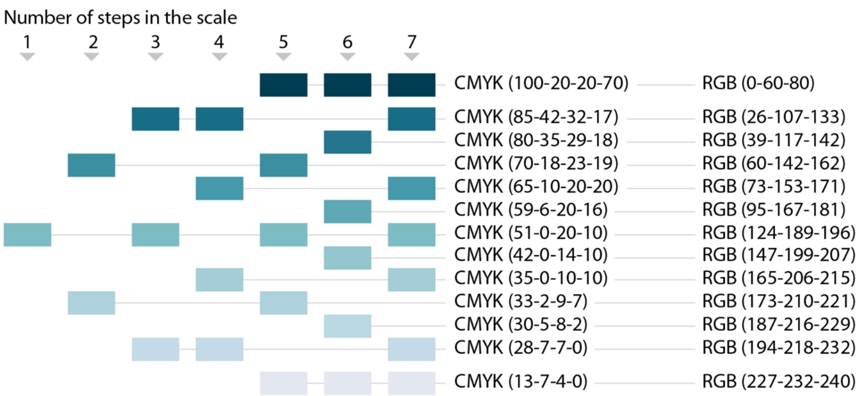
Description
Single variable
When one single variable is represented. It could be any colour. At ECDC we use our corporate green.
CMYK (65-0-100-0) RGB (101-179-46)
Multiple quantitative palette
Increasing density in one colour to represent increasing values in the data.
Green is used for positive increase (e.g. increasing vaccine coverage)



Red is used for negative increase (e.g. increased risk or number of cases)



Blue is primarily used for neutral increase (e.g. differentiating age groups), but also for positive increase in a cold/warm palette (see below)



Cold/warm palette
The cold/warm scale is very effective for representing strong contrasts in data (e.g. areas having met/not met a target value).

A red-to-green palette is often used to represent an indicator with values ranging from good (green) to bad (red). While culturally this 'traffic light' scheme is well understood, such a palette should be avoided, as those who are colour-blind will not be able to interpret the map.

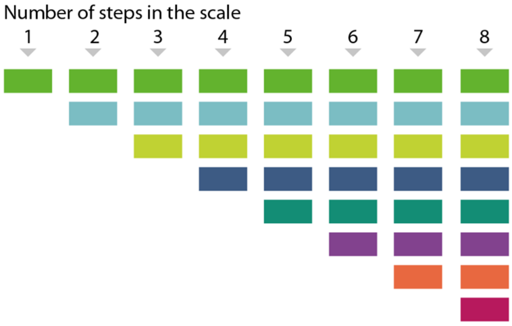
Examples of cold/warm palette

See values in the palettes above

Multiple qualitative palette

Complimentary colours are commonly used to represent qualitative indicators (e.g. a predominant subtype of influenza across countries).

At ECDC we use a combination of different hues, luminance and saturation to improve readability for those who are colour-blind.



CMYK (65-0-100-0)	RGB (101-179-46)
CMYK (51-0-20-10)	RGB (124-189-196)
CMYK (33-0-89-0)	RGB (192-210-54)
CMYK (84-63-30-5)	RGB (62-91-132)
CMYK (83-21-61-5)	RGB (0-140-117)
CMYK (60-84-6-0)	RGB (130-66-141)
CMYK (3-70-77-0)	RGB (232-104-63)
CMYK (30-100-40-0)	RGB (184-26-93)

Greyscale palette

The greyscale is used to depict specific data or areas.

In graphs, average data generally means that totals are in dark grey. Light grey is used for variables with unknown data.

In maps light grey is generally used to depict areas not concerned by the data. Medium light grey is used for unknown data or can also be used to highlight a selected area.

CMYK	RGB	**Data**	**Boundaries in maps**
CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50)	RGB (128-128-128)		
CMYK (0-0-0-70)	RGB (113-113-113)		
CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

4. Tables

4.1 Purpose and usage

Tables can be used to organise data, particularly numbers, more clearly than can be done with words. Tables communicate patterns and messages in your data and can summarise large amounts of data, particularly detailed data to facilitate comparisons. Tables are intended to be assimilated quickly by the reader and they must therefore be presented clearly and include the information needed to interpret the data. In fact, the table must be self-sufficient to enable the reader to make an assessment without having to check other resources. As a general rule, a few numbers can be described in a text, whereas a table should be used whenever more than five numbers need to be reported.

This document contains tips on including tables in scientific articles, as well as examples of common errors and possible solutions. The dummy table section can be used to create tables prior to data analysis, but please refer to the formatting guidelines for tables in the scientific journal you plan to submit to.

4.2 Types of tables

Tables are two-dimensional representations organised in rows and columns. Rows usually represent classification variables used in the analysis (e.g. age categories in Table 5). Columns can be used to display one or several distributions of indicators (e.g. proportion of completeness in Table 4) or number of cases and rate per million, as in Table 4). In addition, a second level of classification can also be presented (e.g. number of cases and rates by gender, as in Table 5). Cells display the value of indicators.

Line-listings are a specific type of table presenting a list of characteristics in columns for a collection of entities presented in rows. Table 3 is an example of a line-listing in which the characteristics of MERS-CoV cases are represented as a listing. Cases are organised in rows, variables are in columns and cells represent individual values for cases.

4.2.1 Line-listing

A line-listing table displays entities as lines, and their attributes as columns. Line-listing information does not usually require a total for lines or for columns. All column headers and values are centred, because line-listings often do not have a row header. A line separates the column header from the body of the listing.

Table 4. Characteristics of the first ten cases of MERS-CoV reported to WHO, April 2012–January 2013

Onset date	Sex	Country	Outcome	Exposure	Age
??/04/2012	Female	Jordan	Dead	Nosocomial	45
??/04/2012	Male	Jordan	Dead	Nosocomial	25
13/06/2012	Male	Saudi Arabia	Dead	Other	60
03/09/2012	Male	Qatar	Dead	Other	49
10/10/2012	Male	Saudi Arabia	Alive	Other	45
12/10/2012	Male	Qatar	Alive	Other	45
30/10/2012	Male	Saudi Arabia	Alive	Other	31
24/10/2012	Male	Saudi Arabia	Dead	Other	39
05/10/2012	Male	Saudi Arabia	Dead	Other	70
24/01/2013	Male	Saudi Arabia	Dead	Other	60

Line-listing can display the distribution of a classification variable in columns for each entity presented as lines. Here the first column represents different entities for which the distribution of the indicator is presented. Row labels are left-aligned, while indicator value columns are centred. Table 4 is an example where the entities are variables in a dataset, presented as lines, and the distribution in columns represents the completeness for different years.

Table 5. Completeness of reporting Legionnaires' disease cases for selected variables, by reporting year, EU/EEA countries, 2009–2012

Variable	2009 %	2010 %	2011 %	2012 %
Date of onset	96	95	97	98
Outcome	68	69	70	71
Cluster	70	63	60	72
Cluster Id	>99	83	98	85
Probable country of infection	97	93	94	92
Place of residence	21	30	35	36
Sequence type	1	1	3	4
Setting of infection	89	89	87	88
Environmental investigation	40	33	37	43
Legionella found	94	96	92	90
Positive sampling site	85	73	83	77

4.2.2 Tabulating distributions

Since the column values in Table 5 represent a distribution, the values can be summed as a total, if expressed as a count, or otherwise averaged. In this case, a total column is optional.

Table 6. Distribution of reported cases and rate per million population of Legionnaires' disease by gender and age group, EU/EEA, 2012

Age (years)	Males		Females		Total	
	Number of cases	Rate/ million	Number of cases	Rate/ million	Number of cases	Rate/ million
0-19	19	0.3	17	0.3	36	0.3
20-29	64	1.9	35	1.1	99	1.5
30-39	225	6.3	71	2.0	296	4.2
40-49	639	16.8	169	4.5	808	10.7
50-59	983	29.2	322	9.2	1 305	19.0
60-69	973	36.8	386	13.3	1 359	24.5
70-79	728	40.7	346	15.1	1 074	26.4
80-89	417	54.5	298	22.0	715	33.8
≥90	52	59.5	61	25.6	113	34.7
Total	**4 100**	**15.6**	**1 705**	**6.2**	**5 805**	**10.8**

4.2.3 Tabulating epidemiological study results

The two most common types of epidemiological study are cohort studies and case-control studies.

In a cohort study, subjects are enrolled on the basis of their exposure and then followed to determine the occurrence of disease. In a case-control study, subjects are enrolled according to whether they have an outcome or not and then evaluated to determine their prior exposure. Analysis begins with a description of the subjects' characteristics. It progresses to calculation of rates, creation of comparative tables (e.g. two-by-two tables) and computation of measures of association (e.g. risk ratios or odds ratios), tests of significance (e.g. chi-square test), confidence intervals or similar. Many epidemiological studies require more advanced analytical techniques, such as stratified analysis, regression or modelling.

Analytical epidemiological studies use standard presentations, as shown in the dummy tables below. The suggested layout is for tables reporting key numerical results for each type of study.

Table 7. Dummy table for case-control study: characteristics of (outcome X) cases among cases and controls, place, time

Exposures	Category	Frequency of exposure		Odds ratio	
		Among cases (n=xx)	Among controls (n=xx)	Estimate	95% confidence interval
Exposure A	Category A1	XX (XX%)	XX (XX%)	XX	XX-XX
	Category A2	XX (XX%)	XX (XX%)	XX	XX-XX
Exposure B	Category B1	XX (XX%)	XX (XX%)	XX	XX-XX
	Category B2	XX (XX%)	XX (XX%)	XX	XX-XX

Table 8. Dummy table for cohort study: incidence of (outcome X) among exposed and unexposed, place, time

Exposures	Category	Incidence		Relative risk (risk ratio or rate ratio)	
		Among cases exposed	Among unexposed	Estimate	95% confidence interval
Exposure A	Category A1	XX (XX%)	XX (XX%)	XX	XX-XX
	Category A2	XX (XX%)	XX (XX%)	XX	XX-XX
Exposure B	Category B1	XX (XX%)	XX (XX%)	XX	XX-XX
	Category B2	XX (XX%)	XX (XX%)	XX	XX-XX

4.3 Specific table attributes

Titles and footnotes are covered in Section 3.2 Data presentation: common attributes.

4.3.1 Headers

Table columns include headers identifying the content of the data presented in the column cells. Each column header defines an indicator and its mode of expression (e.g. '%' or 'Cases per 100 000').

Columns may consist of the juxtaposition of different indicators such as 'Number of cases', 'Number of deaths', 'Case-fatality'. There is no need for a 'Total' column in such cases. However, when columns represent the value of a single indicator for a second classification variable, they usually require a total column to the right of the table. Some tables may include several indicators presented for different values of a classification variable, as in Table 5.

Rows may require headers when they represent different entities, as in Table 5, or a distribution according to a classification variable, as in Table 5. However, if they represent several occurrences of an entity as in a line-listing, they might not require a row header, as in Table 5.

The background colour for headers in tables is the ECDC official green RGB 105-174-35 and the text is always white bold.

4.3.2 Table borders

Tables should not have vertical borders. In fact, if the tables are well aligned, with text aligned to the left and numbers aligned to the right, the table does not need vertical borders. The header is bordered at the top and bottom by horizontal lines. When designing a table, keep in mind that it should help the reader to interpret the data. Avoid large gaps between columns because these make it difficult to read along a row.

It is sometimes difficult to make complex tables easily readable without vertical lines, particularly if the tables contain a very large amount of data or if the header contains both main and sub categories. In these specific cases vertical lines can be used to improve the readability of the table.

The colour for borders in ECDC tables is grey RGB 191-191-191.

4.3.3 Data rows and cells

All cells in a table should be completed. The following conventions are used:

Character	Indication
.	The data are missing
-	Not calculated
0	The data value is zero
UNK	The data are coded as unknown
N.A.	Not applicable

The number of decimal places should be consistent for each variable presented in a table, and the numbers in a column should be aligned according to the decimal point. Percentages are used to show proportions in order to facilitate comparison. For example, 39% is better than 321/815, because it is easier to remember and to compare among groups. If one or two decimal places are presented for percentages, the numbers are no longer easy to remember and compare: 38.76% or even 38.8% is more complex. If the percentage is less than 10 and the figures beyond the decimal point have public health significance, then it may be reasonable to include them.

When rows represent the modalities of a qualitative nominal variable, without an intrinsic sorting order, they can be sorted by decreasing magnitude of the indicator of interest. Sorting rows will help direct the reader to the minimum and maximum values.

4.3.4 Alignment

In general:

- Text in first column and regular text in other columns is aligned to the left
- Pure numbers are aligned to the right to easily compare the values.
- Very short and simple texts (e.g. acronyms) easily comparable texts (e.g. male/female), mix of numbers (e.g. cases + percentage in brackets, confidence intervals), very simple data in simple tables, can sometimes be centred to improve readability.

Keep each set of numbers in its own column with its own alignment. Keep statistics on a single line for comparisons and avoiding inserted lines. The reader can simply move across the row headers to find the corresponding values.

The vertical alignment of the text in high cells is centred.

5. Graphs

5.1 Purpose and usage

A graph is a visual representation for describing, or summarising data. Graphs display patterns in the data using analogies with shapes and lines. This makes it possible to recognise relationships among the values presented according to different classifications.

Graphs represent one quantitative variable (referred to as the indicator and expressed as a count, proportion, rate, ratio or any other measurement) in relation to one or more other variables that can be quantitative, qualitative ordinal or qualitative nominal. When two classification variables are presented, we refer to primary and secondary classifications. The type of indicator variable determines the preferred method for representing the data graphically.

In contrast to tables that can function as a primary repository for a great deal of information that might point to different patterns, a graph can usually show one main message consisting of several patterns. If graphs are clear, the reader will easily see these patterns and the messages.

5.2 Types of graphs

There are many possible graphical representations and the right choice can be difficult. To produce a good graph it is important to understand why the graph is needed. Is it a relationship between two variables, a distribution or a comparison?

A summary of the types of graphs is provided below. Although this is not a comprehensive list, it includes some of the most common graphs and situations. More detailed information about specific graphs used in epidemiology appears below.

5.2.1 Histograms

Histograms are the preferred representation for a quantitative variable expressed as a count, classified by a quantitative variable. Figure 4 presents a distribution by age class with the same intervals. Figure 5 provides an example of a histogram used to display a count of cases by days of the week, which can be considered as a quantitative classification variable. Figure 18 is another example of a histogram.

Figure 4. Distribution of confirmed cases of MERS-CoV by age group worldwide, March 2012–December 2013

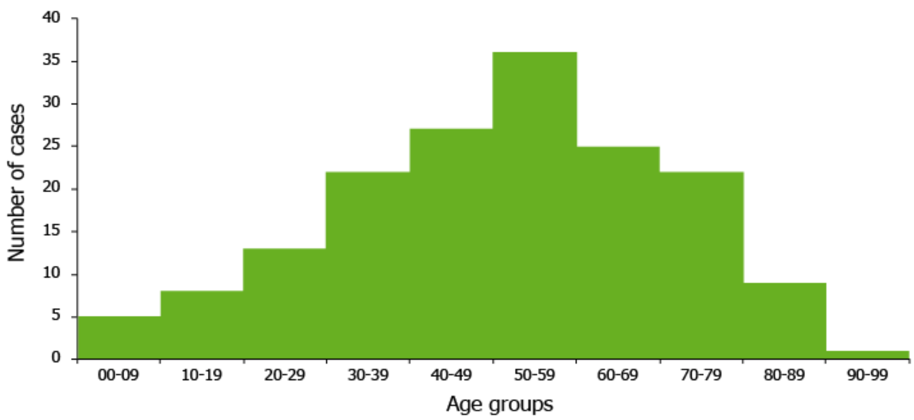
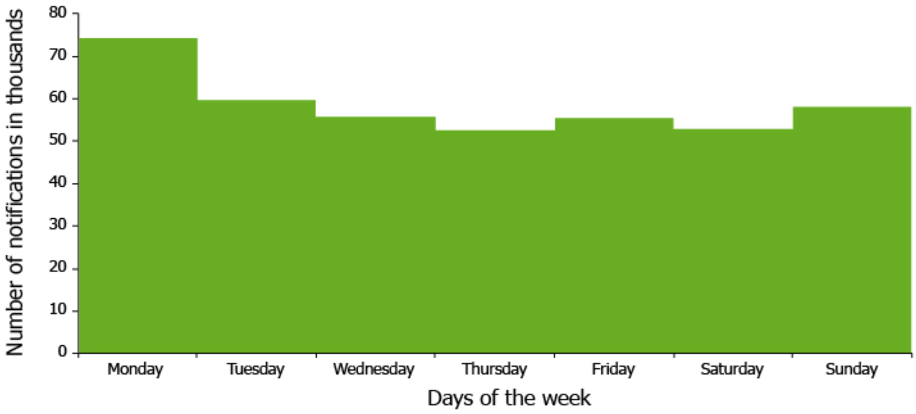


Figure 5. Distribution of notification of *Salmonella* by weekday of onset, EU/EEA countries, January 2007 - June 2013



One property of a histogram is that an area of the same magnitude represents the same number of occurrences. Age pyramids, as presented in Figure 31, are a specific type of histogram. A secondary classification can be represented by colouring parts of the histogram differently, as in Figure 17 where two colours are used to represent MERS-CoV cases infected in and outside of the Middle East.

5.2.2 Frequency polygons

Frequency polygons are used to overlay two or more series of counts. They can be considered as a type of histogram in which the centre parts of the histogram's top bar are joined sequentially. Frequency polygons start and end with 0 cases, closing the polygon, whose appearance and area is then equivalent to a histogram, as in Figure 19.

5.2.3 Line charts

Line charts are mainly used to represent a quantitative variable classified primarily by a quantitative variable expressed in units of time, otherwise known as a time series (see Figures 21 and 22). A secondary classification can be used, resulting in several lines being laid out to represent the trend of the indicator over time for different categories of classification. Figure 26 is an example of such a graph. While there is no general rule, graphs with more than four or five lines tend to become confusing unless the lines are well separated. In a graph with more than one line, different line styles (e.g. solid, dashed) and/or plotting symbols (e.g. asterisks, circles, etc.) should be used. For colour graphs, different coloured lines are the easiest way to differentiate between lines.

Section 5.4 provides more detailed information on the use of line charts for displaying data according to time.

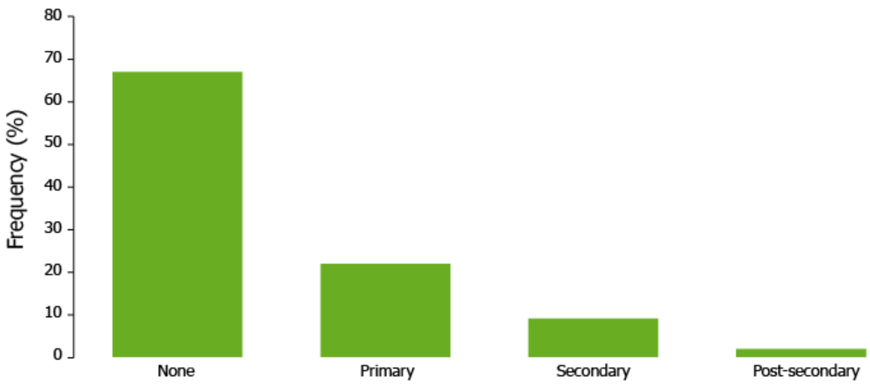
5.2.4 Bar charts

A bar chart is used to compare the number, frequency or other measure (e.g. mean) for different categories or groups. Bar charts are one of the most commonly used types of graph as they are very easy to interpret. The bars can be drawn either vertically or horizontally. The nature of the classification variable guides the choice between horizontal and vertical bars. Vertical bars, from left to right, are well suited whenever the qualitative classification is ordered (from a lower category to a higher one). When the qualitative classification is not ordered, bars are displayed horizontally, ordered from the largest on top to the smallest at the bottom. Bar charts only use one quantitative axis to display the indicator as a bar, either vertically or horizontally, according to the orientation of the bars. The axis containing the modalities of the qualitative classification variable is omitted from such a representation, because it does not convey a mathematical relationship between the modalities.

Vertical bar charts for ordinal classifications

Use a vertical bar graph to show group comparisons in relation to an ordinal classification variable, as in Figure 6 and Figure 9. Categories of the ordinal classifications should be sorted from the lowest on the left to the highest on the right. However, the gradient in the categories is not quantified and therefore the x-axis should not be displayed.

Figure 6. **Vaccine coverage by maternal education, Yamoussoukro, Ivory Coast, 1995**



Horizontal bar charts for nominal classifications

Horizontal bar charts are useful when there is no intrinsic order to the modalities of the qualitative classification variables, as in Figures 7 and 8. Since there is no intrinsic order to the bars representing the modalities, the value of the indicator should be used for sorting. The highest value is plotted at the top and the lowest at the bottom. The vertical axis is not plotted.

Figure 7. **Symptoms reported by cases among wedding attendees, Camberley, UK, 2011**

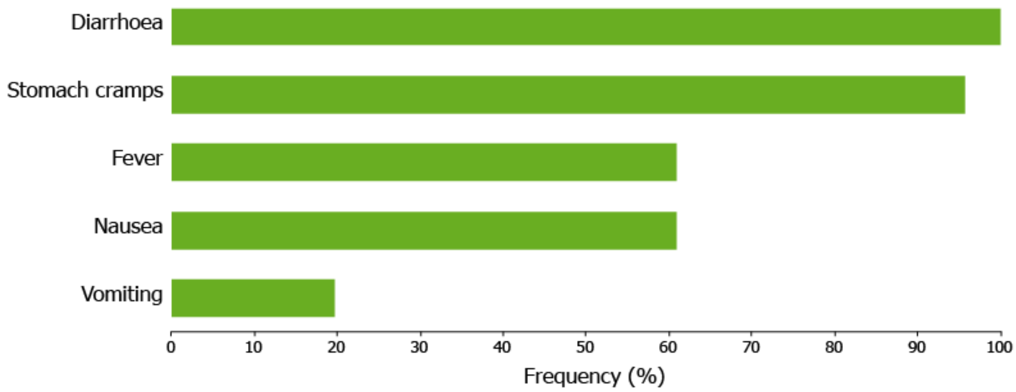
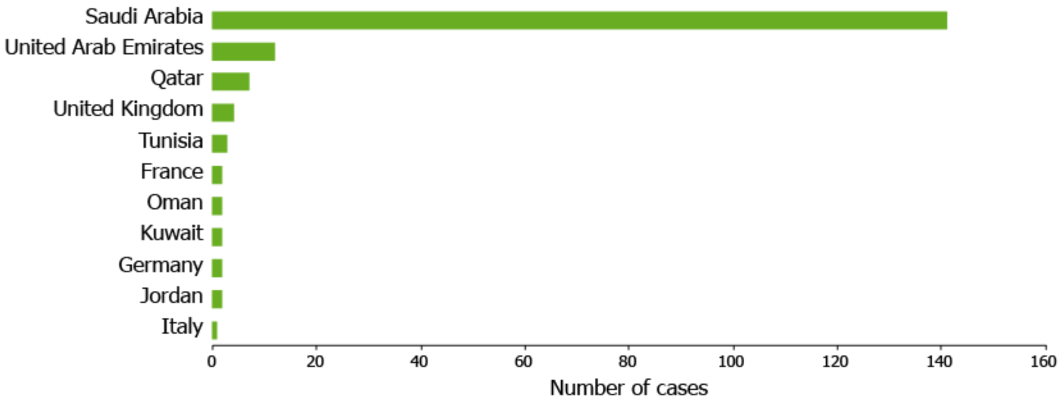


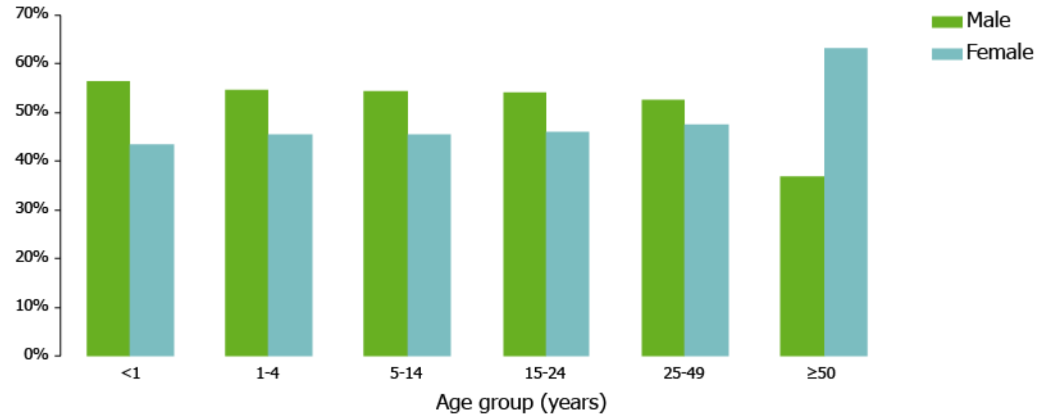
Figure 8. **Distribution of confirmed cases of MERS-CoV by reporting country, March 2012–December 2013**



Second classification variable

When a second classification variable is used, the component of the bars representing each modality of the second classification can be presented as a stack (on top of each other), or as a group (side-to-side).

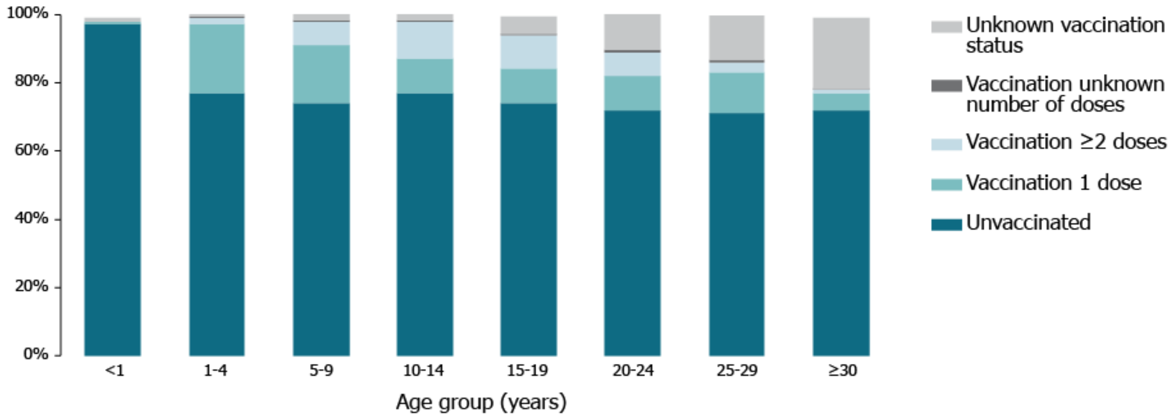
Figure 9. Percentage of confirmed invasive *Haemophilus influenzae* disease cases by age and gender, EU/EEA countries, 2004-2014



Component bar charts

Component bar charts, such as Figure 10, are horizontal or vertical bar charts that include a second classification, with each modality of the second classification presented as a percentage of the total value of the first classification. All the bars representing the first classification are therefore pulled to the same size, enabling a visual comparison of the various components of the secondary classification across the bars. This type of chart is useful for part-to-whole relationships.

Figure 10. Percentage of vaccination status among measles cases by age group, EU/EEA countries, February 2012–January 2013



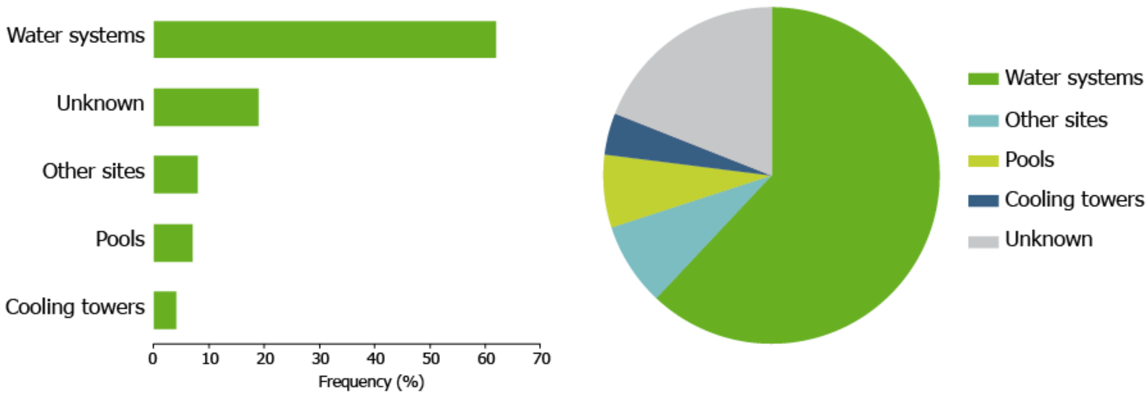
5.2.5 Pie charts

Pie charts can be used to represent the breakdown of a total population into sub-categories according to a nominal qualitative classification variable. The sectors appear sorted in order of magnitude as there is no intrinsic order to the modalities of the nominal classification variable. The colours selected to display the various sectors should use a 'same intensity palette', as presented in Table 3.

Pie charts vs. horizontal bar charts

Pie charts and horizontal bar charts can be used to represent the distribution of an indicator according to the modalities of a nominal classification variable. A pie chart is a good way to display how much of the total each modality represents (e.g. water systems account for almost two thirds of the tested sampling sites in Figure 11), and a horizontal bar chart is preferable for comparing one modality with another (e.g. water systems account for at least three times as many positive samples as any other modality). Both representations are correct but convey a slightly different visual message. Note that both the slices in the pie and the bars in the bar chart are sorted by indicator value.

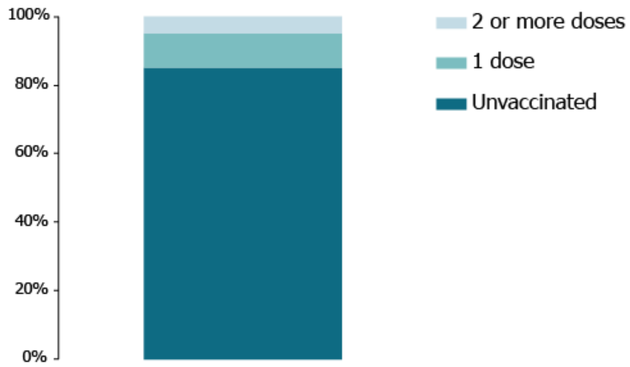
Figure 11. Distribution of sampling sites testing positive for *Legionella*, EU/EEA countries, 2012



5.2.6 Column charts

Column charts represent how a total population can be broken down into sub-categories using an ordinal qualitative classification variable. The order of the sections in the column follows the intrinsic order of the classification variable modalities (as in Figure 12).

Figure 12. Distribution of measles cases by vaccination status, EU/EEA countries, May 2016



Column charts vs. vertical bar charts

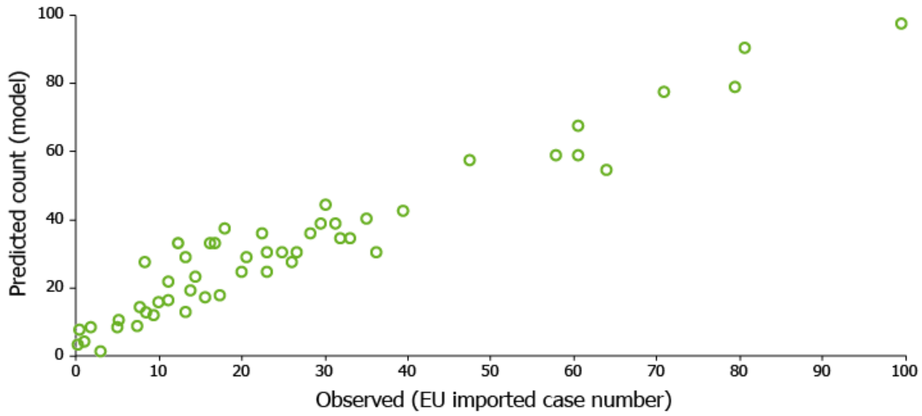
Column charts and vertical bar charts can be used to represent the distribution of an indicator in relation to the modalities of an ordinal classification variable. A column chart is preferable for displaying how much of the total each modality represents, and a vertical bar chart for comparing one modality with another. Both representations are correct, but convey a slightly different visual message.

5.2.7 Other specific graphs

Scatterplots

Scatterplots are a particular type of graph in which two quantitative variables, unrelated in time, are displayed to study their correlation. By default, a dot is plotted at the intersection of the two variable values. The size, shape or colour of the dot can be used to represent a third classification, either qualitative (colour) or quantitative (size). Scatterplots are often represented with a regression line, the slope of which indicates the magnitude of the correlation.

Figure 13. Scatter plot between numbers of imported cases observed and predicted count by the model

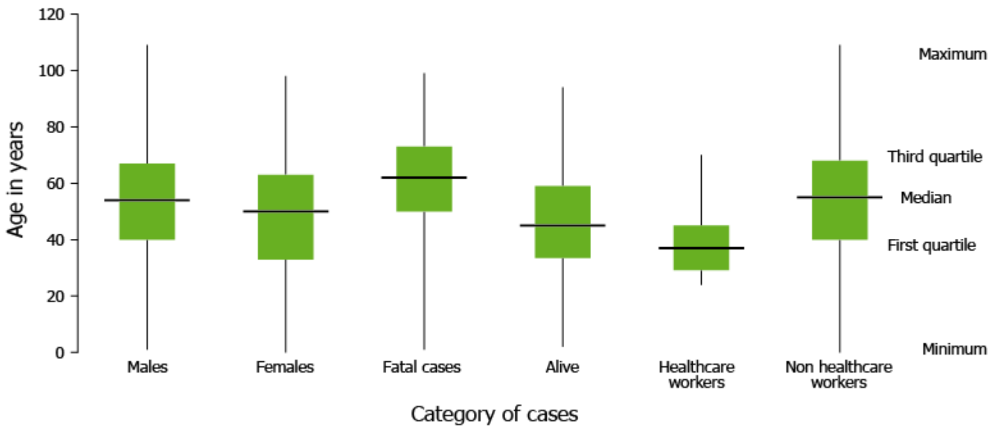


Source: Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International Dispersal of Dengue through Air Travel: Importation Risk for Europe. PLoS Negl Trop Dis 8(12): e3278. 2014, Dec [date cited]. http://dx.doi.org/10.1371/journal.pntd.0003278

Box and whisker

A box and whisker plot (sometimes called a boxplot) is a graphic method of displaying data variation that is ideal for comparing distributions at a glance. Boxplots are named for their design, which includes a box around the lower quartile, median and upper quartile, and two 'whiskers' that extend to the minimum and maximum (Figure 14).

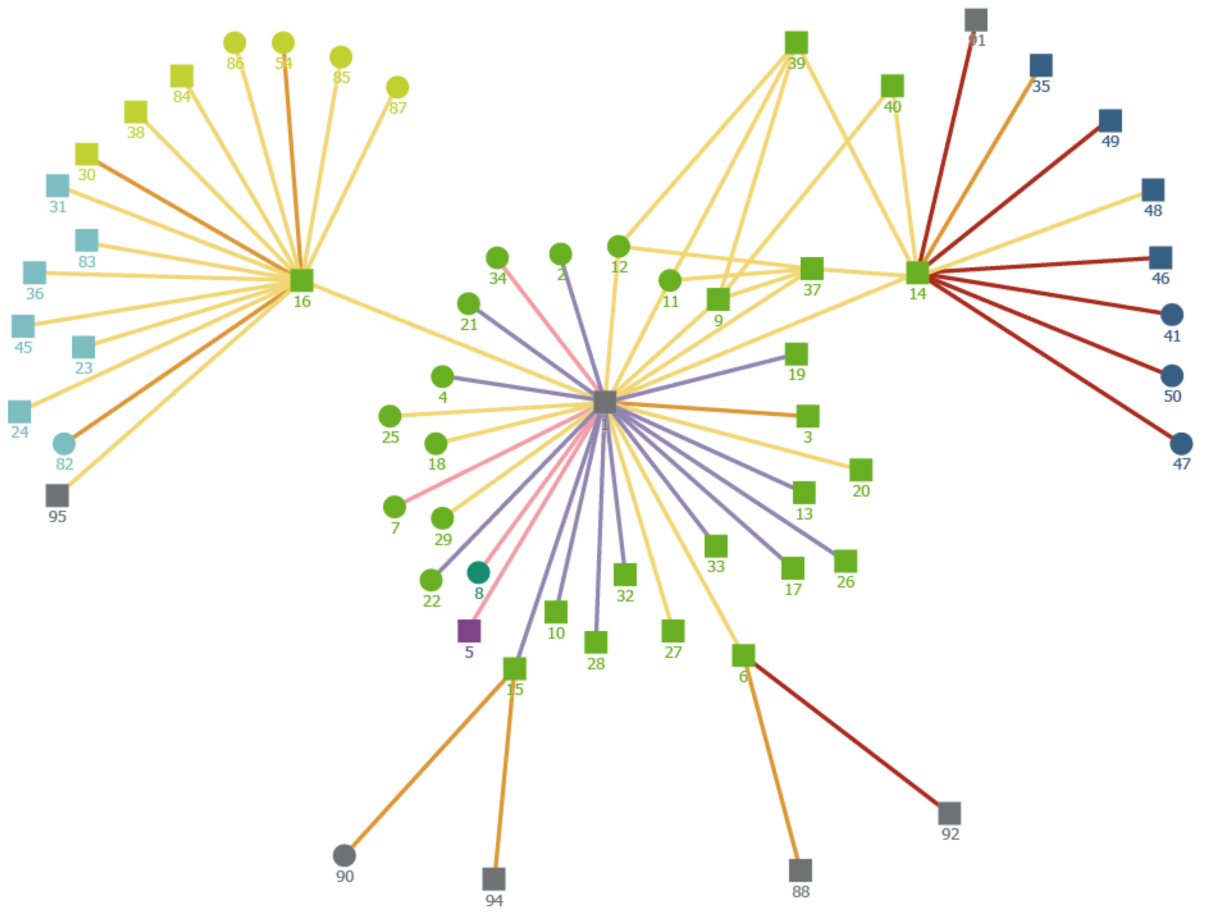
Figure 14. Distribution characteristics of MERS-CoV cases, by case category, worldwide, as of August 2016



Network diagram

Network diagrams are used to show relations consisting of nodes (individuals within the network) and ties (relationships between individuals). Nodes are usually represented as points or other shapes while ties are represented by lines between the nodes. Differences in the shapes or lines of the diagram may be used to represent different characteristics of the individuals or the relationships. Networks analyses in infectious disease control are especially useful in identifying the index or source case and predicting which individuals are more likely to become infected and further infect others.

Figure 15. Network diagram



* Cases with information available on transmission, excluding index case. Source: Ministry of Health, South Korea

Gender
- Male
- Female

Presumed healthcare facility of exposure
- St. Mary's Hospital, Pyeongtaek, Gyeonggi
- Konyang University Hospital, Daejeon
- Konyang University Hospital, Daejeon
- Samsung Medical Centre, Seoul
- Seoul Clinic, Asan, Chungcheongnam
- 365 Yeollin Clinic, Seoul
- Unknown

Presumed mode ox exposure
- Admitted to same emergency room as a case
- Admitted to same room as a case
- Admitted to same ward as a case
- Healthcare worker
- Visiting case in hospital

5.3 Specific graph attributes

Titles and notes for graphs are discussed in Section 3.2.

5.3.1 Legend

A legend is needed to explain a third classification (i.e. one that is not obvious from the x- and y-axes), additional features or anomalies. Labels in the legend should always be mutually exclusive. The placement of the legend should not interfere with the plot area of the representation. The legend is preferably placed in an empty area within the graph and if this is impossible, it should be placed to the right of the graph. There are specific situations for specific graphs when the legend can be put at the bottom of the graph. No frame or background colour is needed. The legend should be omitted when a graph only represents one series of values.

5.3.2 Gridlines

Gridlines are not useful in a graph. They tend to confuse the representation by increasing the 'ink to data ratio'. Gridlines can only be used in a specific situation (i.e. when representing a target or a threshold) to add valuable information to the representation.

5.3.3 Axis

Axes are only shown when they relate to a quantitative variable, because they provide a sense of continuum in the representation. Marks on the axis are displayed on the outer side of the axis in order to avoid interfering with the graph plotting area. Axes must be labelled properly.

An axis related to a quantitative variable should by default start at 0. A histogram or a bar chart axis should always start at 0. However, this rule does not have to apply when plotting a scatterplot.

Where possible, only one y-axis should be used to represent data. The use of two y-axes with different scales of value distorts the representation of the relation in the data series which will no longer be objective.

To ensure good readability and consistency of all ECDC graphs, axes should preferably have a weight of 0.5pt.

5.3.4 Plot area

The plot area should be devoted to presenting data objectively. All unnecessary attributes should therefore be avoided (e.g. frames, grids or colours that increase the 'ink to data' ratio unnecessarily). Data markers on line charts should be avoided as they increase the 'ink to data' ratio unnecessarily. To ensure good readability and consistency of all ECDC graphs, lines should preferably have a weight of 1.5pt.

5.3.5 Data labels

Data labels are unnecessary in the plot area of a graph. The principle of a graph is to allow visual inspection of the data. Adding data labels in the graph's plot area disturbs the visual analysis by mixing digital and analogue modes of conveying information.

5.3.6 Representations in 3D

Volume (3D) is sometimes used to render bar graphs and pie charts more attractive. This should be avoided as it distorts the representation. However, in some instances, a third dimension (classification) can be represented using a three dimensional representation. These are harder to produce or interpret than divided or stacked bar charts and are seldom worth using.

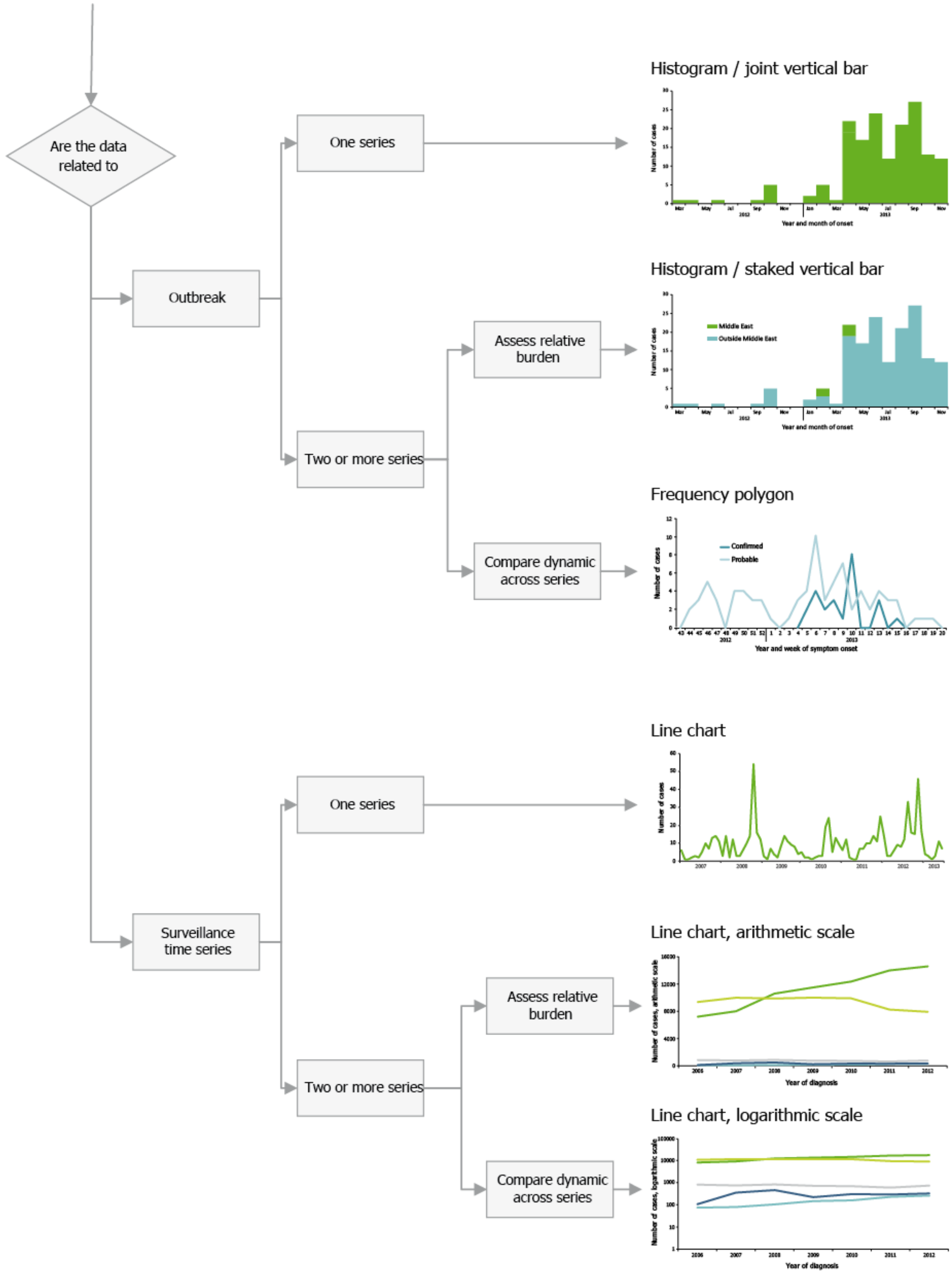
5.4 Graphing time characteristics

Time as a variable plays an important role when you want to create a graph. The algorithm represented in Figure 16 was designed to help select the most appropriate type of graph to represent the distribution of indicators by time in public health surveillance.

There are two possible scenarios, as set out below.

- An outbreak situation represented as an enumeration of cases by time unit. When more than one series is used, the choice between stacked histograms or overlaid frequency polygons depends on whether the representation is to focus on relative burden or to compare the dynamics of the outbreak among different groups.
- A public health surveillance time series represents case counts or rates over time as a line chart. When there is more than one series, an arithmetic scale is used to depict the relative burden across categories, and a semi-logarithmic scale (y-axis only) is used to compare trends over time across the series.

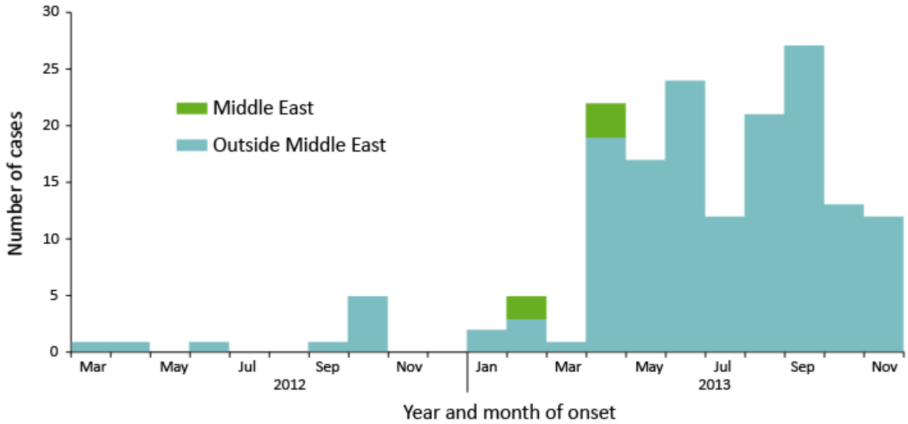
Figure 16. Algorithm for selecting appropriate graph type to represent time data



5.4.1 Epidemic curves

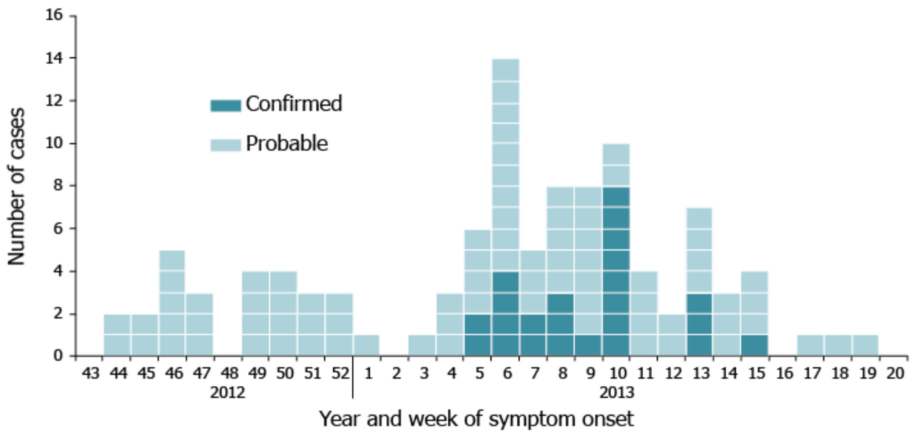
An epidemic curve represents the occurrence of cases by time units during an outbreak. An epidemic curve has a start and an end. When available, the date/time of onset of symptoms is used to represent the data. By default, it is possible to use a fraction of the incubation period as a time unit for aggregation (usually a quarter of the incubation period). When there are few cases per time unit, or when the outbreak extends over a long period (persistent source outbreak or sustained person-to-person transmission), the time unit for aggregation may be chosen differently, for example by month as in Figure 17. Representation of the time interval immediately preceding the first case on the x-axis, as for 2012–02 in Figure 17, objectifies the beginning of the outbreak.

Figure 17. Distribution of confirmed MERS-CoV cases by month and place of probable infection, worldwide, March 2012–December 2013



A case plot, as in Figure 18, should be used to represent the epidemic curve of an outbreak with a small number of cases. Each case occurrence is represented by a square box. All boxes corresponding to the same time unit are stacked. No space is allowed between the boxes along the x-axis, unless no cases have occurred in a given time unit. For each box, a fill or line colour, hatching or a symbol could be used to represent an additional classification, such as confirmed or probable cases, cases with or without fatal outcome, symptomatic or asymptomatic cases, primary or secondary cases, etc.

Figure 18. Distribution of probable and confirmed cases of hepatitis A with travel history to Egypt, by confirmation status and week of symptom onset, EU/EEA countries, November 2012–April 2013



Source: Adapted from J Sane and al. Multistate foodborne hepatitis A outbreak among European tourists returning from Egypt–need for reinforced vaccination recommendations. November 2012 to April 2013.

The first case occurrence should be represented (often the primary case), and the last time unit on the x-axis should be the last one for which information is available. Most graphic packages are not suitable for producing case plots. Excel is often used and can give good results if Excel cells are used to represent cases. However, this is subject to error, because each individual case needs to be represented manually.

If there are many cases in an outbreak, it becomes difficult to represent each single case as a box. Histograms or joint vertical bars can be used, as in Figure 17. However, the same rules apply to these representations. Stacked joint bars can be used to represent an additional classification.

As a rule of thumb, epidemic curves are used to represent cases. Coloured areas of similar size should represent the same number of cases.

Construction of a frequency polygon from a histogram

Sometimes a frequency polygon can replace the histogram. A frequency polygon is drawn by using a line that connects the middle of the coordinates at the centre of each interval and the count in each interval (see Figure 19). Frequency polygons are useful for comparing two data sets, as in Figure 20. While a case plot stacks cases with different characteristics (Figure 18) to visually represent the proportion of each category over time, the frequency polygon (Figure 20) overlays cases, making it possible to visually compare values at different times.

Figure 19. Distribution of probable and confirmed cases of hepatitis A infection with travel history to Egypt, by status EU/EEA countries and week of symptom onset, November 2012–April 2013, with frequency polygon

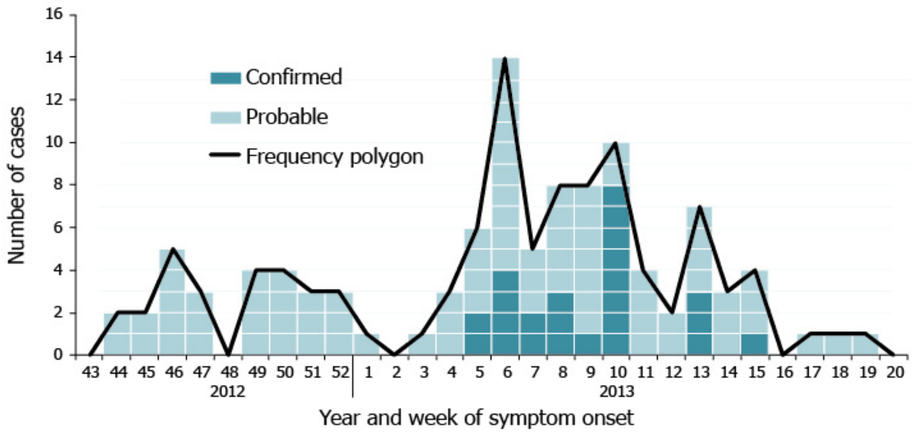
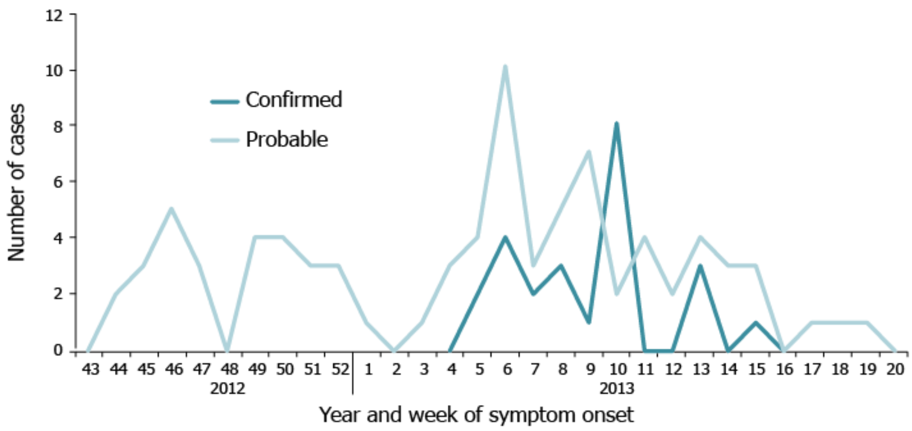


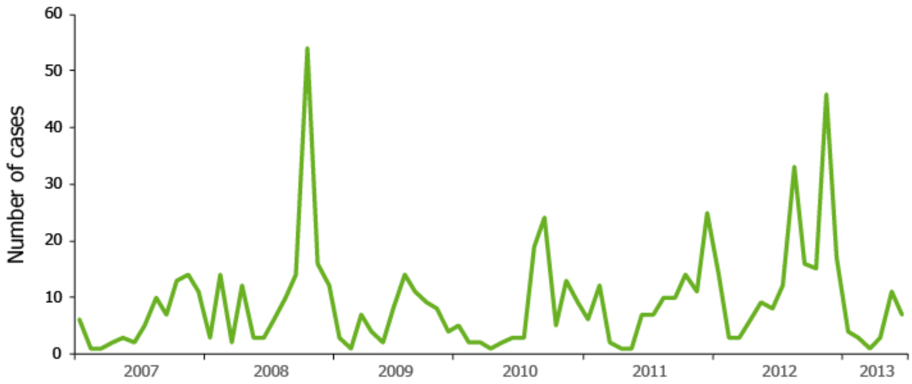
Figure 20. Distribution of probable and confirmed cases of hepatitis A with travel history to Egypt, by confirmation status and week of symptom onset, EU/EEA countries, November 2012–April 2013



5.4.2 Time series

A time series is an ordered sequence of values of a variable at equally spaced time intervals (such as the weekly number of cases of *Salmonella* Mikawasima EU/EEA, 2007–2013). A time series chart is one of the most efficient means of displaying large amounts of data for meaningful analysis. The typical time series line chart is a scatterplot chart, with time represented on the x-axis and lines connecting the data points (Figure 21).

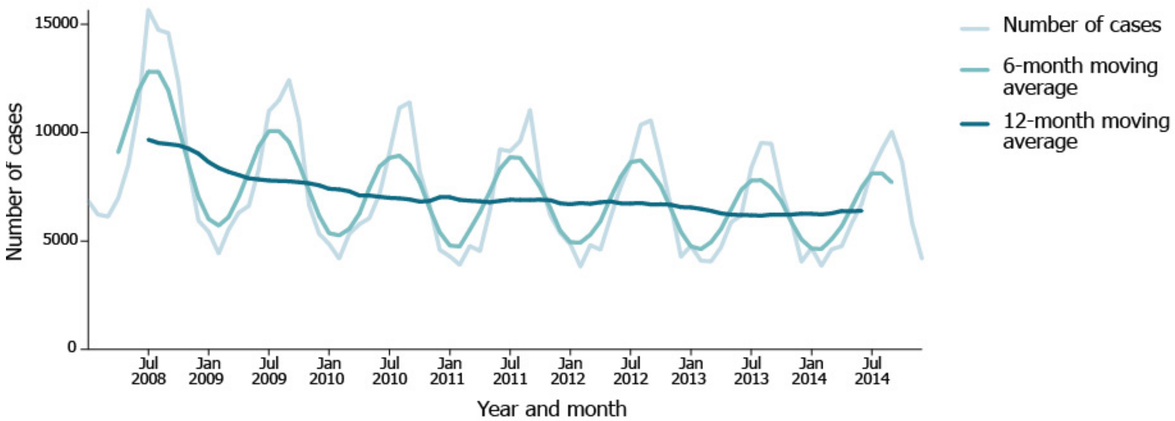
Figure 21. Distribution of *Salmonella* Mikawasima by month, EU/EEA countries, January 2007–June 2013



One or several time series can be plotted on the representation to reflect additional classification variables (e.g. time series by gender, or age groups). Data are usually aggregated over a certain time period. The choice of time granularity (the time aggregation level) depends on the objective of the representation and the number of cases in the series. Time series in epidemiology are usually represented in weeks, months, quarters or years. Shorter time periods are preferred when the objective is to detect changes in the epidemiological pattern over time (outbreaks). However, when assessing the long-term trend of a disease, yearly data are sufficient. A granularity that is too fine may result in only one or no case per time unit being displayed, preventing easy identification of underlying patterns in the data. A line graph is the preferred method for representing a time series. When plotting a time series, it is important that the reader can clearly distinguish the lines for separate data series and is also aware of scaling effects.

When the variation in the number of cases is high, smoothing techniques can be used to highlight an underlying pattern, as in Figure 22. This may be particularly effective for showing the underlying secular trend in a time series (12-month moving average), or the magnitude of seasonal variations (6-month moving average).

Figure 22. Distribution of confirmed salmonellosis cases by month, EU/EEA countries, 2008–2014

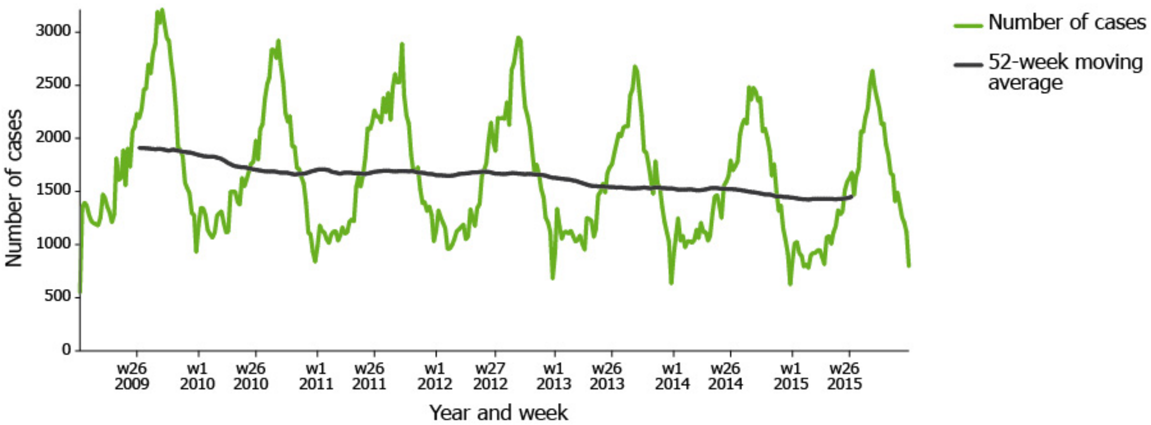


The moving average is a simple method for smoothing out a time series. In statistics, a moving average is an algorithm that calculates the mean of n observations. The parameter n is commonly named 'window size', because the algorithm can be seen as a window that slides over the observations. In the analysis of the time component, moving averages are useful in order to see:

- trends: long-term pattern or direction of the time series, often called secular trend;
- seasonal variation: cycles that occur over short periods of time, usually less than one year, often called seasonality.

The moving average method is simple to understand and easy to implement, and gives a correct picture of the long-term trend (Figure 23).

Figure 23. Distribution of the notifications of salmonellosis by week used for statistics, EU/EEA countries, weeks 2009-05 to 2015-52

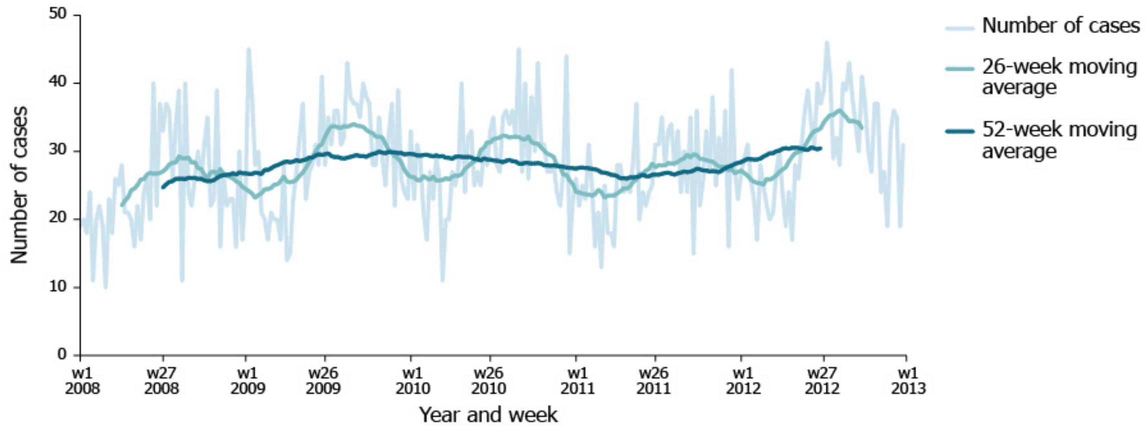


The selection of the window size depends on the objective of the smoothing-out, the amount of variability (noise) in the data and the unit set for aggregating the data.

- When the objective is to explore and highlight seasonality, it is important to smooth out the short-term variation in the data by applying a narrow window. The size of the window is increased until the short-term variation has been smoothed out appropriately.
- When the objective is to explore and highlight a trend, it is important to smooth out the noise and possible seasonality by setting a window size that encompasses an entire year. Therefore, a window of 52 units is used for data represented on a weekly basis, or a window of 12 units for data represented monthly (Figure 24).

Figure 24 illustrates an example using weeks as the time unit. The moving average with a 52-week window indicates that there is an apparent non-linear downward trend in the observed data and a 26-week window highlights the annual seasonal pattern.

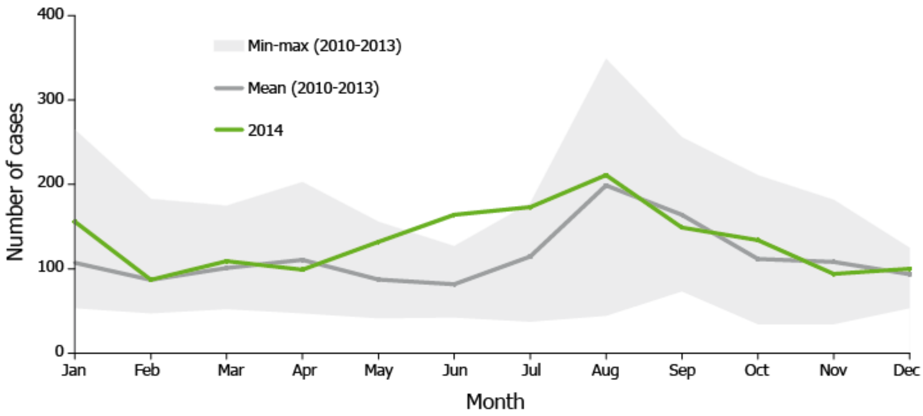
Figure 24. Distribution of listeriosis notifications by week used for statistics, EU/EEA countries, weeks 2008-05 to 2012-52



5.4.3 Seasonality

Seasonality expressed as counts of cases can be represented in a histogram or line chart, because months are considered to be a quantitative classification variable. When the indicator is expressed as a rate, seasonality needs to be represented as a line chart. In Figure 25, additional indicators are presented to allow comparison with the average for past years as well as the range, expressed as an area graph.

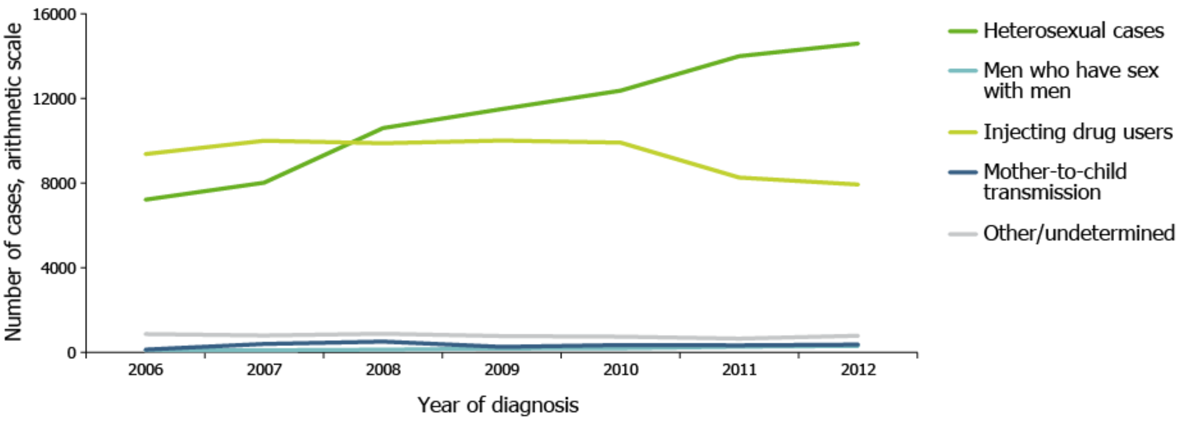
Figure 25. Distribution of reported dengue cases by month in 2014 compared with 2010–2013, EU/EEA countries



5.4.4 Comparing time series

If a secondary classification variable is used several series can be represented in a single time series line graph. In Figure 26, HIV diagnoses by year are plotted for different modes of transmission using different shades of green and markers to differentiate the modalities of the classification. Note that if colours were available, the use of different colours for each line would be a more effective way to convey the information.

Figure 26. Distribution of reported HIV diagnoses, by transmission mode and year of diagnosis, adjusted for reporting delay, WHO European Region, East, 2006–2012



This representation describes the trend in the different groups by mode of transmission and enables the magnitude of the problem to be compared in these groups.

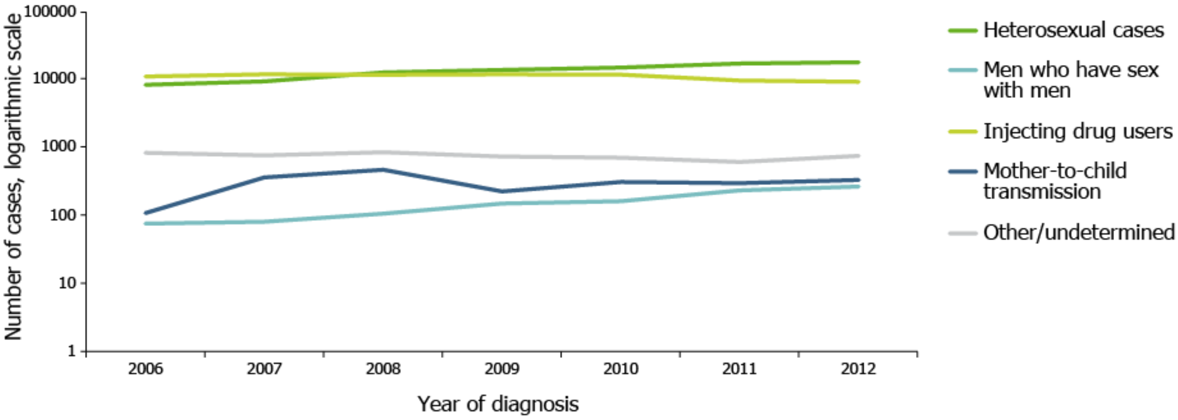
5.4.5 Use of logarithmic scale when comparing trends in time series

By default, an arithmetic scale is used on the y-axis of a time series to represent an indicator. In an arithmetic scale, the same length on the axis represents the same interval in the quantity displayed. Arithmetic scales therefore display the burden of a disease over time.

However, when the objective of the representation is to study a trend at different time intervals, or to compare trends of several series over time, a logarithmic representation on the y-axis is preferable, because it facilitates a comparison of the rates of changes for series of different magnitude. On an arithmetic scale, a 50% decrease from 200 to 100 cases results in a much steeper slope than a 50% decrease from ten to five cases. On a logarithmic scale, the same length on the axis corresponds to the same rate of change in the data. A decrease from 200 to 100 cases and from ten cases to five will therefore result in two parallel lines, indicating that the rate of change over time (i.e. the trend) is similar.

Figure 27 represents the data in Figure 26 on a logarithmic scale for the y-axis. The trend among heterosexual cases in Figure 27 seems to be increasing much more than in the group of men who have sex with men. However, the different magnitude of the values hinders a visual comparison of the trends in the two series. Figure 27 uses a logarithmic scale on the y-axis. Paradoxically, the depiction of the two trends shows a slight convergence of the series over time, indicating that the increasing trend among men who have sex with men is occurring slightly more quickly than among heterosexual persons. Note that the ratio of relative magnitude for each group cannot be estimated visually using a logarithmic scale.

Figure 27. Distribution of reported HIV diagnoses, by transmission mode and year of diagnosis, adjusted for reporting delay, WHO European Region, East, 2006–2012



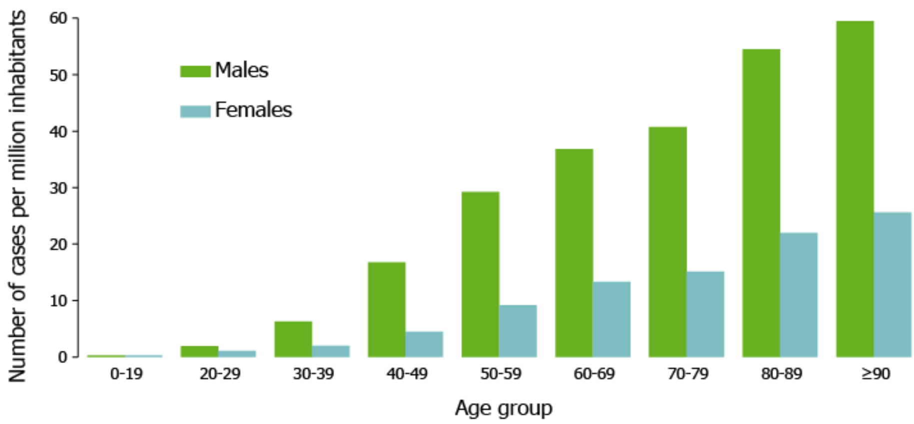
Both representations are correct. An arithmetic scale allows the comparison of absolute values, but not the comparison of trends. A logarithmic scale allows the comparison of trends, but not the comparison of the relative burden. Both representations can be included in a presentation and a report if both the relative magnitudes and the relative trends across series convey important information. In that case, a narrative can highlight for the reader what needs to be seen.

5.5 Graphing age and gender

5.5.1 Age

Age is a quantitative variable. Age may be grouped in age classes of equal or unequal sizes. When grouped in equally-sized age classes, it remains a quantitative variable and can be represented by a histogram, if the case count is represented as shown in Figure 4. Any other indicator, such as rate or proportion, would not be appropriately represented by a histogram, because a similar area under the histogram would not represent the same number of cases. In this case, vertical bars are preferable, as in Figure 28. A line graph should not be used because it would convey the message of a continuum between the different age classes, which might not be the case.

Figure 28. Distribution of reported notifications of Legionnaires' disease per million inhabitants, by gender and age group, EU/EEA countries, 2012

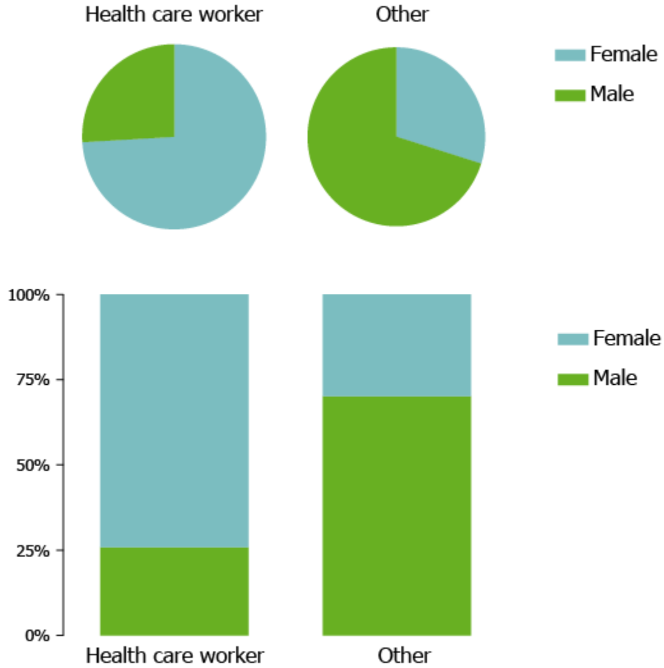


When grouped in an unequal age class, the variable becomes a qualitative ordinal variable. The order is represented along the x-axis, with the lowest age group on the left, and the highest on the right. In such graphs, the x-axis is no longer an axis with a mathematical dimension and therefore should be removed. In this case disjointed vertical bars are used.

5.5.2 Gender

Gender is a qualitative nominal variable with no intrinsic ordering for which pie chart representation may seem appropriate. However, if you have two proportions or distributions to compare, vertical component bar charts are more effective, because they align the proportions for direct comparison (Figure 29).

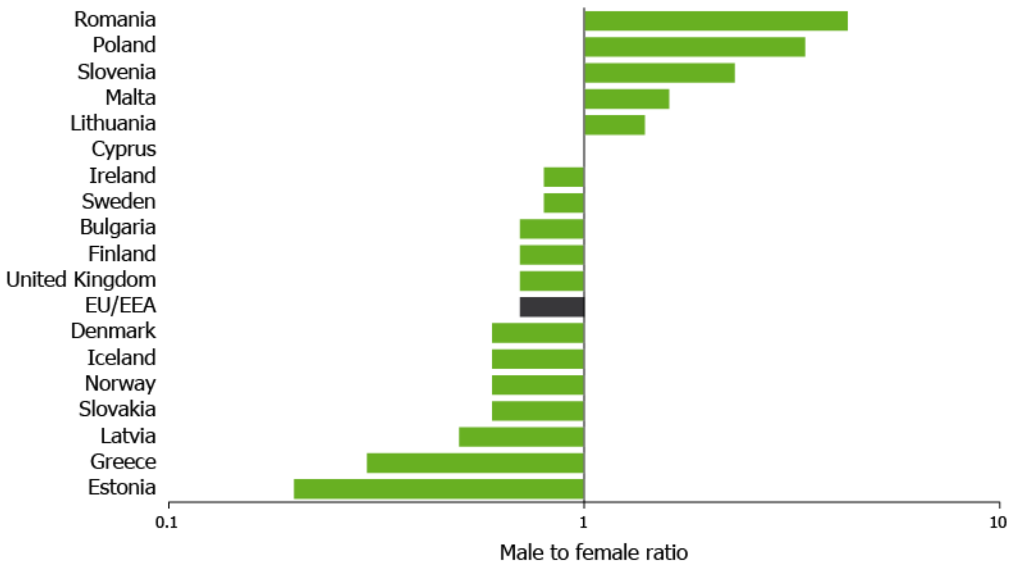
Figure 29. Distribution of confirmed cases of MERS-CoV by gender and healthcare worker status, Jeddah, Saudi Arabia, March 2012–December 2013



The male-to-female ratio is a common descriptive measure in epidemiology, facilitating comparison between the occurrence of the disease in males and females by summarising the difference between the two groups. Bar charts are often used to present male-to-female ratio for different countries (Figure 30) and a horizontal bar chart is preferable. The graph in Figure 30 shows the ratio of males to females for syphilis cases in EU/EEA during the period 2002–2011, by country. The countries should be sorted according to the rank of values so they are easier to compare.

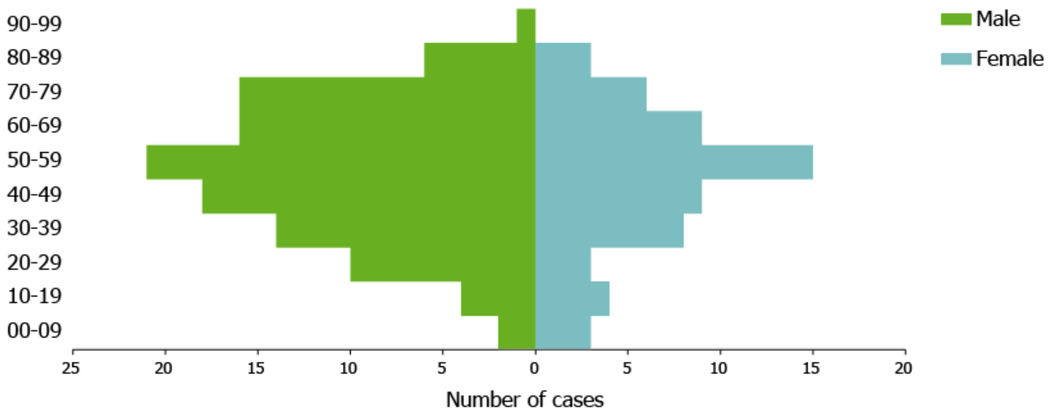
One problem with ratios is their lack of symmetry, therefore the use of the logarithmic scale is recommended. The further towards zero the ratio value is, the larger the visual discrepancies become, and displaying these values on a linear scale is very misleading.

Figure 30. Distribution of male-to-female sex ratio for syphilis cases, by countries with comprehensive surveillance systems reporting consistently, 2002–2011, EU/EEA



To aid understanding, the graph stroke around bars can be used to increase readability.

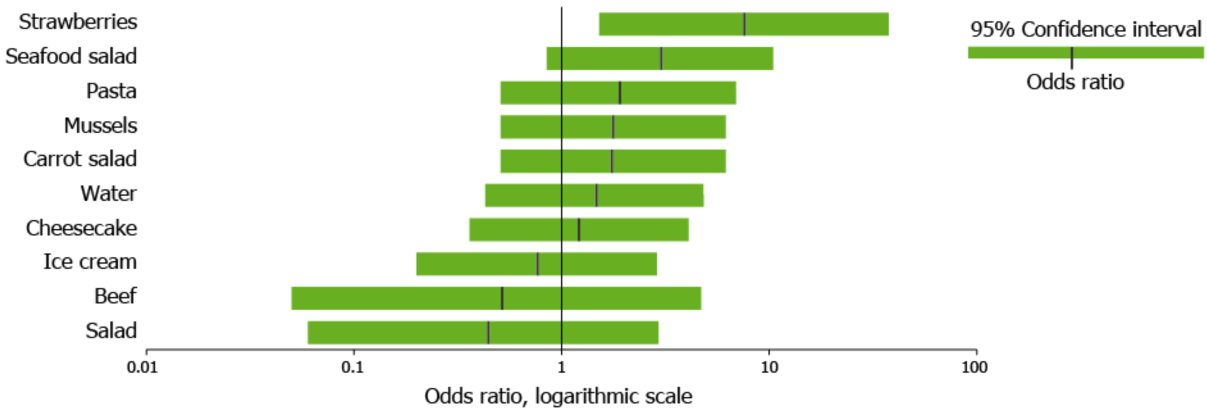
Figure 31. Distribution of confirmed cases of MERS-CoV by month of onset and place of probable infection, Saudi Arabia, March 2012–August 2013



5.6 Graphing measures of associations

Measures of association in an epidemiological investigation are expressed by relative risks, odds ratios and prevalence ratios, depending on the design used to estimate them. A log scale is best suited to the presentation, because there are as many measures of association between zero and one as there are between one and infinity (i.e. an odds ratio of 0.1 is equivalent to 10, 0.01 to 100, etc.) Note that most readers are not familiar with a graphic representation of measures of association. Tables are the best reference method, because they allow for the presentation of other important indicators (e.g. the proportion of exposed cases) that may be relevant for other calculations. However, a graphic representation, such as Figure 32, might be useful for a slide in an oral presentation to convey a specific point (e.g. the degree of uncertainty).

Figure 32. Odds ratio and 95% confidence interval for exposure to selected food item, outbreak of gastroenteritis, place X, time Y



6. Maps

6.1 Purpose and usage

Maps in epidemiology are visual representations of data distribution by geographical unit, providing an overview of geographical patterns. They are mainly used for explanatory purposes. Disease maps can be used to survey high-risk areas, to help policy makers or to decide on resource allocation in specific areas.

Disease maps can present data in several ways:

- area maps apply colour shades or hatching to geographical areas for representation. These maps are referred to as choropleth maps;
- symbol maps use the size or the colour shade of a symbol plotted on the location where it has been calculated;
- chart maps use a graph to present the distribution of a variable, plotted on the location where it has been calculated.

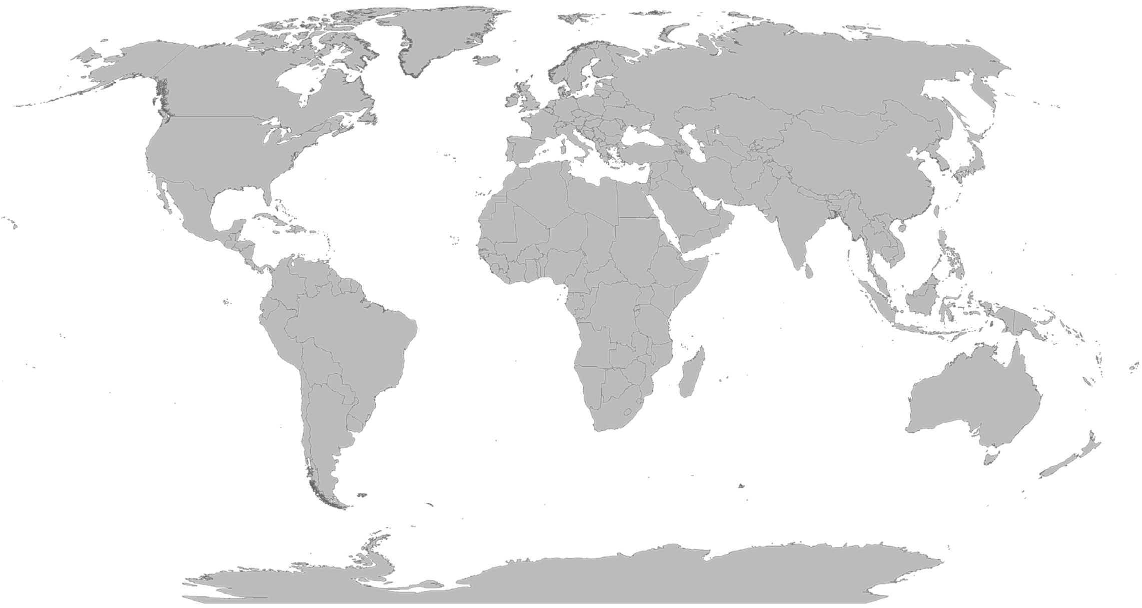
6.2 Considerations for mapping

6.2.1 Projections

Since the earth is a sphere, projections need to be applied in order to represent the countries on a flat map. The following projections are used at ECDC (Figure 33):

- 'Robinson' for mapping the world;
- 'Lambert Azimuthal Equal Area' for EU/EEA countries and the WHO European Region.

Figure 33. Type of data presentation on maps



World: Robinson projection

EU/EEA countries: Lambert Azimuthal Equal Area

WHO Euro Region: Lambert Azimuthal Equal Area

For EU/EEA maps it is important to ensure that Turkey is fully visible as it is a candidate country.

WHO has its own projection to depict its European Region. Use as much as possible their projection but always ask permission from WHO. If this permission cannot be granted for any reason, use the Lambert Azimuthal Equal Area protection.

The recommended projections should be used for producing static maps and, when possible, also in the production of web-based interactive maps.

6.2.2 Geocoding

Geocoding is the process that enables map representation by defining the geographical units used for mapping. ECDC uses two sources for geocoding:

- Nomenclature of Territorial Units for Statistics (NUTS[i]) from EUROSTAT, covering EU/EEA countries;
- Global Administrative Unit Layers (GAUL[ii]) from the Food and Agriculture Organization of the United Nations, covering the world.

NUTS geocoding

NUTS geocoding applies to European Union countries. It is based on a three-level hierarchical classification that subdivides EU Member States into NUTS1, NUTS2 and NUTS3 regions. This classification is a hierarchical system for dividing up the economic territory, aiming to harmonise EU regional statistics. The advantage of the NUTS geocoding is that it aims to ensure that regions are comparable. The disadvantage is that it does not exactly match the administrative boundaries used in the Member States. Table 9 presents the NUTS geocoding scheme according to 2013 classification, however some deviations exist for particular circumstances[iii].

Table 9. The NUTS geocoding scheme

Level	Number of units	Population range
NUTS1	98	3 to 7 million
NUTS2	276	800 000 to 3 million
NUTS3	1342	150 000 to 800 000

GAUL geocoding

GAUL is a global geographical layer containing administrative boundaries for all countries, maintained by the Food and Agriculture Organization of the United Nations (FAO). The GAUL dataset includes three levels of administrative boundaries and units.

Table 10. The GAUL geocoding scheme

Level	Number of units
Level 0	International or country boundaries
Level 1	First-level administrative boundaries
Level 2	Second-level administrative boundaries

By default, ECDC uses the NUTS geocoding scheme to map epidemiological data for the EU and EEA countries. For specific maps, including areas beyond the EU/EEA countries (e.g. for mapping West-Nile-affected areas), ECDC uses a hybrid geocoding scheme, where NUTS codes are used for EU/EEA countries, and GAUL-ADM2 for non-EU/EEA countries.

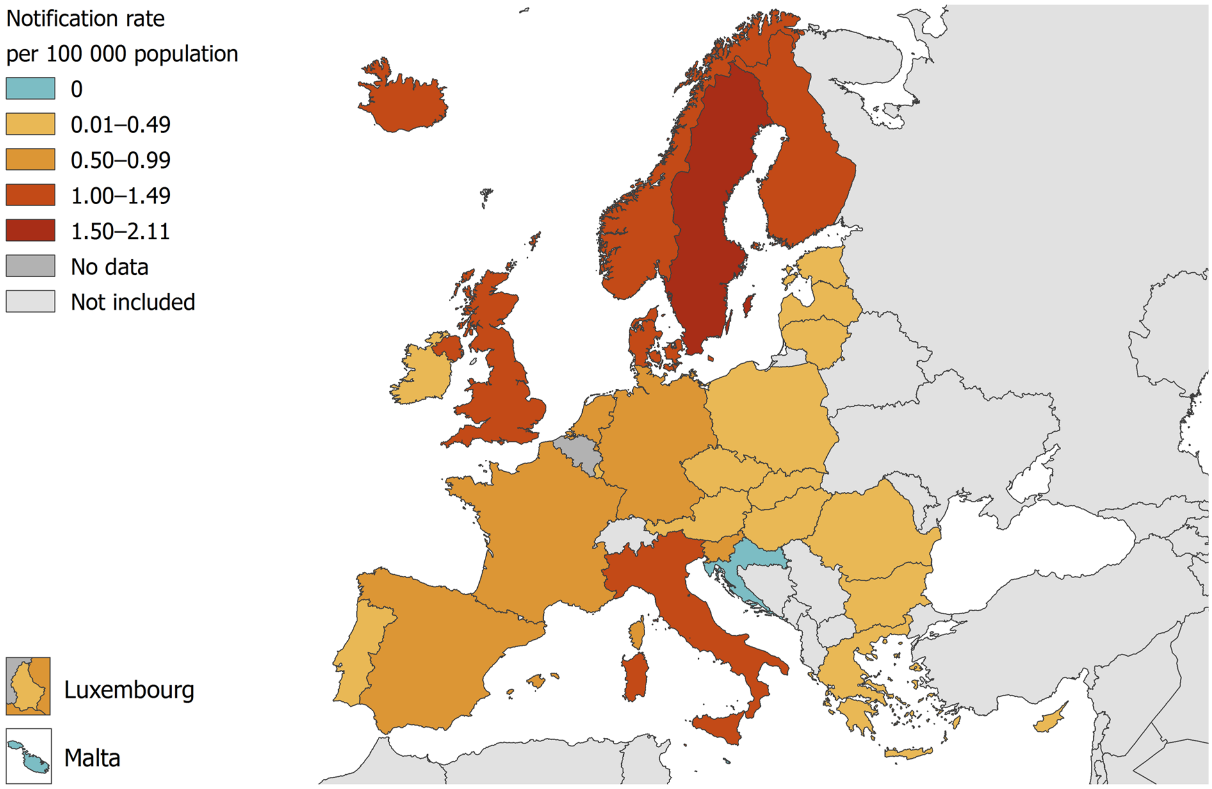
[i] Available at http://ec.europa.eu/eurostat/web/nuts/overview
[ii] Available at http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691
[iii] Available at http://ec.europa.eu/eurostat/documents/3859598/6948381/KS-GQ-14-006-EN-N.pdf/b9ba3339-b121-4775-9991-d88e807628e3

6.3 Types of maps

6.3.1 Choropleth maps

A choropleth map uses a range of colour shades to represent the values of the indicator to be mapped. Indicators expressed as quantities are best represented using a shade gradient of a single colour (Figure 34).

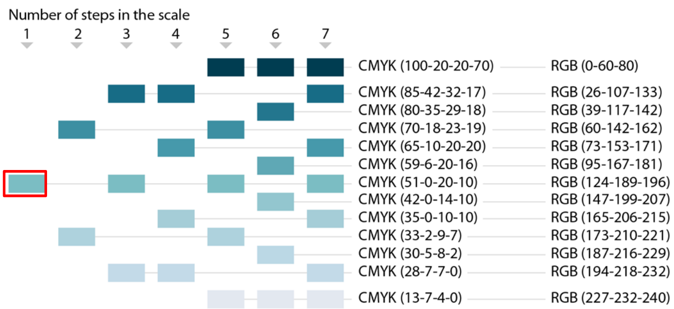
Figure 34. Map displaying notification rate by reporting country, EU/EEA



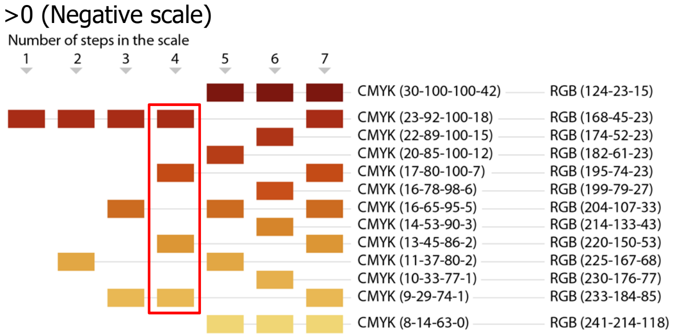
ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Fill colour:
  - 0 (Neutral scale)



  - >0 (Negative scale)



- Base map colours:
  - Fill colour:

	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

Classification of data

Choropleth maps require that data are classified into classes or categories. There are no specific rules for setting the number of classes used in a choropleth map. However, it is generally agreed that between four and eight classes are appropriate. Choropleth maps are more suitable for calculating indicators such as rates, than for enumeration (counts of cases).

The method used to select classes will strongly influence the perception of the geographical distribution of the indicator being mapped. The choice of classes may be specific to the indicator being mapped – for example in relation to a target value, such as reaching 95% vaccine coverage. Otherwise, the method used for categorising data will depend on the distribution of the values to be mapped. Therefore, when deciding on the classification method to use, the first step will be to plot the distribution of values for the indicator being mapped. The shape of the distribution will help determine the most appropriate method (Table 11).

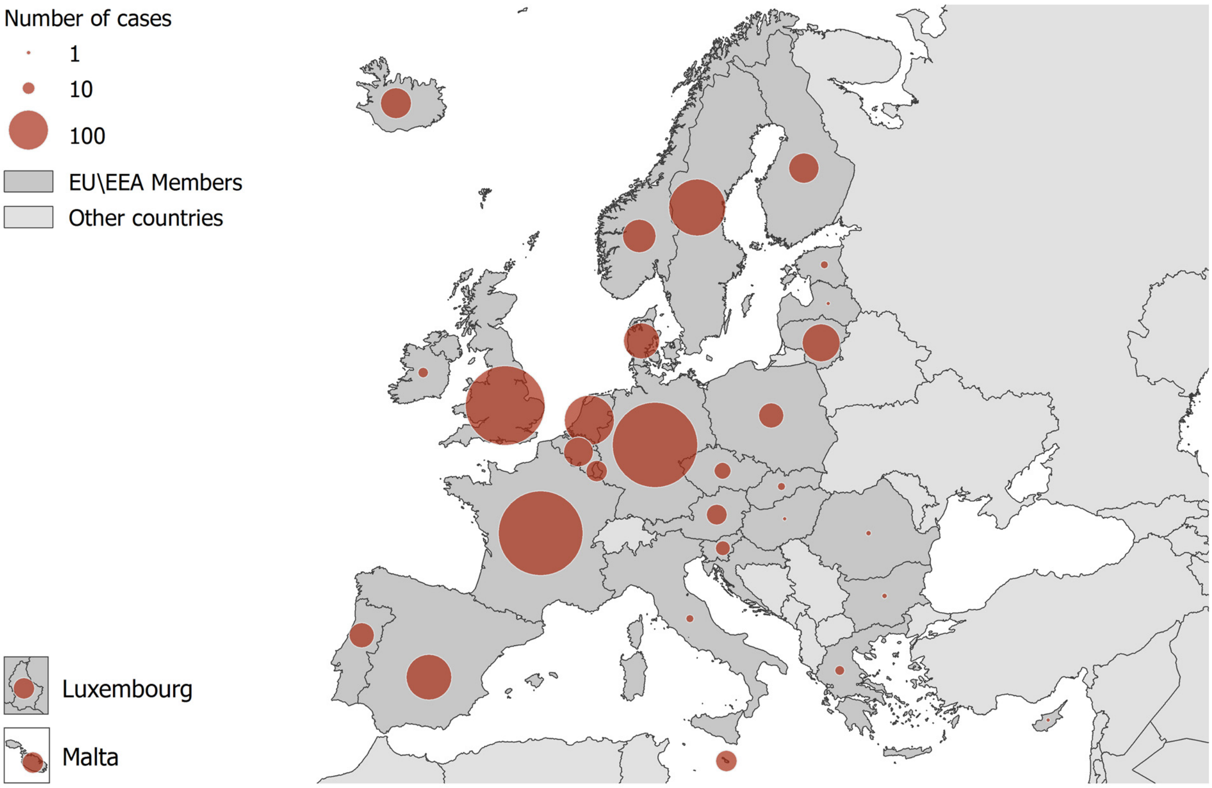
Table 11. Methods for categorising data in a choropleth map

Method	Type	Description	Consideration
Equal intervals	Standard classification	The range of values is split into classes of equal size. For example, five categories for an indicator ranging from 20 to 100 would consist of 20–35, 36–51, 52–67, 68–83, 84–100.	Can be used when values are distributed homogeneously. Otherwise, some classes may not be represented.
Quartiles	Standard classification	The range of values is split in order to allow the same number of geographical units in each class.	Can be used when values are distributed homogeneously. Otherwise, areas with close values may appear in different categories, while some areas with large differences may appear in same category.
Standard deviation	Standard classification	The range is split into intervals expressed in standard deviation of the distribution. For example, five categories may use <-2SD, -2 to <-1, -1 to <1, 1 to <2, >2.	Can be used when values are normally distributed.
Natural breaks	Standard classification	The range of values is split according to natural groupings in the data. The method relies on minimising the variance within groups and maximising the variance across groups.	Can be used when values are heterogeneously distributed. It accurately represents the geographical pattern in the data.
Geometrical interval	Standard classification	It creates class breaks based on class intervals that have a geometrical series. The geometric coefficient in this classifier can change once (to its inverse) to optimize the class ranges. The algorithm creates geometric intervals by minimizing the sum of squares of the number of elements in each class. This ensures that each class range has approximately the same number of values with each class and that the change between intervals is fairly consistent.	This algorithm was specifically designed to accommodate continuous data. It is a compromise method between equal interval, Natural Breaks (Jenks), and quantile. It creates a balance between highlighting changes in the middle values and the extreme values, thereby producing a result that is visually appealing and cartographically comprehensive.
Defined interval	Standard classification	Defined interval allows you to specify an interval size used to define a series of classes with the same value range.	The number of classes will be determined based on the interval size and the range of all field values.
Manual interval	Custom classification	If you want to define your own classes, you can manually add class breaks and set class ranges that are appropriate for your data.	You can start with one of the standard classifications and make adjustments as needed.

6.3.2 Symbol maps

Symbol maps use the size or the colour of one symbol to represent the indicator to be mapped. The most common use of symbol maps is to represent a case count as a circle, the area of which represents the number of cases.

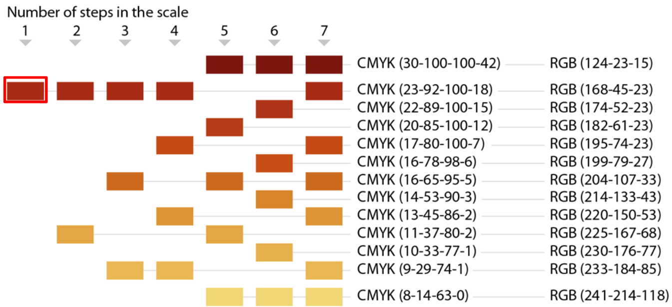
Figure 35. Map displaying number of cases by reporting country, EU/EEA countries



ECDC. Administrative boundaries: ©EuroGeographics

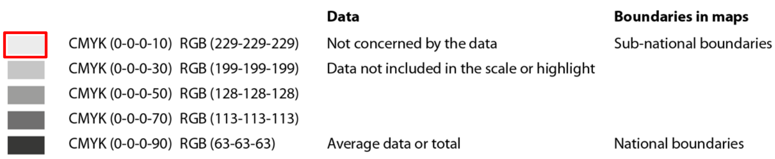
Settings used for the map production:

- Symbol fill colour (Negative scale):



Transparency is recommended when the bubbles overlap polygons borders. In this case, 30% transparency is recommended. However, in general, layers with effect such as transparency cannot be exported as a vector and may be lost, for example when exporting in pdf format.

- Symbol borders colour:



- Base map colours:
  - Fill colour:

		Data	Boundaries in maps
CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50)	RGB (128-128-128)		
CMYK (0-0-0-70)	RGB (113-113-113)		
CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

		Data	Boundaries in maps
CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50)	RGB (128-128-128)		
CMYK (0-0-0-70)	RGB (113-113-113)		
CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

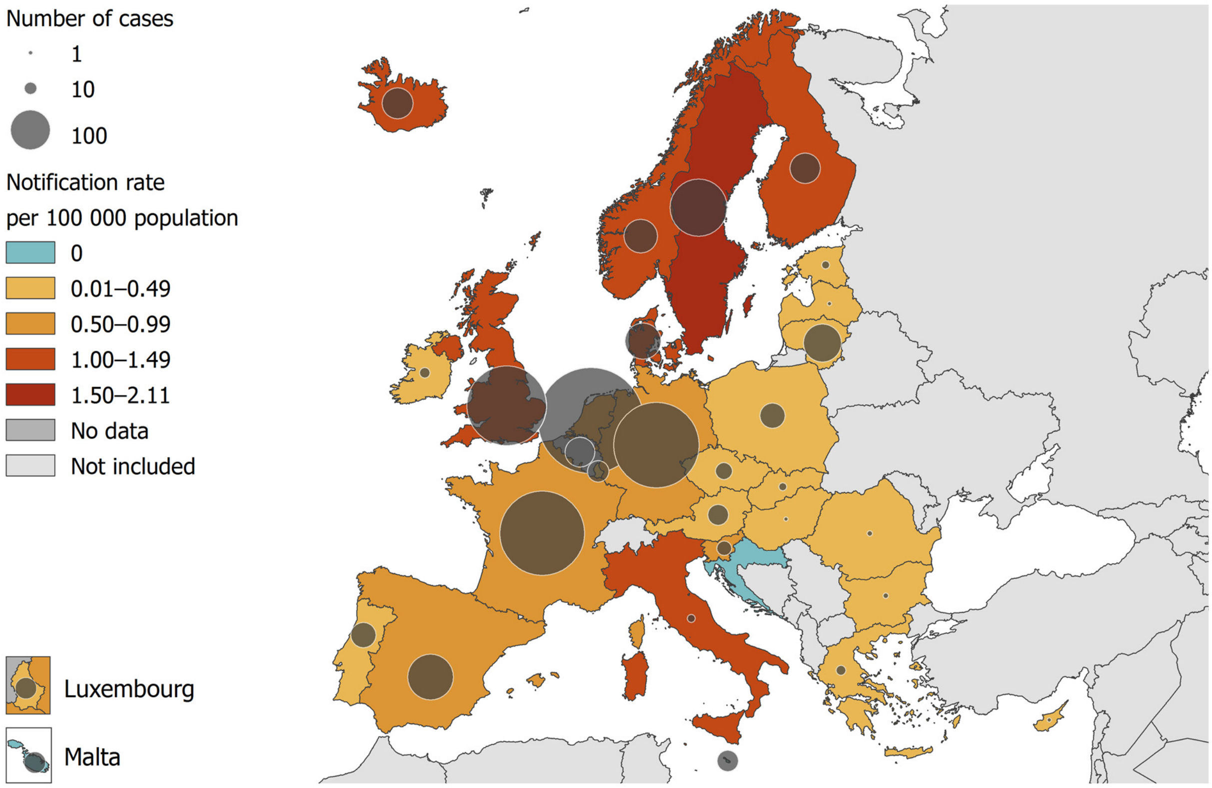
Graduated versus proportional symbols

Symbols should always be proportional to the indicator being mapped. Graduated symbols categorise data into classes assigned to a specific size of symbol. The use of graduated symbols to represent epidemiological data should be avoided as such symbols do not accurately represent the range of values in the dataset.

6.3.3 Multiple layer maps

Different map types can be overlaid to show more than one variable. For example, a choropleth map with the notification rate can be combined with a proportional symbol map with the number of cases (Figure 36).

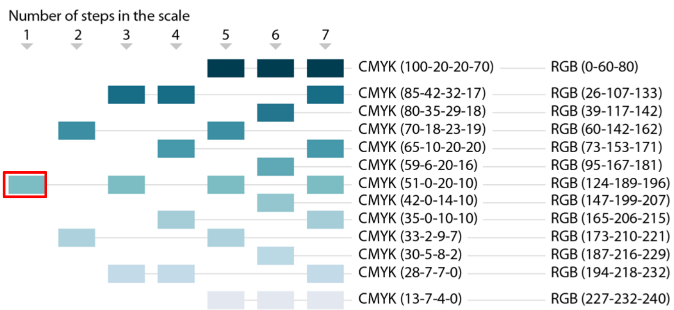
Figure 36. Map displaying number of cases and notification rate by reporting country, EU/EEA



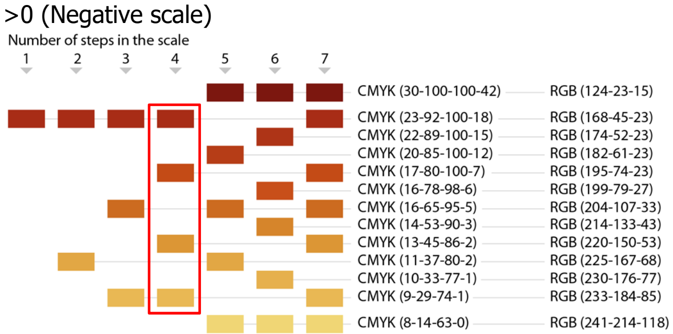
ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Fill colour:
  - 0 (Neutral scale)



  - >0 (Negative scale)



- Proportional symbol fill colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

Transparency: 30%

- Proportional symbol outline colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

- Base map colours:
  - Fill colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

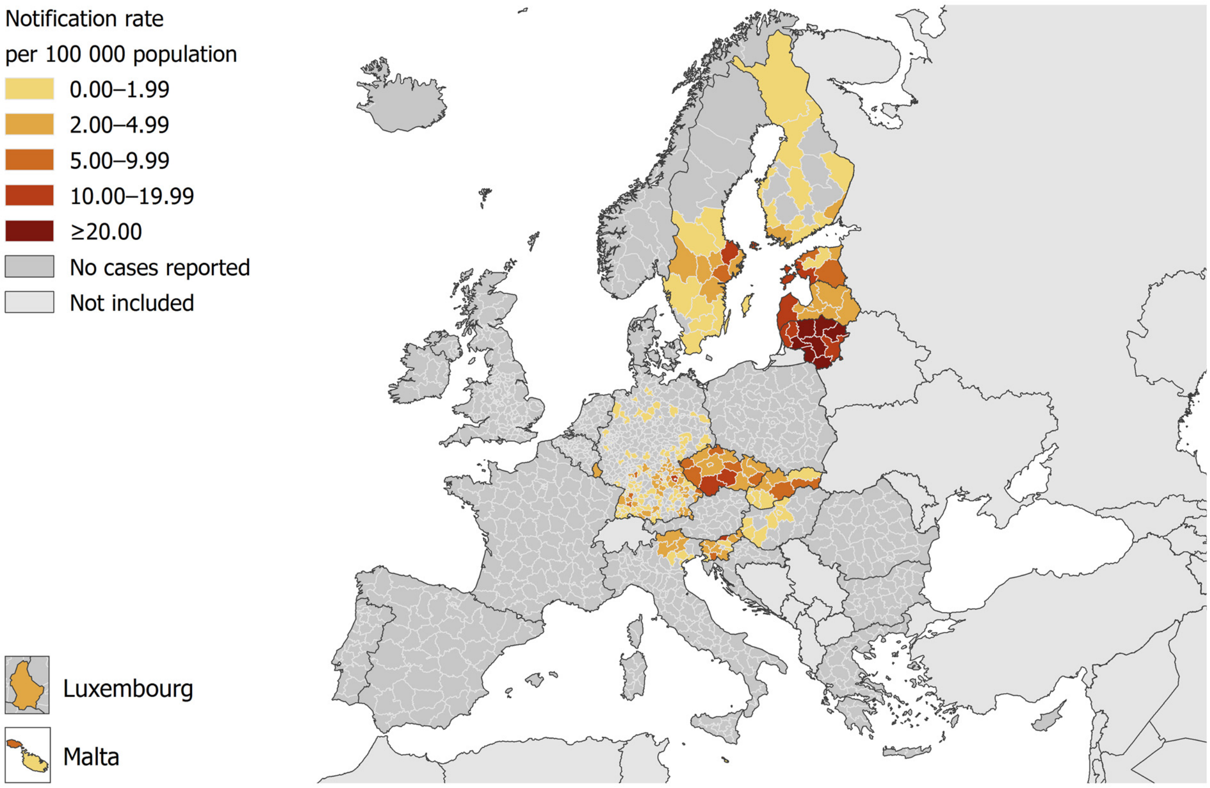
  - Outline colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

6.3.4 Subnational level

These type of maps contain aggregated data using administrative areas lower than the national level (e.g. NUTS3). Figures 37 and 38 show examples of maps produced using NUTS3 level for both types of data: quantitative and qualitative, respectively.

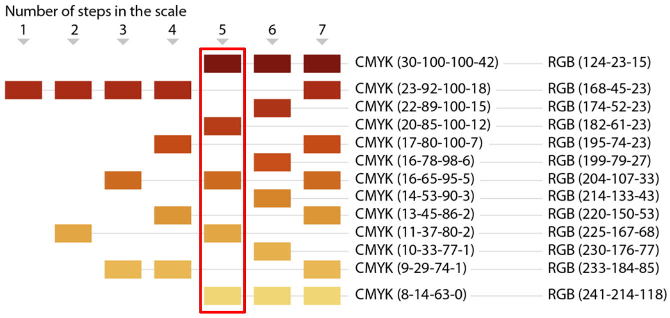
Figure 37. Map displaying notification rate by NUTS3, EU/EEA



ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Choropleth colours (Negative scale):



- Subnational borders colour:

			Data	Boundaries in maps
	CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
	CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
	CMYK (0-0-0-50)	RGB (128-128-128)		
	CMYK (0-0-0-70)	RGB (113-113-113)		
	CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

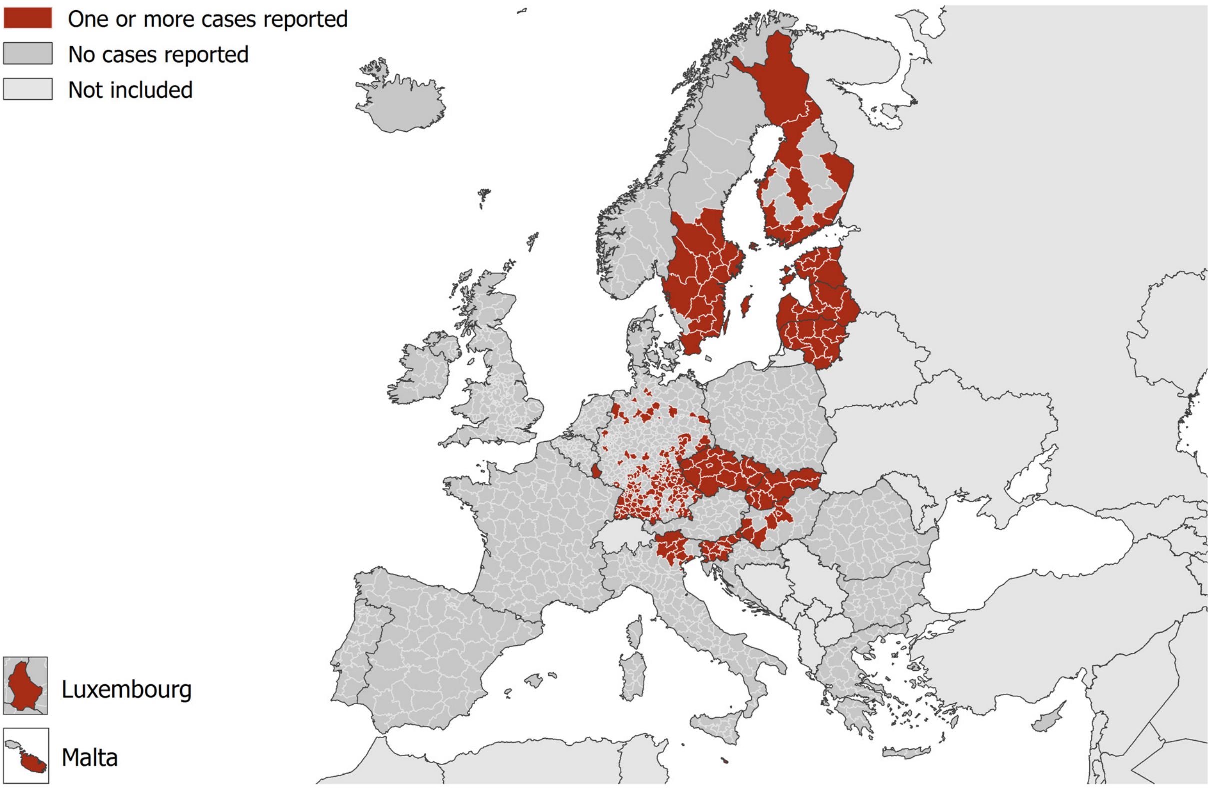
- Base map colours:
  - Fill colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

Figure 38. Map displaying presence of cases by NUTS3, EU/EEA



ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Fill colours (Negative scale):

Number of steps in the scale

Colour	1	2	3	4	5	6	7
CMYK (30-100-100-42) RGB (124-23-15)					■	■	■
CMYK (23-92-100-18) RGB (168-45-23)	■	■	■	■			■
CMYK (22-89-100-15) RGB (174-52-23)						■	
CMYK (20-85-100-12) RGB (182-61-23)					■		
CMYK (17-80-100-7) RGB (195-74-23)				■			■
CMYK (16-78-98-6) RGB (199-79-27)						■	
CMYK (16-65-95-5) RGB (204-107-33)			■		■		■
CMYK (14-53-90-3) RGB (214-133-43)						■	
CMYK (13-45-86-2) RGB (220-150-53)				■			■
CMYK (11-37-80-2) RGB (225-167-68)		■			■		
CMYK (10-33-77-1) RGB (230-176-77)						■	
CMYK (9-29-74-1) RGB (233-184-85)			■	■			■
CMYK (8-14-63-0) RGB (241-214-118)					■	■	■

- Subnational borders colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

- Base map colours:
  - Fill colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

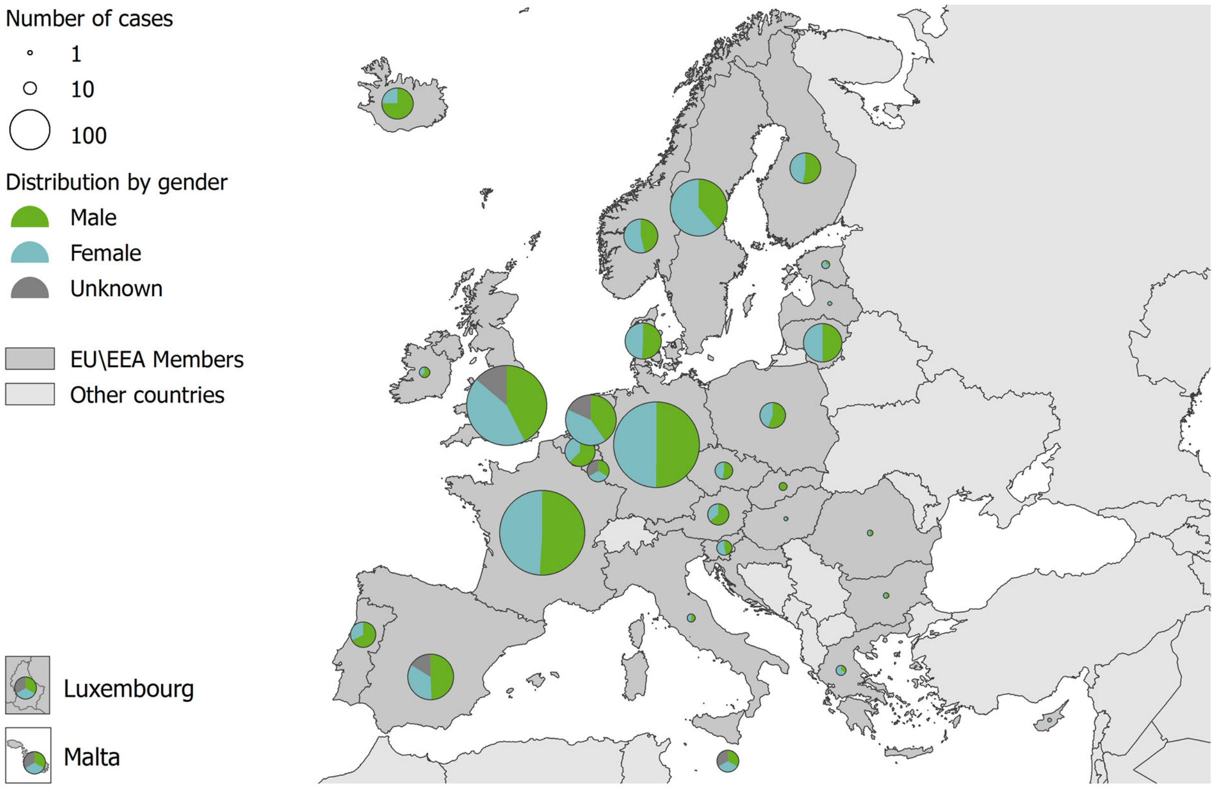
Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

6.3.5 Chart maps

A chart map will overlay geographical areas with small distribution charts to display the different values of an indicator across geographical areas. Pie charts (Figure 39) and bar charts (Figure 40) are usually preferable, although it is possible to use any of the charts described in Chapter 5.

The map in Figure 39 displays a chart map layer using pie charts to represent distribution by gender. Additionally, the overall size of the pie is used as proportional symbol to represent the number of cases.

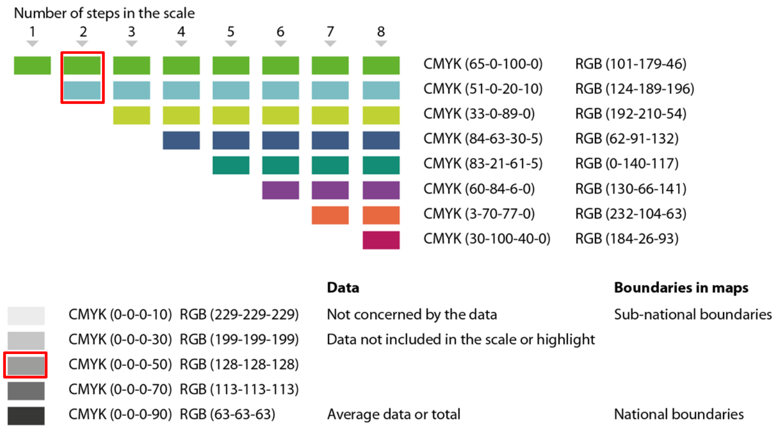
Figure 39. Map displaying gender distribution pie chart sized by number of cases, EU/EEA



ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Pie chart colours:



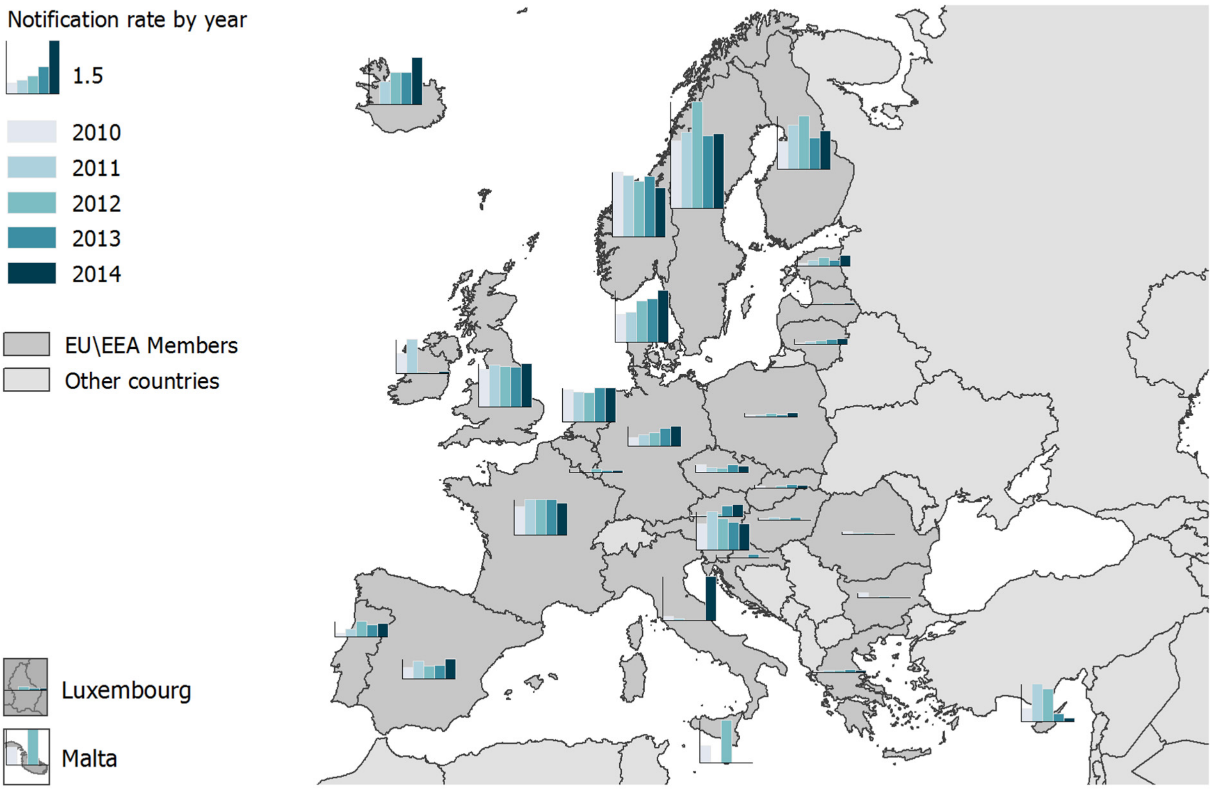
- Base map colours:
  - Fill colour:

Colour		Data	Boundaries in maps
CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50)	RGB (128-128-128)		
CMYK (0-0-0-70)	RGB (113-113-113)		
CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

Colour		Data	Boundaries in maps
CMYK (0-0-0-10)	RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30)	RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50)	RGB (128-128-128)		
CMYK (0-0-0-70)	RGB (113-113-113)		
CMYK (0-0-0-90)	RGB (63-63-63)	Average data or total	National boundaries

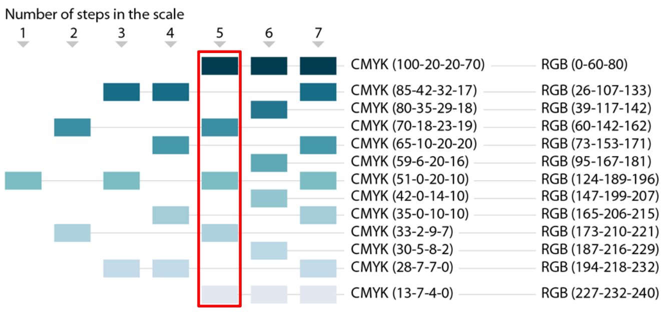
Figure 40. Map displaying notification rate bar chart by year, EU/EEA, 2010-2014



ECDC. Administrative boundaries: ©EuroGeographics

Settings used for the map production:

- Bar chart colours (Neutral scale):



- Base map colours:
  - Fill colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

  - Outline colour:

Colour	Data	Boundaries in maps
CMYK (0-0-0-10) RGB (229-229-229)	Not concerned by the data	Sub-national boundaries
CMYK (0-0-0-30) RGB (199-199-199)	Data not included in the scale or highlight	
CMYK (0-0-0-50) RGB (128-128-128)		
CMYK (0-0-0-70) RGB (113-113-113)		
CMYK (0-0-0-90) RGB (63-63-63)	Average data or total	National boundaries

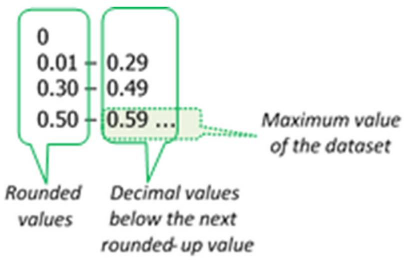
6.4 Specific map attributes

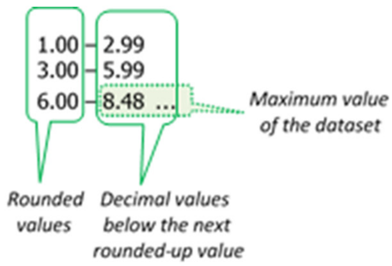
6.4.1 Legend

The legend should be located on the upper left of the map, if the area is focused on the EU Member States, and in the bottom left, if a worldwide map. In specific circumstances, based on the need to display data on the map, the legend can be placed differently.

There should not be any frame around the map, each layer name should be written and the label accompanying each symbol should be written out in ordinary text.

The legend rates should start with rounded values and end with decimal values just below the next rounded-up value. The final value of a legend rate should be the maximum value for the dataset. It is recommended that this value be used instead of the expression 'and (&) above'.





6.4.2 Non-visible countries

Non-visible countries are those that are not clearly visible in the map, such as Luxembourg and Malta for EU Member States. Liechtenstein, an EEA country, is usually not represented on a map. Non-visible countries should always be represented on ECDC maps.

6.4.3 Other cartographic elements

All maps produced at ECDC are in the public domain and can be reproduced freely, provided that the source is acknowledged. The words 'ECDC, 2017' [year of map creation] should appear directly underneath. The logos of tools or other systems used to create the map should not be included.

A scale or an arrow indicating North are not necessary for epidemiological maps, unless in very specific situation such as for mapping a small area in the context of an outbreak investigation.

7. Bibliography

Armitage P, Berry G, Matthews J. Statistical methods in medical research. Oxford: Blackwell Scientific Publications; 2002.

Bell SB, Hoskins RE, Pickle L, Wartenberg D. Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public. BioMed Central Ltd. BioMed Central Ltd, 2006.

Blazes D, Lewis SH. Disease surveillance: technological contributions to global health security. CRC Press; 2016

Bradley N. Essentials of Environmental Public Health Science: A Handbook for Field Professionals. New York: Oxford University Press; 2014.

Cleveland W. The elements of graphing data. Monterey, California: Wadsworth Advanced Books and Software; 1985.

European Commission. Regions in the European Union: Nomenclature of Territorial Units for Statistics NUTS 2013/EU-28. Luxembourg: Publications Office; 2015.

Few S. Show me the numbers: Designing tables and graphs to enlighten. Burlingame, California: Analytics Press; 2012.

Geography and Geospatial Science Working Group. Cartographic Guidelines for Public Health. Atlanta, GA: US Centers for Disease Control and Prevention; 2012

Gordis L. Epidemiology. Fifth Edition. Philadelphia (PA): Elsevier Saunders; 2014.

Itten J. The Elements of Color. John Wiley & Sons, 1970.
https://books.google.se/books?id=ofvRhNBgoCoC&source=gbs_similarbooks

Last J. A Dictionary of Epidemiology. Fifth Edition. New York: Oxford University Press; 2008.

Melnick AL. Introduction to Geographic Information Systems in Public Health. Gaithersburg, MD: Aspen Publishers; 2002.

Miller J. Organizing Data in Tables and Charts: Different Criteria for Different Tasks. Teaching Statistics.2007;29(3):98-101.

Miller J, 2004. The Chicago Guide to Writing about Numbers. The Chicago Guides to Writing, Editing and Publishing. University of Chicago Press, Chapters 5 and 6.

Peat J, Barton B, Elliott E. Statistics Workbook for Evidence-based Health Care. Chichester: Wiley-Backwell; 2008.

Plaza-Rodríguez C, Appel B, Kaesbohrer A, Filter M. Discussing State-of-the-Art Spatial Visualization Techniques Applicable for the Epidemiological Surveillance Data on the Example of *Campylobacter spp*. in Raw Chicken Meat. Zoonoses and Public Health. 2016;63(5):358-69.

Robert W. Statistical Persuasion: How to Collect, Analyze, and Present Data…Accurately, Honestly, and Persuasively. Thousand Oaks, California: SAGE Publications, Inc; 2010.

Rothman K, Greenland S, Lash T. Modern Epidemiology. Third Edition. Philadelphia (PA): Lippincott Williams & Wilkins; 2008.

Silva I. Cancer Epidemiology: Principles and Methods. Lyon, France: International Agency for Research on Cancer; 1999.

Simons NE, Menzies B, Matthews M. A Short Course in Soil and Rock Slope Engineering. London: Thomas Telford Publishing; 2001.

Tufte E. The Visual Display of Quantitative Information. Second Edition. Cheshire, Conn: Graphics Press; 2001.

United Nations. Making data meaningful. New York: United Nations; 2009.

United Nations. Making data meaningful: Part 2. New York: United Nations; 2009

U.S. Department of Health and Human Services. Principles of epidemiology in public health practice: An introduction to applied epidemiology and biostatistics. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2006.

Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Hoboken, NJ: John Wiley & Sons; 2004.

Wallgren A. Graphing statistics & data: Creating better charts. Thousand Oaks, California: SAGE Publications, Inc; 1996

FEM wiki links:

https://wiki.ecdc.europa.eu/fem/w/wiki/choosing-a-method-of-data-display.aspx

https://wiki.ecdc.europa.eu/fem/w/wiki/tables.aspx

https://wiki.ecdc.europa.eu/fem/w/fem/graphs-charts-diagrams.aspx

European Centre for Disease Prevention and Control (ECDC)

Postal address:
Granits väg 8, SE-171 65 Solna, Sweden

Visiting address:
Tomtebodavägen 11A, SE-171 65 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union and the ECDC Independence Policy, ECDC staff members shall not, in the performance of their duties, deal with a matter in which, directly or indirectly, they have any personal interest such as to impair their independence. Declarations of interest must be received from any prospective contractor(s) before any contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Publications Office