MEDICAL UNIVERSITY – PLEVEN
FACULTY OF PUBLIC HEALTH

DEPARTMENT OF PUBLIC HEALTH SCIENCES
CENTRE FOR DISTANT LEARNING

LECTURE No1

INTRODUCTION TO STATISTICS

Assoc. Prof. Gena Grancharova, MD, PhD

Medical Statistics

Academic year 2019/2020

Second year – IV semester

Horarium:15/15

Lecturer:

Assoc. Prof. G. Grancharova, MD, PhD

Tuesday, second week, 12:15 – 13:45

Lecture hall “Galen”

Practical assignments:

Assoc. Prof. Eleonora Mineva

	THEMATIC PLAN OF LECTURES
Lecture 1 17.09.2019 2 hours	**Introduction to Statistics.** Population and sample. Types of study. The research process. **Sources and types of data.** Organising and presenting of data.
Lecture 2 01.10.2019 2 hours	**Distributions.** Normal distribution. Standard scores and standard normal curve. Asymmetric distributions. **Simple descriptive statistics for categorical data.**
Lecture 3 15.10.2019 2 hours	**Descriptive statistics for quantitative data. Measures of central tendency.** Characteristics and comparison of the mean, mode, median. **Quantiles.** Characteristics and determination of different quantiles. **Percentiles.**
Lecture 4 29.10.2019 2 hours	**Descriptive statistics for quantitative data. Measures of spread.** Characteristics of the range, interquartile range, standard deviation, variance. The concept of "norms" and "normal groups' limits".
Lecture 5 12.11.2019 2 hours	**Correlation.** Types of correlation. Characteristics of the correlation coefficients. Pearson's and Spearman's correlation coefficients. Correlation and causation. **Regression.**
Lecture 6 26.11.2019 2 hours	**Inferential statistics. Statistical estimation: from sample to population.** Point and interval estimation. Basic concepts: standard error, probability, probability coefficient, degree of freedom, confidence interval. Basic steps in rates' and proportions' estimation.
Lecture 7 10.12.2019 3 hours	**Inferential statistics. Hypothesis testing.** Basic concepts: null and alternative hypotheses, directional and non-directional hypotheses, statistical significance, types I and II errors. Basic steps in hypothesis testing. **Parametric tests** – one-sided and two-sided t-test. **Inferential statistics. Hypothesis testing. Nonparametric tests -** chi-square ($\chi 2$). Summarisation

Tuesday 12:15-13:45 – according to the above schedule
Lecturer: Associate Prof. Gena Grancharova, MD, PhD

Lecture plan

1. Definition and major objectives of statistics
2. Basic concepts
3. Types of studies and research process
4. Sampling and sample types
5. Classification of variables
6. Organizing and presenting the data
6.1. Table presentation
6.2. Graphical presentation

1. Definition and major objectives of Statistics

Statistics is the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data, and that, by use of mathematical theories of probability, imposes order and regularity on aggregates of more or less disparate elements.

According to Webster's Dictionary "STATISTICS" means:

1. facts or data of a numerical kind, assembled, classified and tabulated so as to present significant information about a given subject.
2. the science of assembling, classifying and tabulating such facts or data.

But it is more than this. Statisticians and researchers should analyse data in order to make generalisations and decisions.

Therefore, **statistics is** the science of collecting, summarizing, presenting and interpreting data, and of using them to estimate the magnitude of associations and to test hypotheses.

In short, **statistics is** a science that deals with **mass events**, e.g. events that are characterized by many cases.

In medicine, there are many problems expressed by mass events, e.g. diseases, births, deaths, disabilities, admissions in and discharges from hospitals, etc.

For this reason, statistics is of utmost importance in medicine.

Two kinds of statistics

1. Descriptive statistics

2. Inferential statistics

STATISTICAL ACTIVITIES

- **Statistical description – the process of summarizing the characteristics of data under study (at the sample or population level). This process is called descriptive statistics.**
- **Statistical relationship analysis - the process of analysis of relationship between dependent (effect) and one or more independent (causes) variables.**

STATISTICAL ACTIVITIES

- **Statistical inference – the process of generalization from a sample to a population, when the observation is performed in a representative sample, usually with calculated degrees of uncertainty; we call this process inferential statistics.**

STATISTICAL ACTIVITIES

- **If we observe the total population we perform only the methods of descriptive statistics.**
- **When only the sample is available we usually need to perform description and inference.**
- **Relationship analysis could be performed in both situations.**
- **Which method is to be used depends on statistical features of variables.**

Major objectives of statistics

- **1. To make inferences about a population, by analysing sample data;**
- **2. To make assessments of the extent of uncertainty in these inferences; and**
- **3. To design the process and extent of sampling so that the observations allow us to draw valid and accurate inferences.**

2. Basic concepts

POPULATION AND SAMPLE

- **POPULATION - A statistical population** is the complete set of possible measurements corresponding to the entire collection of units for which inferences are to be made. So, ***the population includes all members of a defined group.***

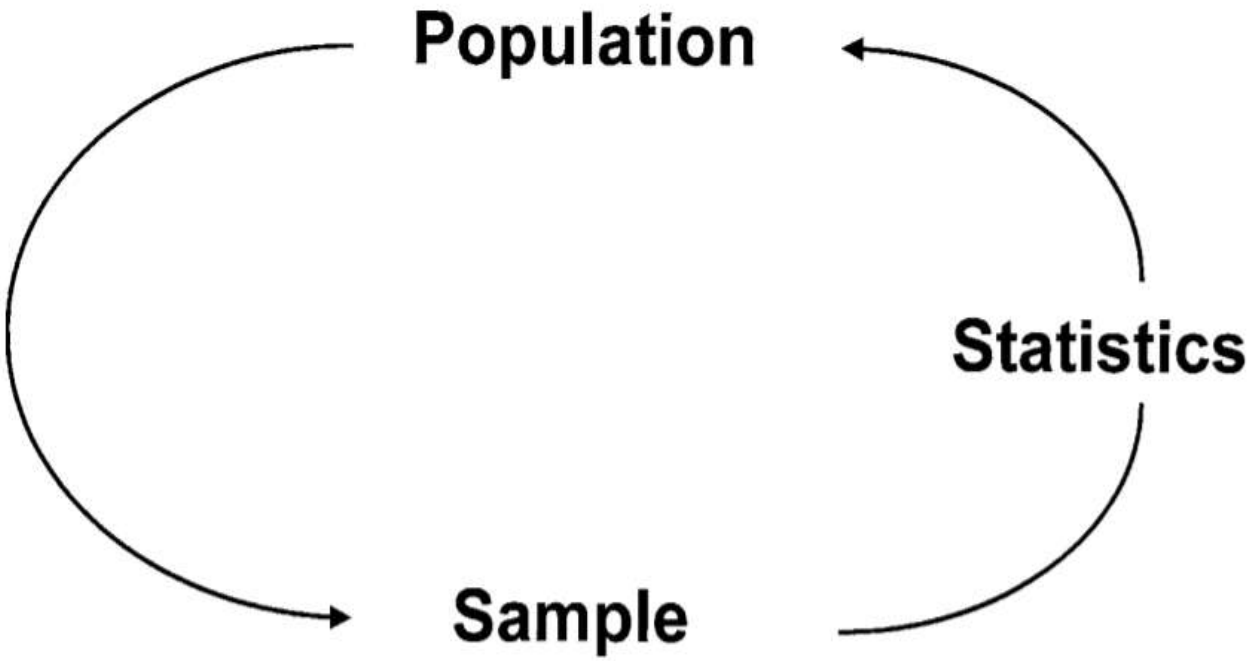


Diagram to show the role of statistics in using information from a sample to make inferences about the population from which the sample was derived

POPULATION AND SAMPLE

The population represents the target of an investigation, and the aim of the process of data collection is to make inferences (draw conclusions) about the population.

Examples of populations:

- all patients with a certain disease,

- all inhabitants of Bulgaria

POPULATION AND SAMPLE

For meaningful measurements on a single patient, it is desirable to compare them with the distribution of all such measurements on the complete population of diseased persons in the same categories (sex, age group, geographic area, and so on).

But it is obviously impossible to obtain such data on complete populations; therefore, investigations are to be carried out on a **representative subset called a *sample.***

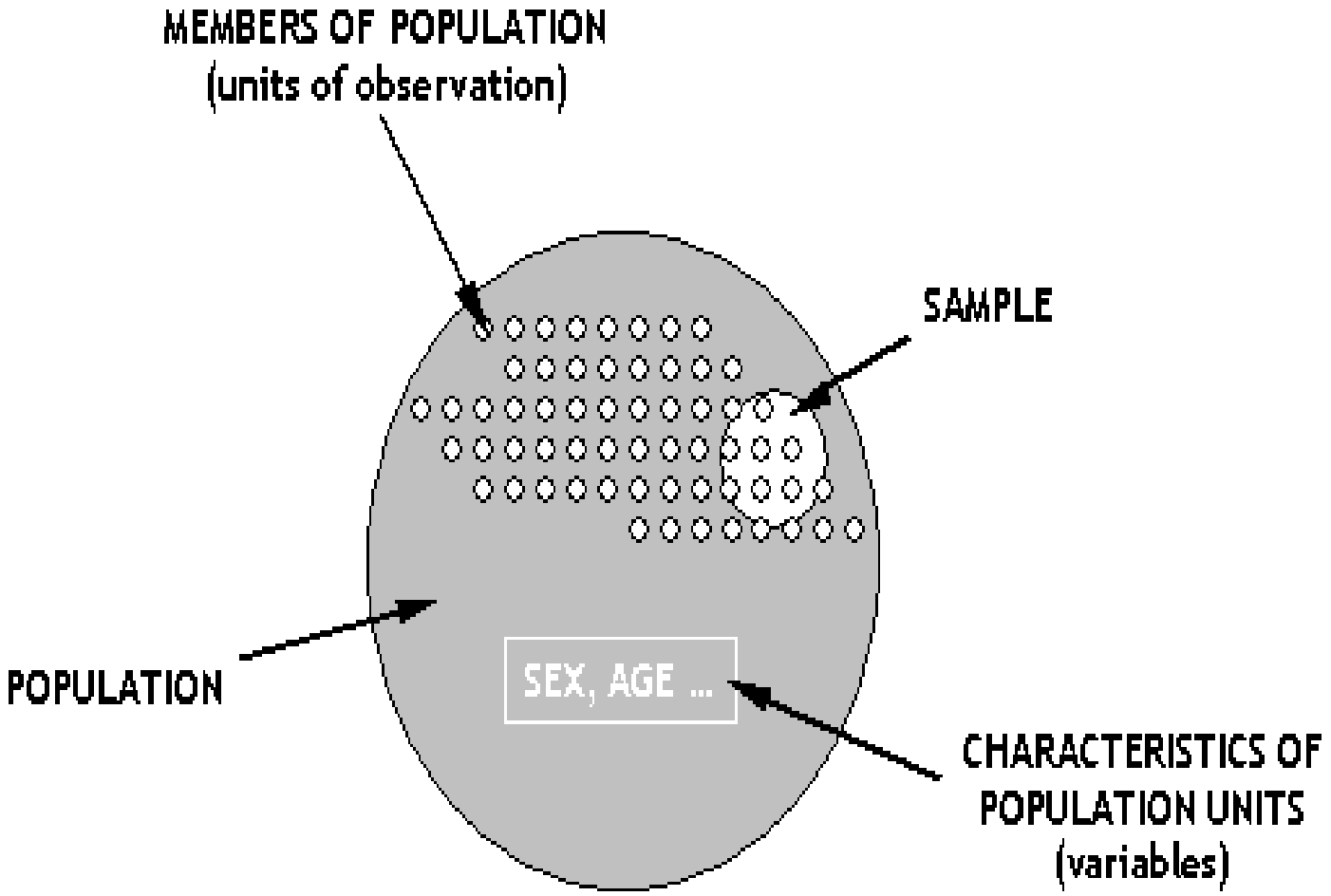
POPULATION AND SAMPLE

Thus, a **SAMPLE is a subset or a fraction of the population. It is a small group drawn from a larger population.**

- *Examples of samples:*

- *50 patients with a certain disease from one regional hospital*
- *100 newborns from a neonatal clinic*

POPULATION AND SAMPLE



POPULATION AND SAMPLE

- The raw data of an investigation consist of observations made on individuals. In many situations the individuals are people, but they need not be. For instance, they might be red blood cells, urine specimens, rates, or hospitals.
- The number of individuals in a sample is called a **sample size.**

3. Types of studies and the research process

TYPES OF STUDY

- **1. “GENUINE” SAMPLE STUDIES**

For most studies, the sample size falls between the above two extremes. In such studies, called **representative studies,** statistics is most useful.

They are the most common in medical practice and science as there are millions of births, deaths, diseases, etc.

TYPES OF STUDY

- **2. 100% STUDIES -** the study of the entire population. There is only the need for summarizing the data and no inferences are to be made as all the information has been gathered.

Example: The decennial census in different countries (like in Bulgaria).

- **3. N=1 STUDIES (MONOGRAPHIC STUDIES) -** ***Example:*** An outbreak of salmonella food poisoning

THE RESEARCH PROCESS

⇒ **PLANNING**

⇒ **HYPOTHESES OR AIMS**

⇒ **RESEARCH DESIGN**

⇒ **DATA COLLECTION**

⇒ **ORGANISATION AND PRESENTATION OF DATA**

⇒ **DATA ANALYSIS**

⇒ **INTERPRETATION AND CONCLUSIONS**

RESEARCH PLANNING

- The planning must take into consideration both the previous research evidence as well as the ethical and economic factors before the appropriate research strategy is selected, and the precise research hypothesis or aim is stated.
- ***Hypothesis*** **–** a proposition about relationship between variables or about differences between groups that are to be tested.
- ***Variable*** **-** a property, attribute or measurement that varies from case to case (age of patients, temperature, blood pressure, height, weight, etc.)

RESEARCH DESIGN (STRATEGY)

The most ***vital part of any investigation is its design***, just as a house must be built upon solid foundations.

RESEARCH STRATEGIES

- **EXPERIMENTAL STUDIES** or ***randomized studies*** - the group units are determined by a chance - ***"randomization"***

Randomization allows to make the groups as similar as possible except in one respect that is being studied. So, any differences observed later can be attributed to this factor alone.

Example: controlled clinical trials to compare new treatment with standard one.

- The randomized, double-blind, controlled clinical trial is the ***"gold standard"*** of medical research.

RESEARCH STRATEGIES

- **NON-EXPERIMENTAL STRATEGIES** or non-randomized studies, called ***"observational studies".*** There is no assignment of cases to experimental groups (e.g. people themselves decide whether or not to smoke).

Two main types of observational studies:

- ***case-controls studies***

- ***cohort studies***

They also may be***: prospective and retrospective.***

4. Sampling and sample types

SAMPLING

- Research in health sciences usually involves data collection on a ***sample*** of cases, rather than on the entire population as ***it is often impossible or extremely costly to study complete populations***.
- ***The aim of all sampling methods is to draw a representative sample*** from the population which allows later to generalize findings from the sample to the population.

SAMPLING

- If the sample is biased (not representative) the generalization will be less valid, and the conclusions or inferences about the population might be incorrect.
- So, the main point is **how to draw a representative sample.**
- **The main principle of sampling**: ***every individual should have a known chance of being included in the sample.***

SIMPLE RANDOM SAMPLE

It is suitable for small-scale studies. Each individual among all who could be sampled, say N, has equal chance (1/N) of inclusion in the sample.

Example: To select a simple random sample of 50 cases from a set of 800 births, we could read a ***table of random numbers***: 12454 45730 07944 73506 81149……. Then select numbers 124, 544, 573, ……. or we could use an equivalent computer programs.

SYSTEMATIC RANDOM SAMPLE

- This involves working through a list of the entire population and choosing, for instance, every tenth (for 10% sample of N available) or twentieth case (for 5%) for inclusion in the sample.

- This approach is useful when cases are automatically time-ordered, such as arrival or discharge of hospital inpatients.

- In the simple and systematic sample there is a need of a list of all population – this is not always possible.

STRATIFIED RANDOM SAMPLE

- Sometimes it is known in advance that there are important subgroups within the population that may affect the results (for instance, males and females, different age groups, etc.).
- The proportions of these subgroups in the sample must be the same as in the population. In this case the sample will be representative to the population.
- A list of all members of the population, their characteristics and the proportions of the important groups within the population need to be known.

Example 1. Say that it is known that coronary disease occurs twice as frequently among males as females and three times more commonly among over 50 year-olds than those under 50. Given a stratified sample of 480, how many females over 50 would you expect in the sample?

Example 2. Say that it is known that coronary disease occurs twice as frequently among males as females and three times more commonly among over 50 year-olds than those under 50. Given a stratified sample of 240, how many males over 50 would you expect in the sample?

MULTISTAGE CLUSTER SAMPLE

- This approach is used when we want to sample for a large-scale study spread over a wide geographic area. As its name implies, it involves multiple stages of sample selection.

- ***Example:*** To obtain a random sample of all babies born in maternity units across Bulgaria, we can firstly choose a random sample of health districts, than a random sample of hospital units within those districts, then select wards within those units, but at the final level, again, we will choose a simple random sample in each ward.

OTHER SAMPLES

- The above four methods are all “good” for choosing representative samples.
- Other methods that are less reliable (less good) for getting correct conclusions:
- ***Convenience sample*** – it includes subjects who are easiest to select (e.g. first 50 people on the street at one time)
- ***Self-selected sample -*** in postal surveys, *non-responders may bias the results*
- *These two types of samples are not representative and not recommended to be used.*

SAMPLE SIZE

- There is no magic number that we can point to as an optimum sample size.
- It depends on the characteristics of an investigation.
- The sample size must be adequate for making correct inferences from a sample to a population.
- It relates to the concept of sampling error.

Sampling error

The value of the standard error depends:

- **on the variability of individual measures in the sample (expressed by the standard deviation – s) and**
- **on the sample size (n).**

It is expressed by the formula:

$\bar{x}$

Sampling error

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where:

$s_{\bar{x}}$ - the standard error of the sample mean

s – the standard deviation of the sample mean

n – the number of cases in the sample

SAMPLING ERROR

- **Thus, sampling error** is directly proportional to the variability of individual measures expressed by the standard deviation.
- The higher the standard deviation, the higher the sampling error.
- That is why the problem of how the sample is being chosen is of utmost importance.

SAMPLING ERROR

- **Secondly, sampling error** is inversely proportional to the square root of the sample size.
- For example, if a researcher would like to minimise the sampling error he/she should increase the number of units in the sample.
- A 4-fold increase in sample size would result in only 2-fold reduction in the sampling error.
- If we want to decrease the sampling error 3-fold the sample size should be 9 times larger.

5. Classification of variables

VARIABLE

- ***Variable*** - a property or attribute that varies.

Each variable has:

- ***variable values*** – every single variable can take two or more different values;
- ***distribution of variable values*** – the complete summary of the frequencies of the values of a single variable.

Classifying variables

1. Regarding their values	
Quantitative (numeric)	Qualitative (categorical)

- ***Quantitative (numerical) variables*** – values of which are expressed by numbers (e.g. weight, number of patients per day);
- ***Qualitative (categorical) variables or attributes*** – values of which are expressed only by description (e.g. gender, residence, blood group, profession, marital status, ethnic group, etc.

Quantitative (numeric) variables

2. Regarding the possibility of infinitive number of their values may be:

Continuous variables	Discrete variables

Continuous variables – with potentially infinite number of possible values along a continuum; set of observations that may theoretically lie anywhere within a specified interval on the number scale; ***the process of measurement produces continuous data; any value inside a range is possible.***

Continuous variables may be presented on:

- ***interval scale*** - no true zero (e.g. temperature)
- ***ratio scale*** - has a true zero (time, weight, height).

Discrete variables – values of which could be arranged into naturally or arbitrarily selected groups of values; set of observations that may lie only on certain isolated points on a number scale; ***the process of counting produces discrete data;*** e.g. number of patients per day, number of children in a family, number of live births, number of deaths, etc.

Qualitative (categorical) variables

3. Regarding the orderliness of values:

Ordinal variables	Nominal variables

- ***ordinal variables*** – values of which are classified into ordered categories, the measurements are on an ***ordinal scale***; e.g. pain intensity (excruciating, severe, moderate, mild, no pain), education (primary, secondary, higher), etc.

- nominal variables - there is no natural ordering of categories; the measurements are on a ***nominal scale***; they can be reduced to “yes” or “no”, e.g. gender (dichotomous); blood group (polychotomous), residence, profession, marital status, etc.

Classifying variables

4. Regarding the number of distinct values:

Dichotomous or binary	**Polychotomous**

- ***dichotomous or binary variables*** – with only two possible values, often containing information of having the characteristic of interest or not;
- ***polychotomous variables*** – with more than two possible values.

Classifying variables

5. Regarding the relationship between two or more variables:

Dependent variables	Independent variables

- ***dependent variables*** – values of which are depending on the effect of other variables (independent variables) in the relationship under study; they describe the results or the outcome;
- ***independent variables*** – that are hypothesized to influence the values of other variables (dependent variables) under study; they describe the factors or causes.

Classifying variables

- All these classifications could be linked to each other. When we put the classification on numerical and categorical variables in the central position and link it to all other classifications, then we can say that:
  - **numerical variables** are continuous or discrete, only ordinal and polychotomous, dependent or independent;
  - **categorical variables** are only discrete, dichotomous or polychotomous, ordinal or nominal, dependent or independent.

TYPES OF VARIABLES

In summary, we usually classify variables into ***four main types of variables***:

- Numerical continuous variables
- Numerical discrete variables
- Categorical ordinal variables
- Categorical nominal variables

6. ORGANIZING AND PRESENTING DATA

SOURCES OF DATA

- ***Primary data*** - data we collect; how best to collect and how much is needed.
- ***Secondary data*** - data that someone else has already collected; there are various sources of routine health data, published by the NHS or after Population Census, health data collected at regional and district level, etc.

ORGANIZING AND PRESENTING THE DATA

- For any kind of a research it is of basic importance to have data well organized and prepared for description and analysis. As the methods for both kind of activities are statistical methods, it is very important to follow the rules of preparing the data in an adequate structure for statistical analysis.

SUMMARIZING AND PRESENTING DATA

- Before the interpretation of the information, the raw data must be organized and presented in a clear and intelligible fashion, so the objective here is to convert masses of numbers - raw data - into meaningful summaries, called **statistics.**
- ***Statistic*** - a particular number obtained by the mathematical treatment of specific data; **a number resulting from the manipulation of the raw data.**

SUMMARIZING AND PRESENTING DATA

- **How this objective is fulfilled depends on:**
  - the purpose of the investigation;
  - the type of data involved; and
  - the intended audience.

  For descriptive purposes and internal consumption - pictorial presentation of data is enough.

  For external consumption and/or inferences to a wider population - numerical summaries are needed..

After the collection of raw data they should be organized and presented in a meaningful way.

Frequency distributions give a general picture of the pattern of the observations but sets of measurements cannot be adequately described only by the values of all individual measurements.

For many purposes, the overall summary of a group's characteristics is of utmost importance.

There are two basic methods of summarization:
numerical and graphical.

The objective of the numerical approach is to convert masses of numbers (raw data) into meaningful summary statistics (indices), reduced to a single number, that convey information about the average (typical) degree of a given variable and the degree to which observations differ (the degree of dispersion or spread).

6.1. Table presentation

ORGANIZING AND PRESENTING THE DATA

- The adequate structure is DATA MATRIX. This is a structure in which data of all observational units and all observed attributes of units are organized in a **table.**
- The basic element of table is a **cell.**
- The cells are organized in **rows** and **columns.** The meaning of elements of the table are as follows:

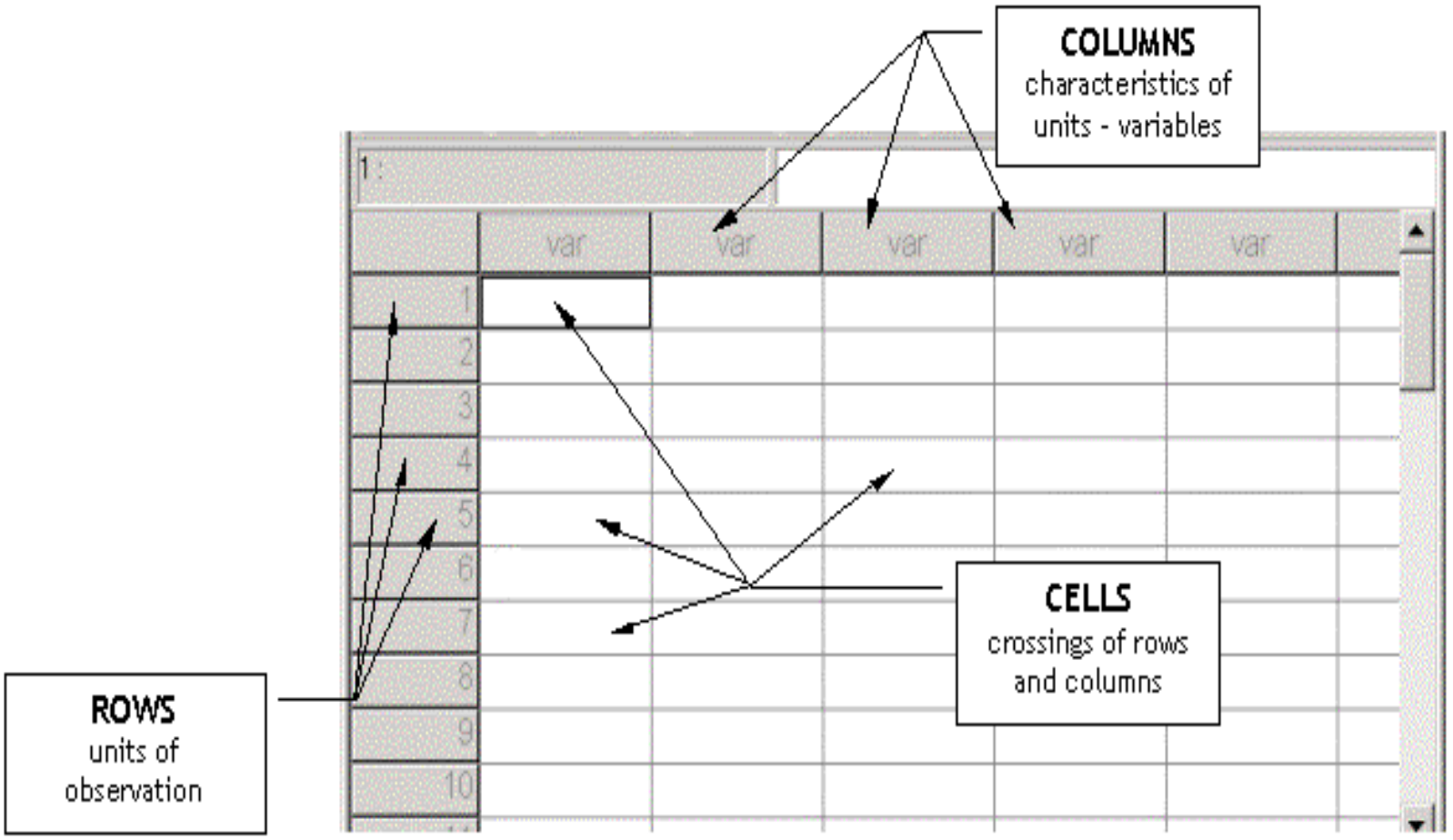
ORGANIZING AND PRESENTING THE DATA

- ***Cell*** – the record of a piece of information (lat. datum) on single attribute (variable) of a single observational unit (statistical unit).

- ***Row*** - the record of values of all variables for a single unit.

- ***Column*** - the record of values of all units for a single variable.

ORGANIZING AND PRESENTING THE DATA



DATA PRESENTATION

TABLES

- **2 x 2 Contingency tables** - two nominal variables having two categories each
- For such tables a measure called the ***odds ratio (OR)*** can be calculated which is used in epidemiology to evaluate the risk

	Yes	No	Total
Males	12	24	36
Females	10	34	44
Total	22	58	80

DATA PRESENTATION

TABLES

- **R x C tables - when a variable has 2 or more outcomes, it can be cross-classified with another one into a larger contingency table with R rows and columns**
- **In such tables percentages are often given with cell counts. It is important to know which percent - row, column or all.**

	Y1	Y2	Total
X1			
X2			
X3			
X4			
Total			

6.2. Graphical presentation

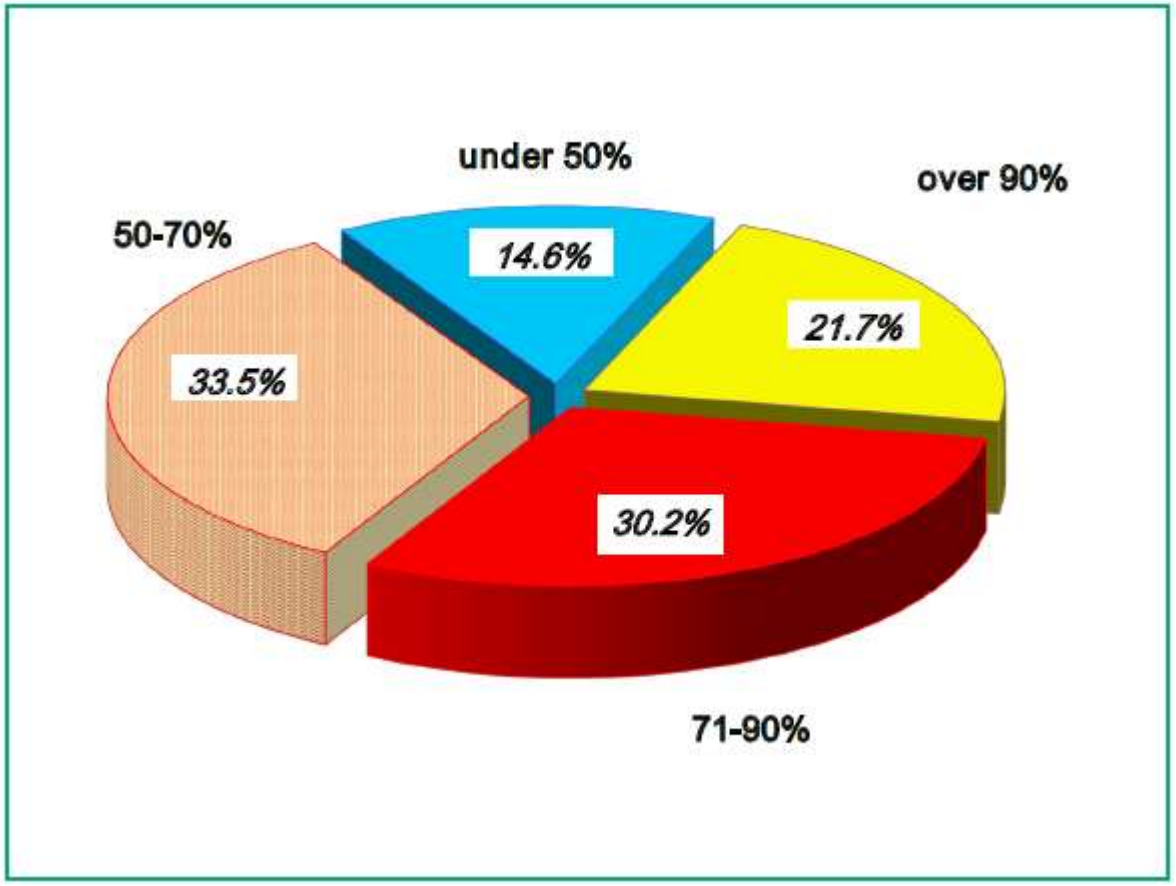
FOR QUALITATIVE VARIABLES

FOR QUALITATIVE VARIABLES

The most appropriate graphical presentation is a pie diagram.

It is constructed very easily:
- the circle is equal to 100 %;
- we calculated the proportion of each part, e.g. the proportion of men and women in a dataset;
- the sum of all proportion is to be equal to 100%.

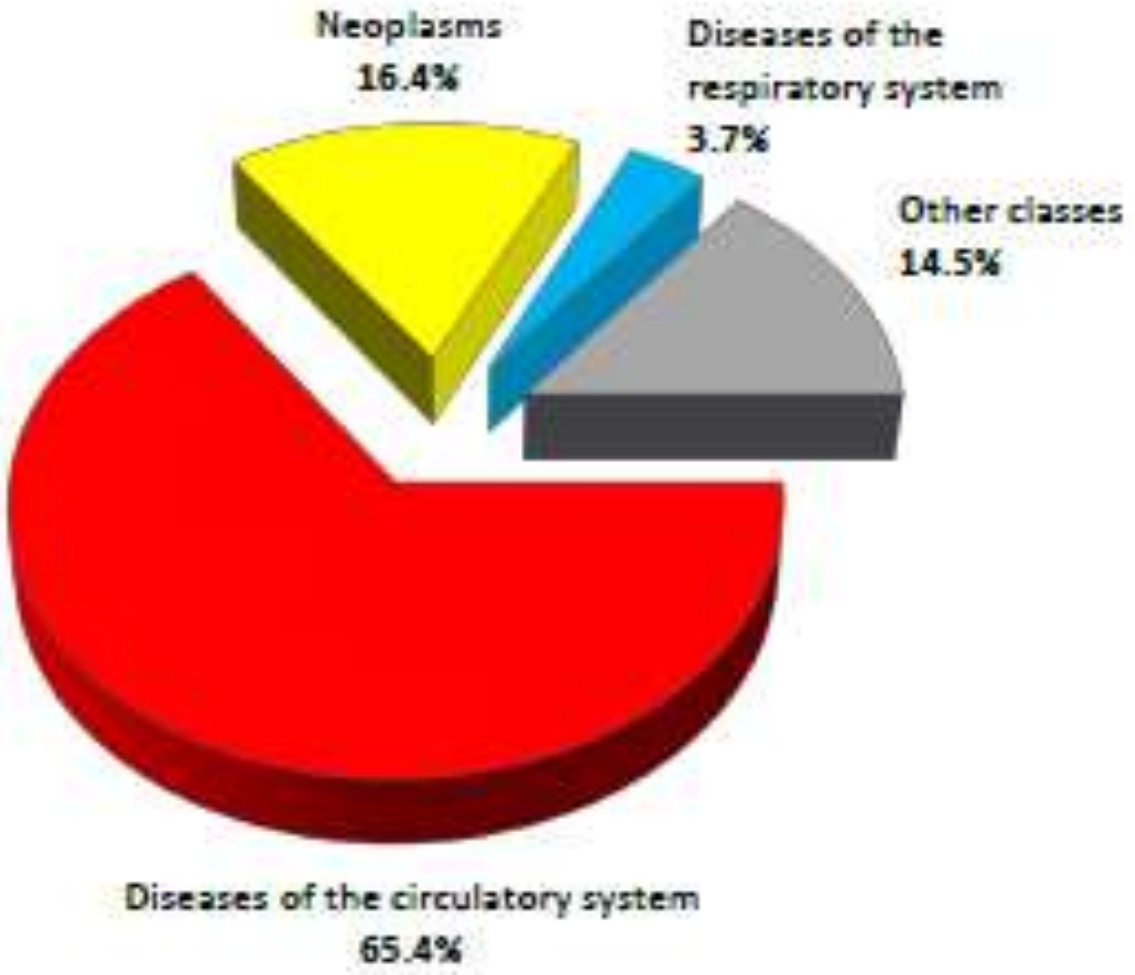
Primary Disability of Persons over 16 Years of Age in Bulgaria, 2015 (in %)



Source: Bulgarian health statistics 2016. National Centre of Public Health and Analyses

For students: Read and explain the diagram

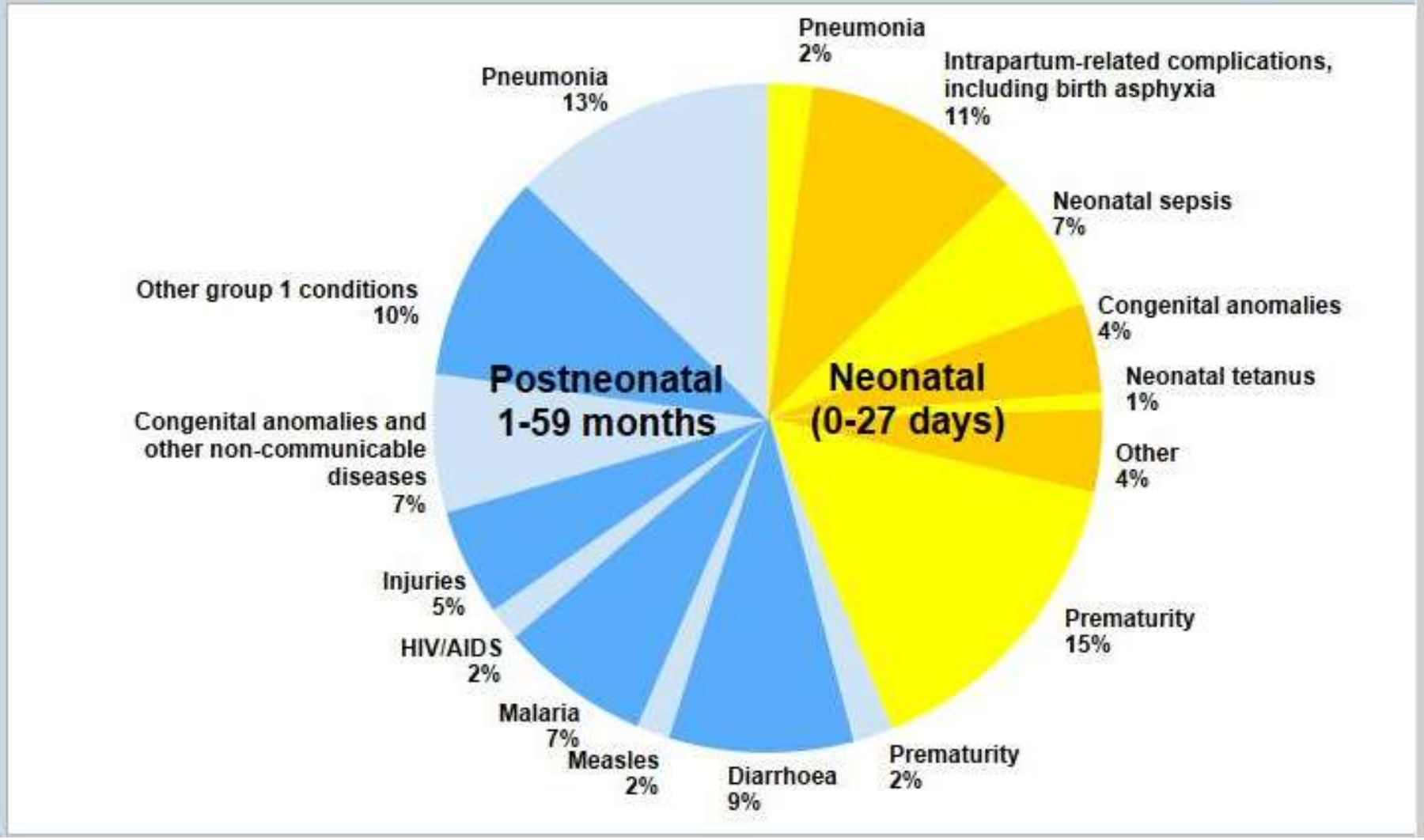
Mortality by Causes in Bulgaria, 2015 (in %_

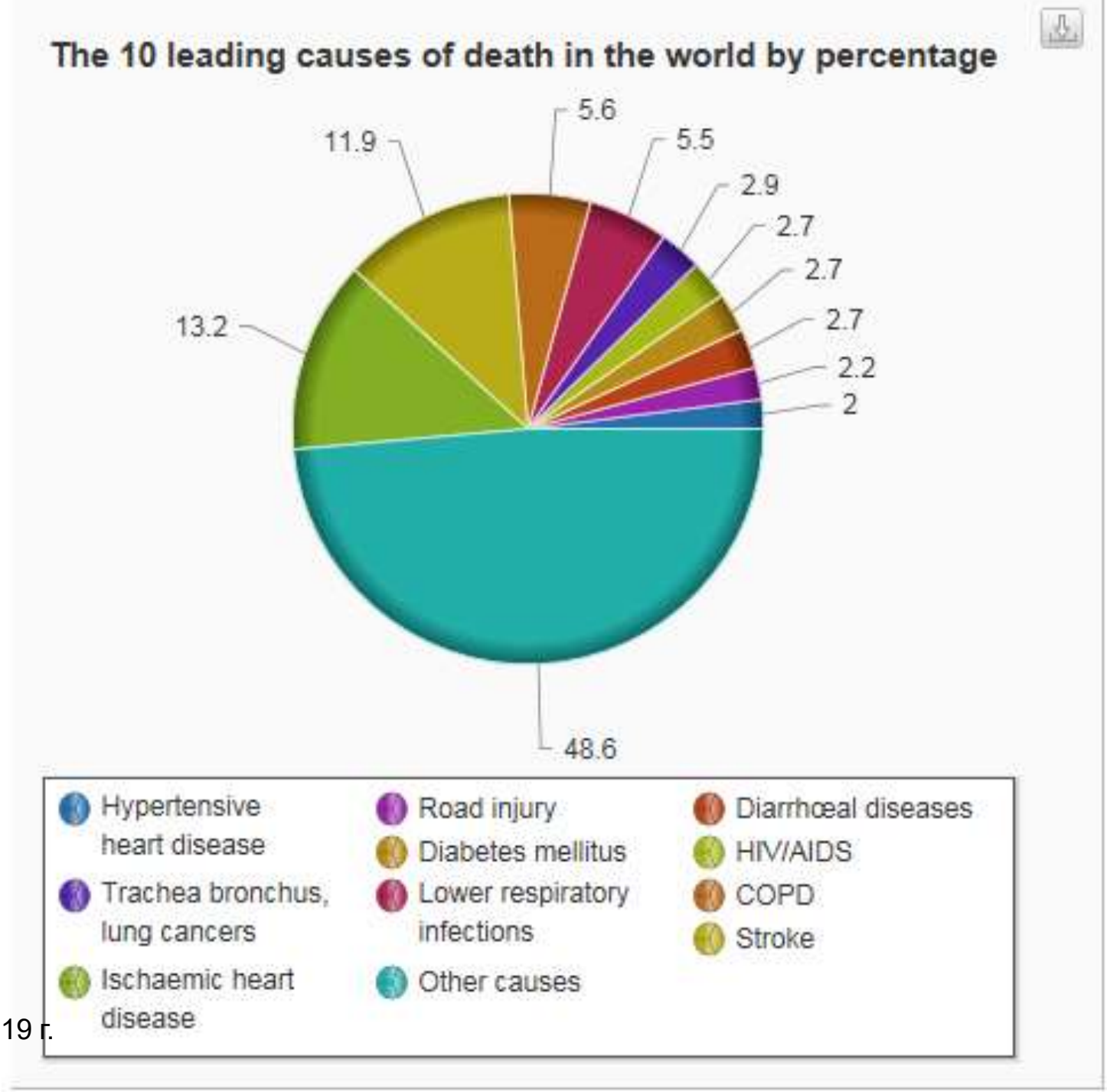


Source: Bulgarian health statistics 2016. National Centre of Public Health and Analyses

For students: Read and explain the diagram

Causes of deaths among children under 5 years, 2013




The 10 leading causes of death in the world by percentage
11.9
5.6
5.5
2.9
2.7
2.7
2.7
2.2
2
13.2
48.6
Hypertensive heart disease
Trachea bronchus, lung cancers
Ischaemic heart disease
Road injury
Diabetes mellitus
Lower respiratory infections
Other causes
Diarrhœal diseases
HIV/AIDS
COPD
Stroke

FOR QUANTITAIVE VARIABLES

FOR QUAUNTATIVE VARIABLES

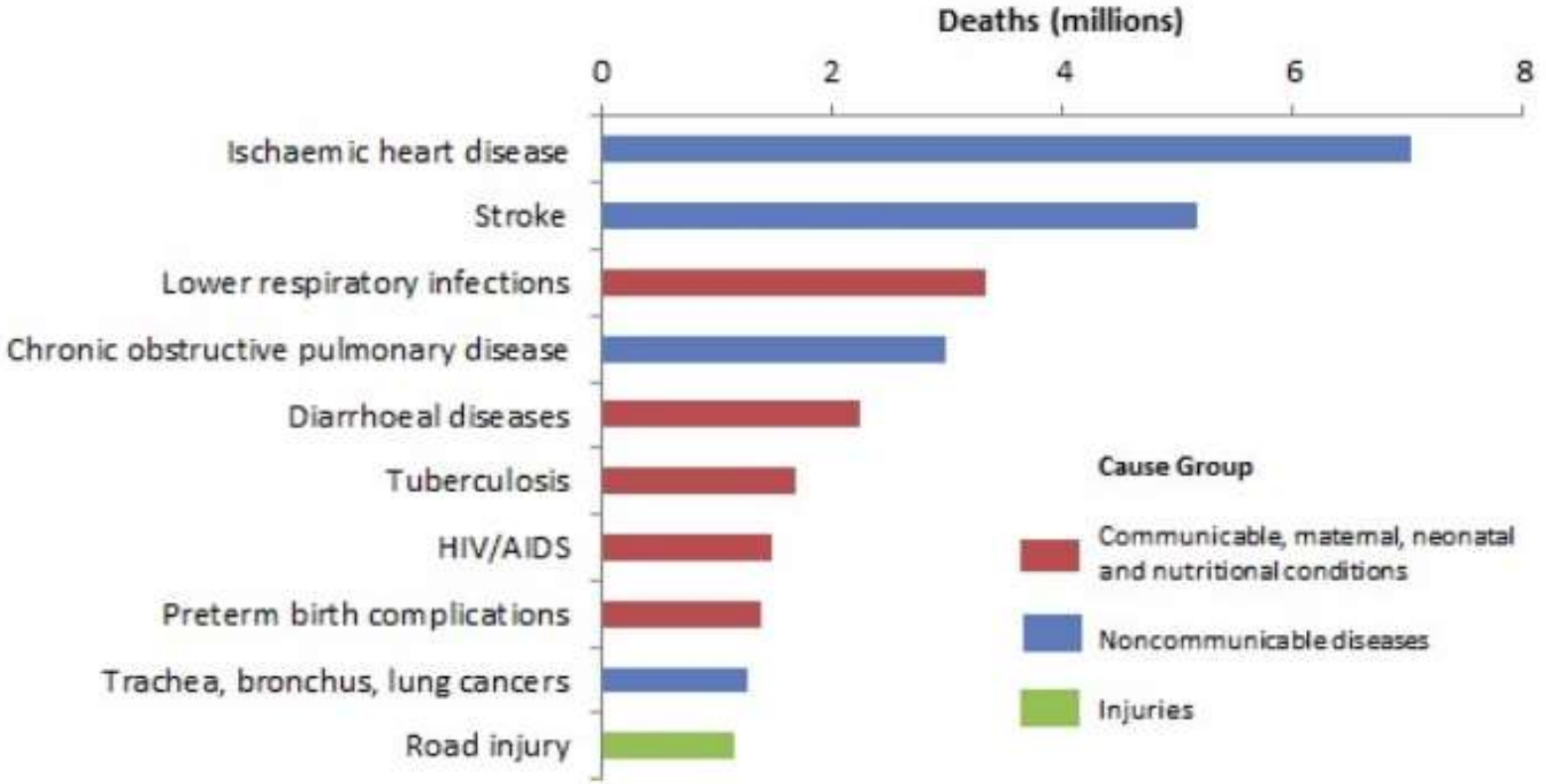
The most appropriate graphical presentations are the following:

- **histograms;**
- **bar charts;**
- **linear diagrams;**
- **map diagrams.**

In the bar charts all bars are separated.

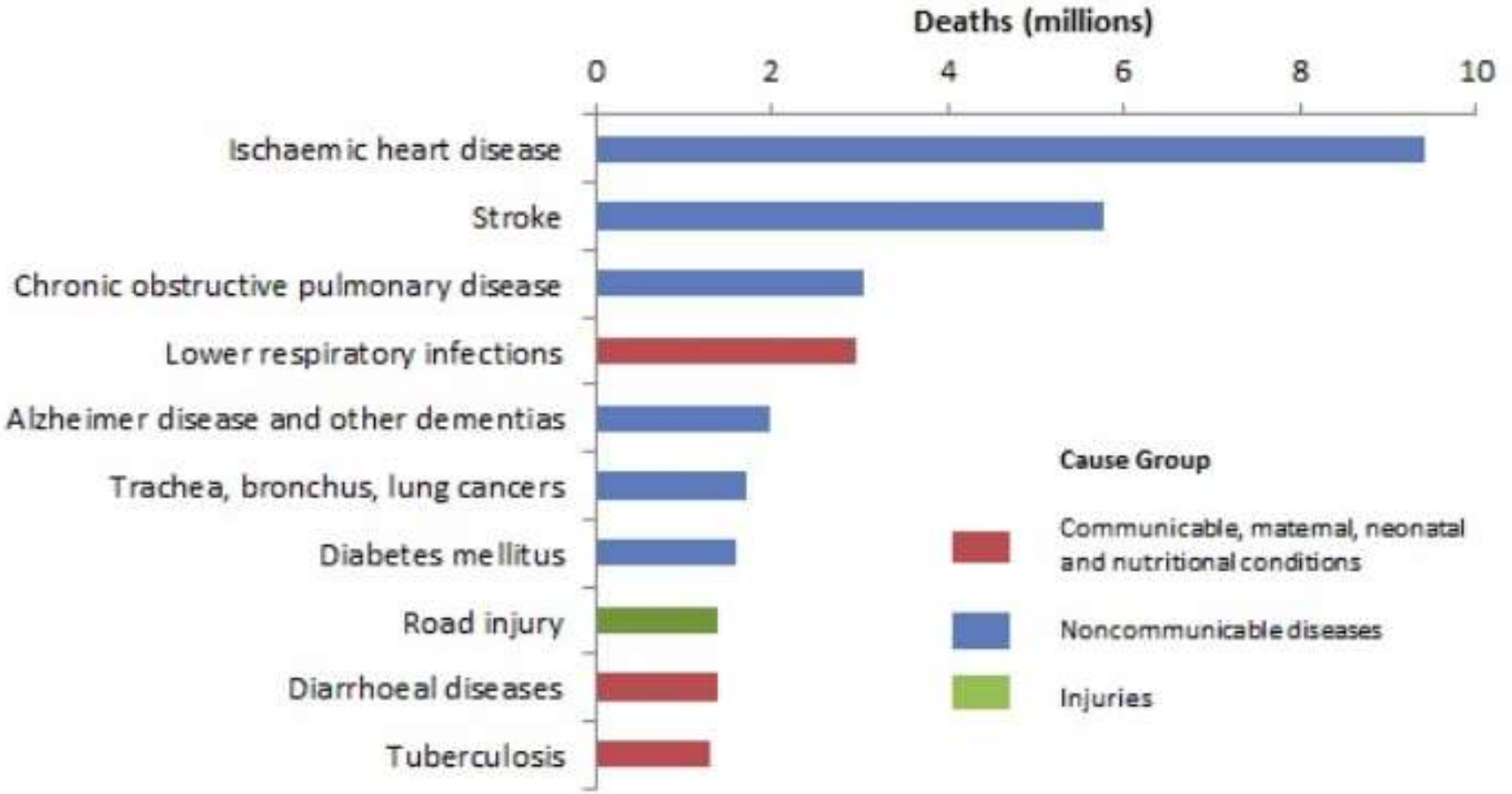
They are appropriate to express changes in rates over time or levels of rates in different areas (countries, regions, etc.)

Top 10 global causes of deaths, 2000



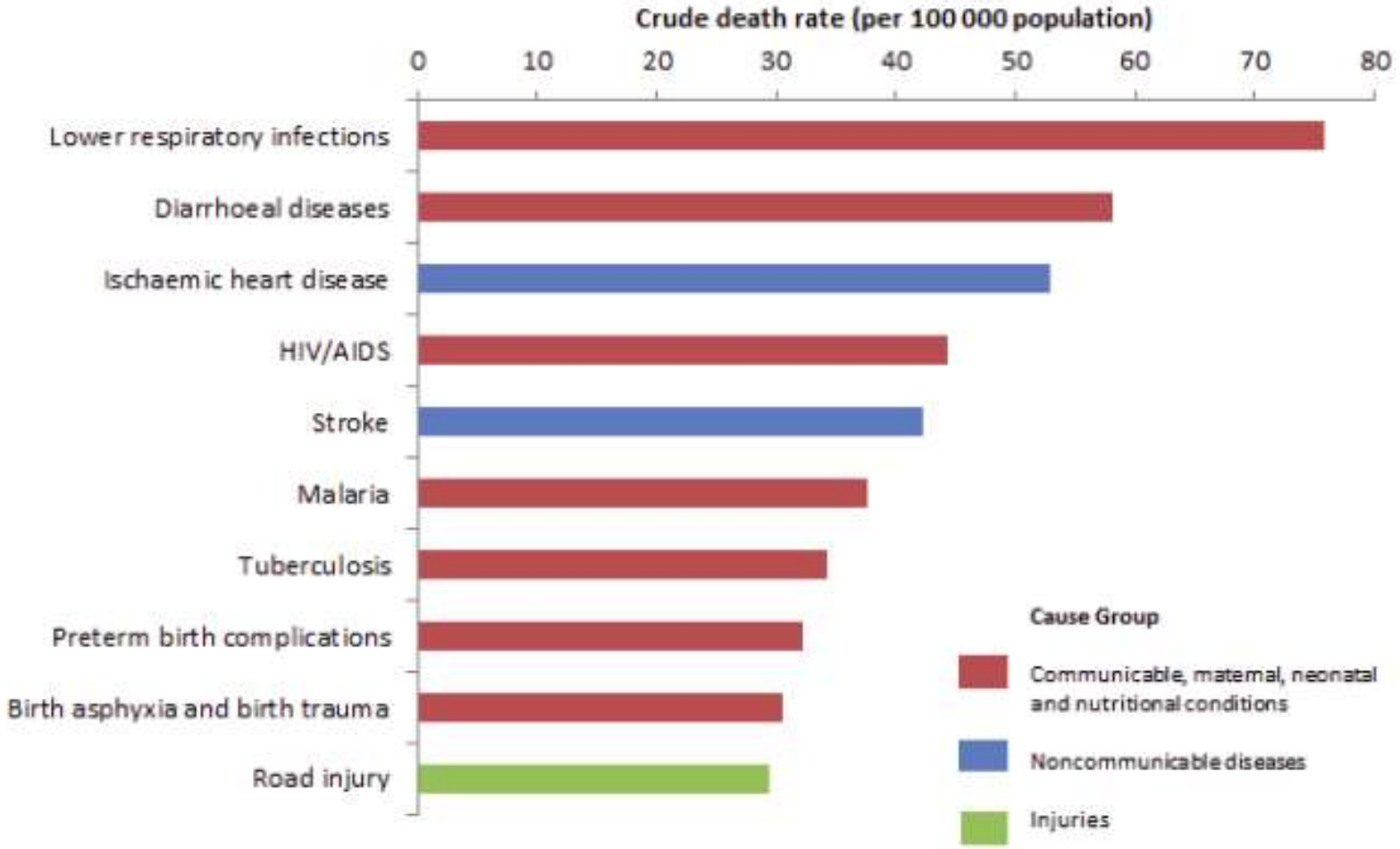
Source: Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.

Top 10 global causes of deaths, 2016

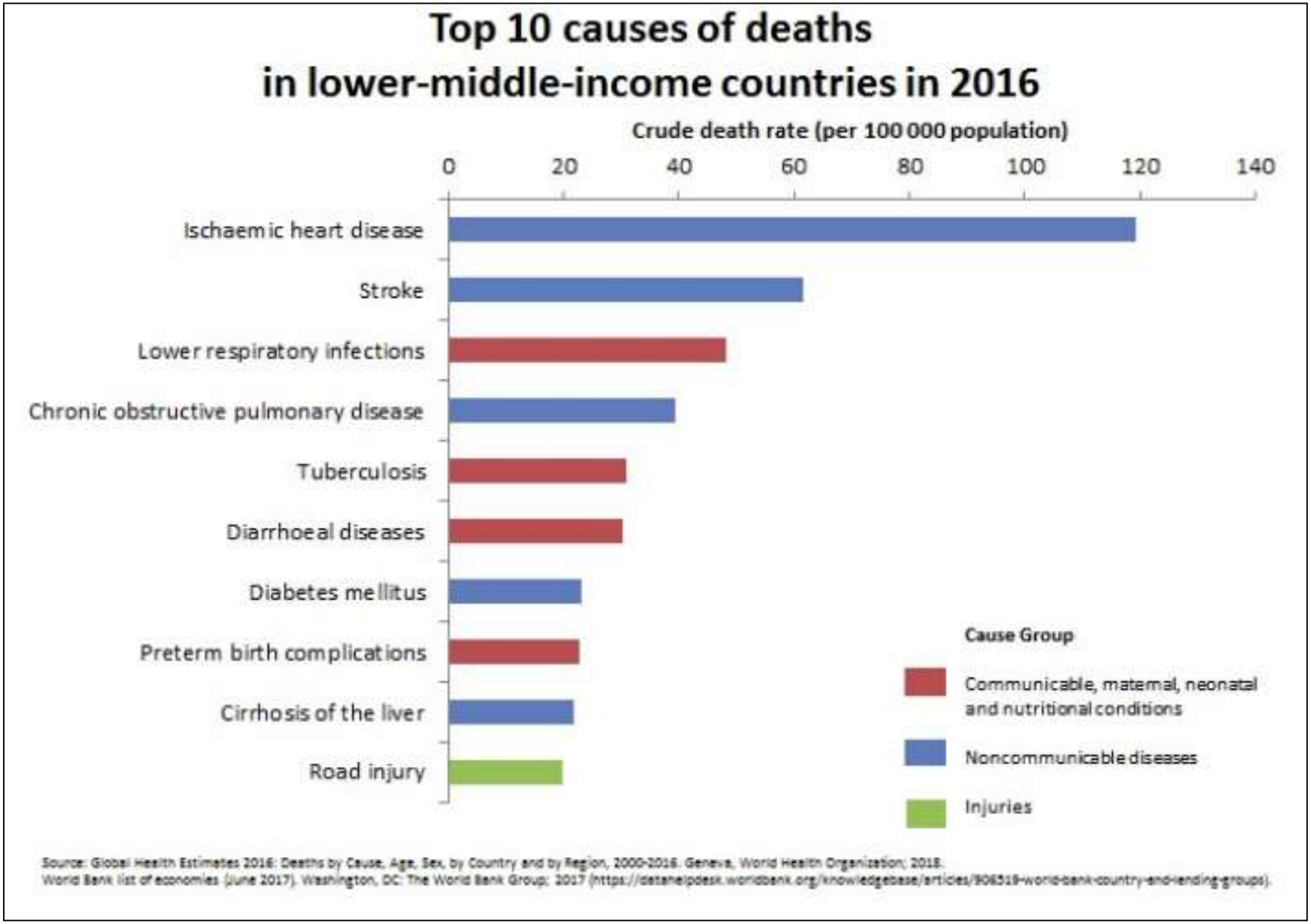


Source: Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.

Top 10 causes of deaths in low-income countries in 2016



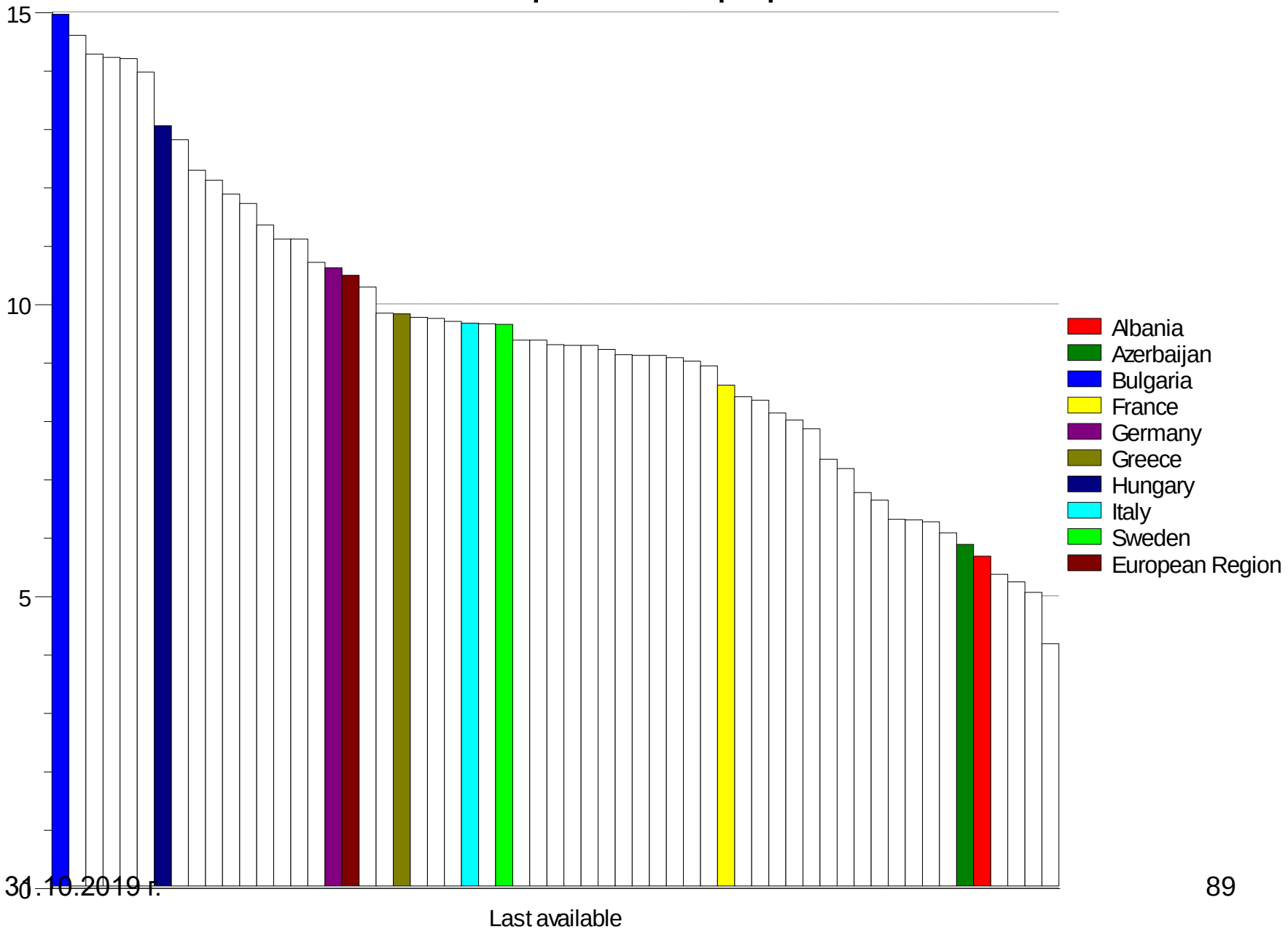
Source: Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.
World Bank list of economies (June 2017). Washington, DC: The World Bank Group; 2017 (https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups).


Top 10 causes of deaths
in lower-middle-income countries in 2016
Crude death rate (per 100 000 population)
0
20
40
60
80
100
120
140
Ischaemic heart disease
Stroke
Lower respiratory infections
Chronic obstructive pulmonary disease
Tuberculosis
Diarrhoeal diseases
Diabetes mellitus
Preterm birth complications
Cirrhosis of the liver
Road injury
Cause Group
Communicable, maternal, neonatal and nutritional conditions
Noncommunicable diseases
Injuries
Source: Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.
World Bank list of economies (June 2017). Washington, DC: The World Bank Group; 2017 (https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups).

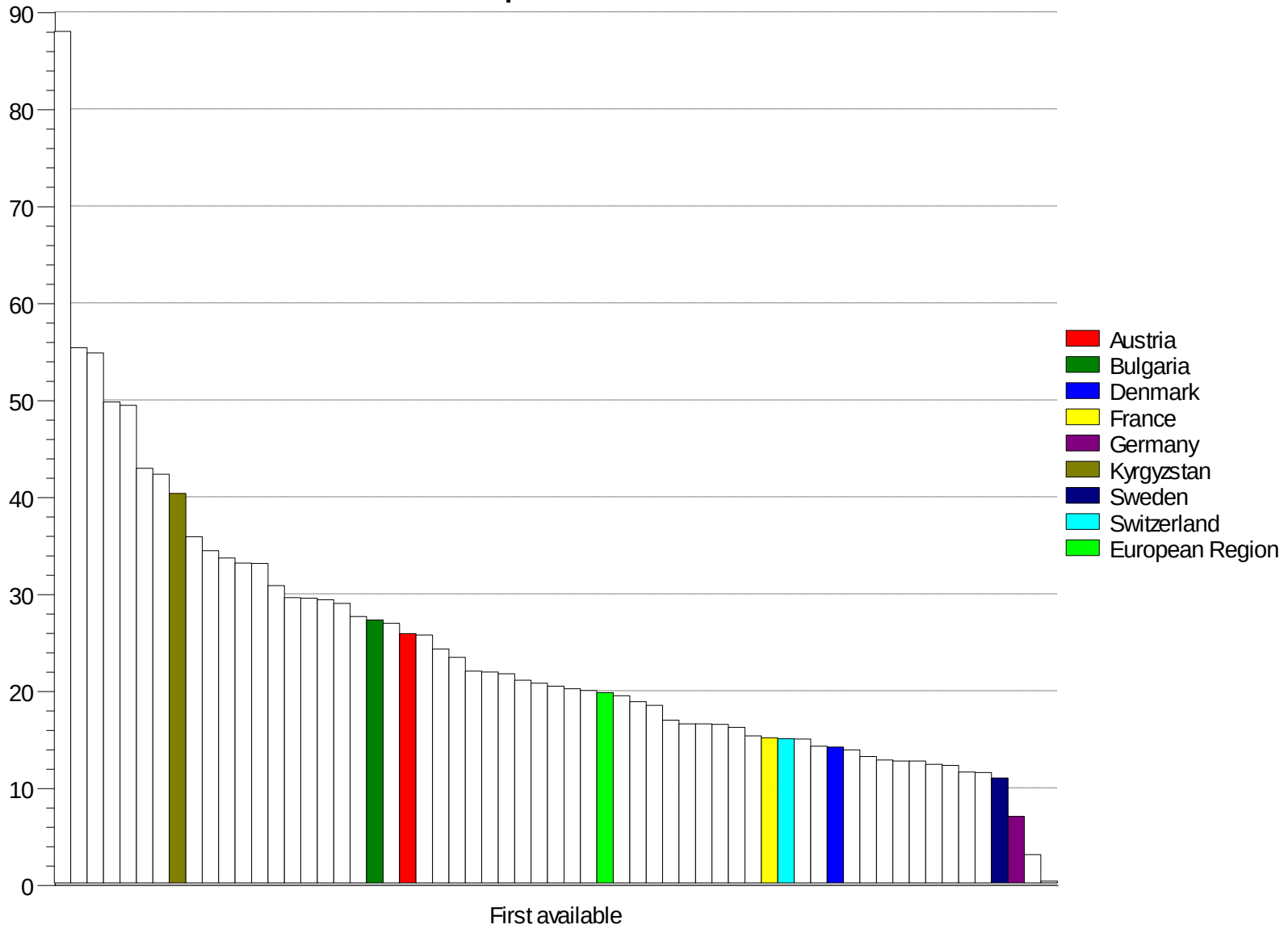
In the histograms all bars are linked to each other.

They are appropriate to express changes in rates over time or levels of rates or proportions in different areas for the same time (countries, regions, etc.)

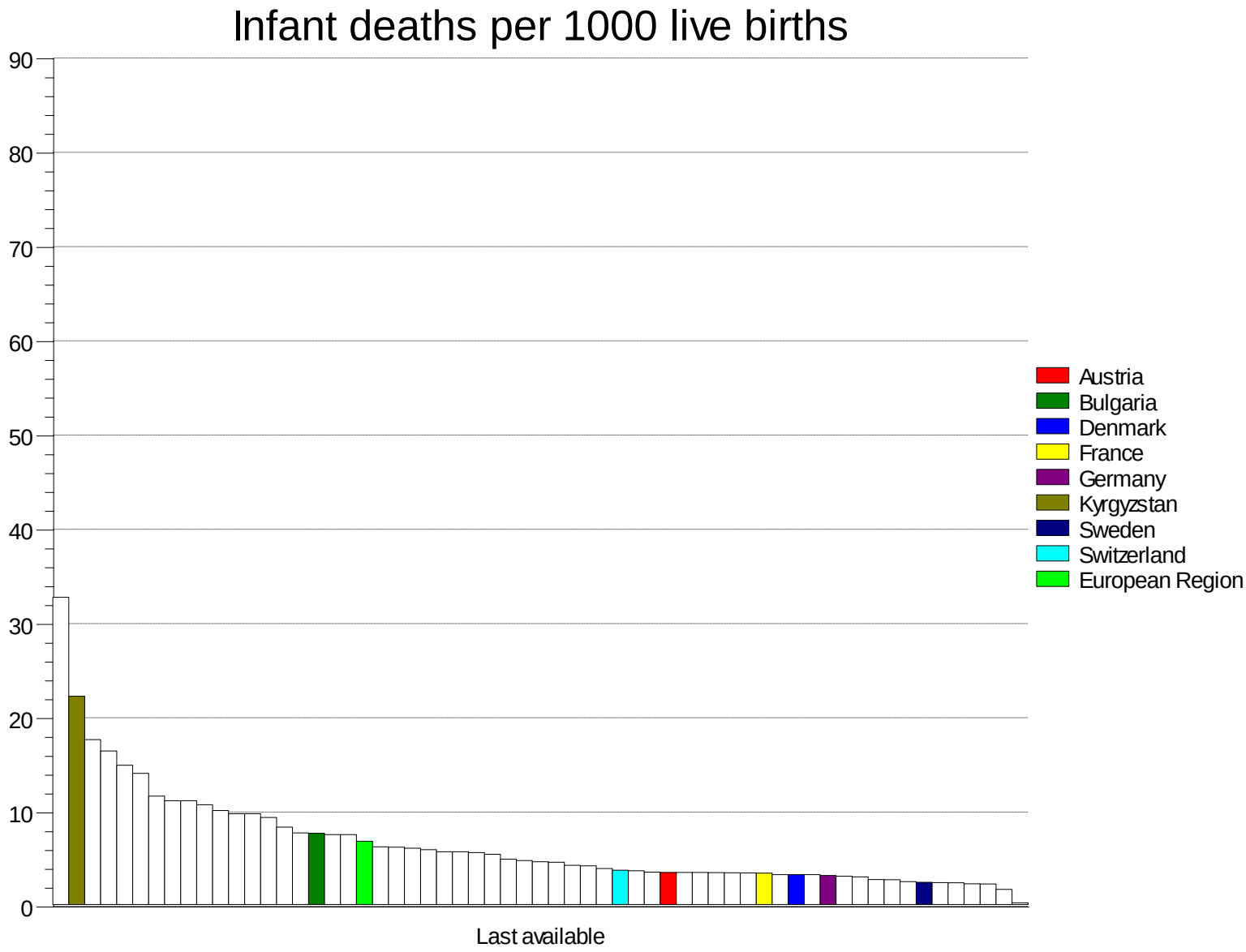
Crude death rate per 1000 population



Infant deaths per 1000 live births

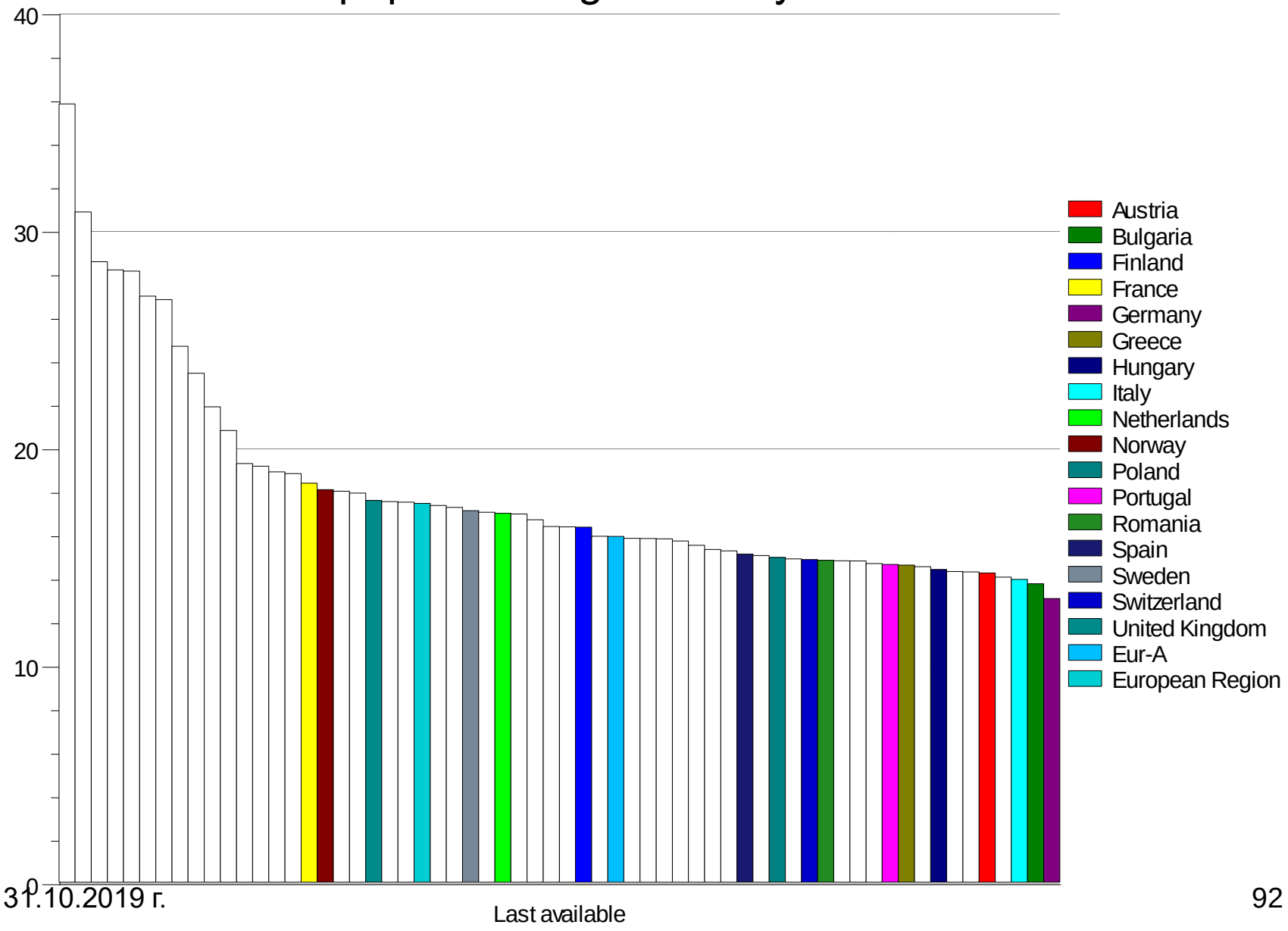


Source: European Health for All database

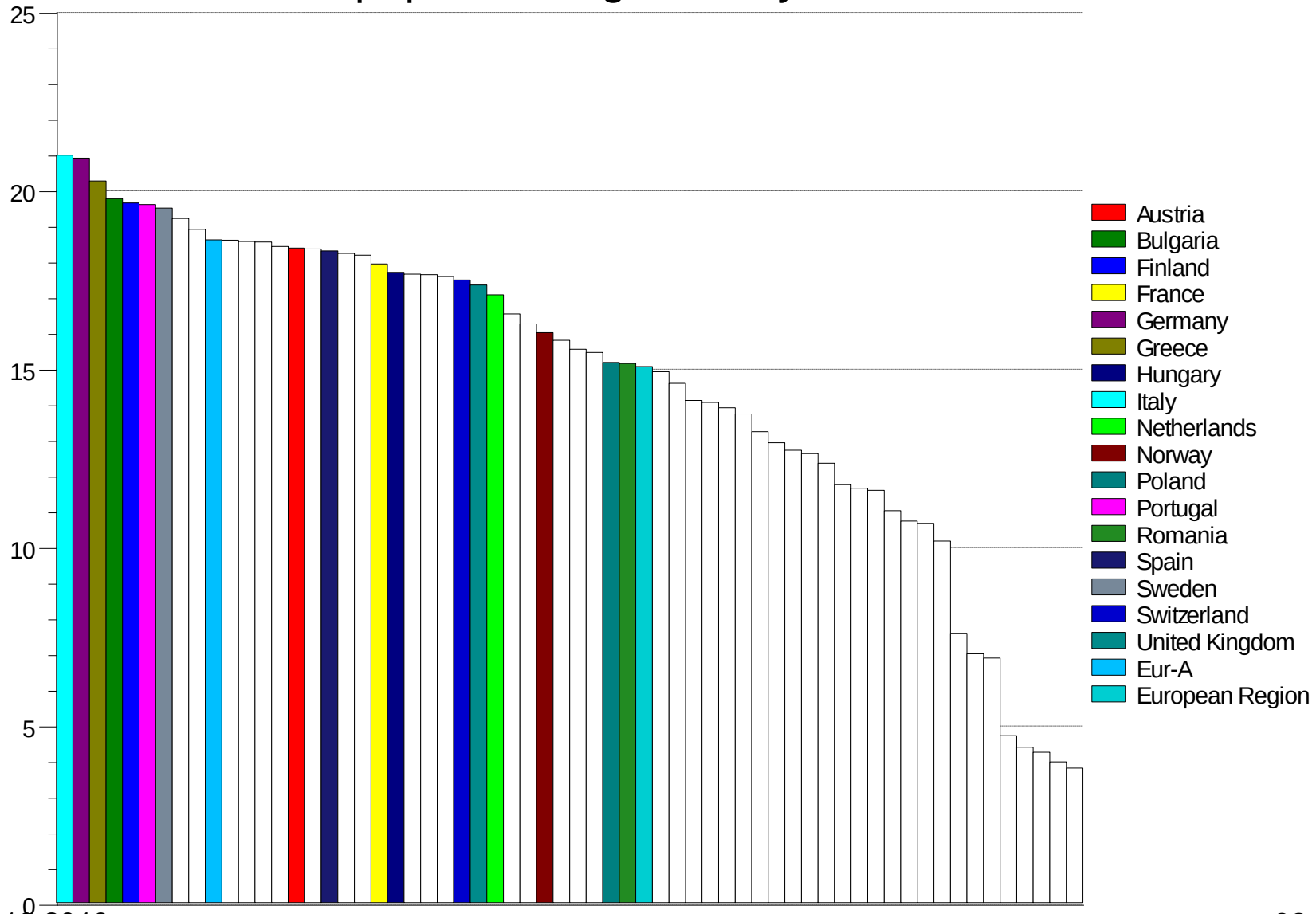


Source: European Health for All database

% of population aged 0–14 years

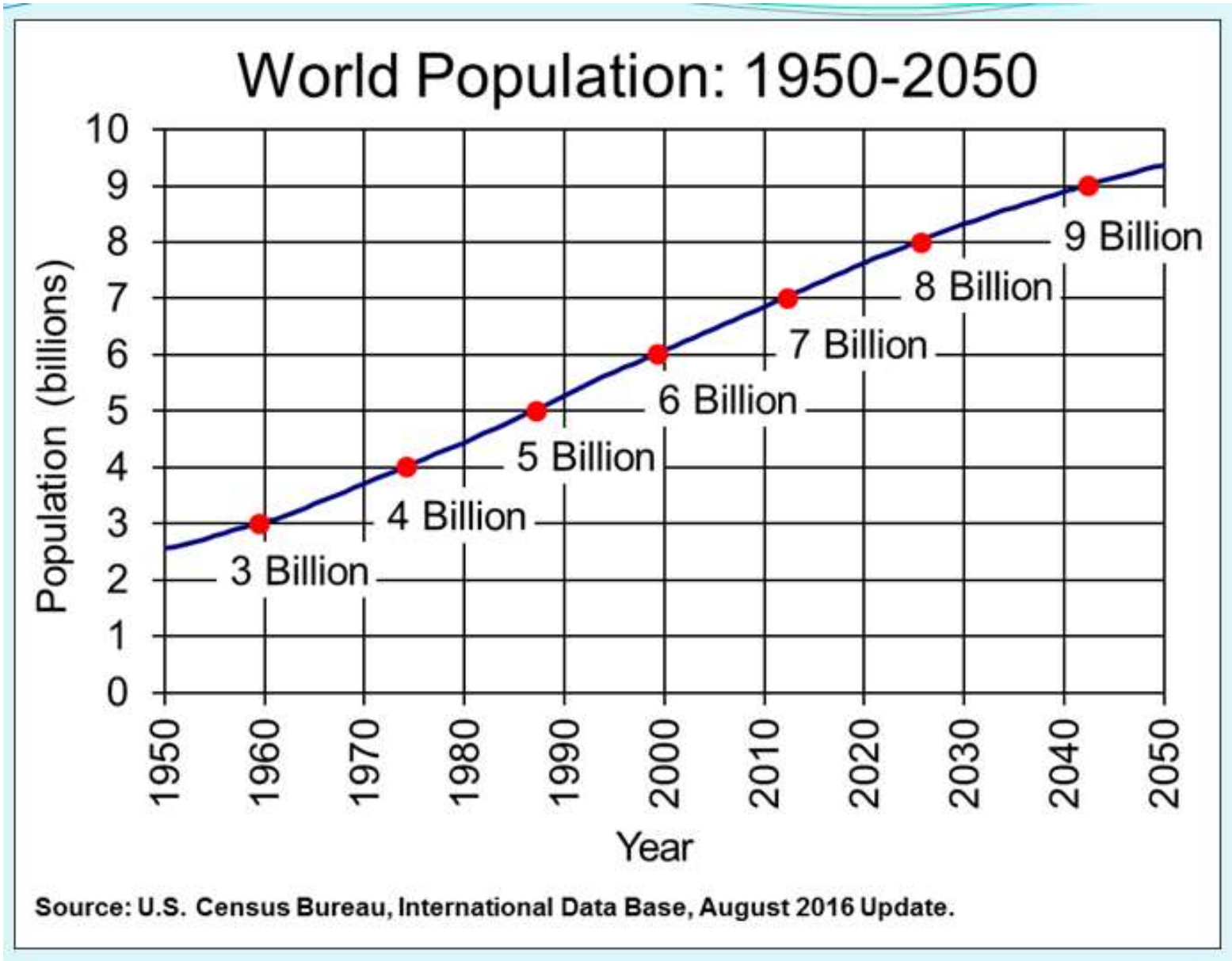


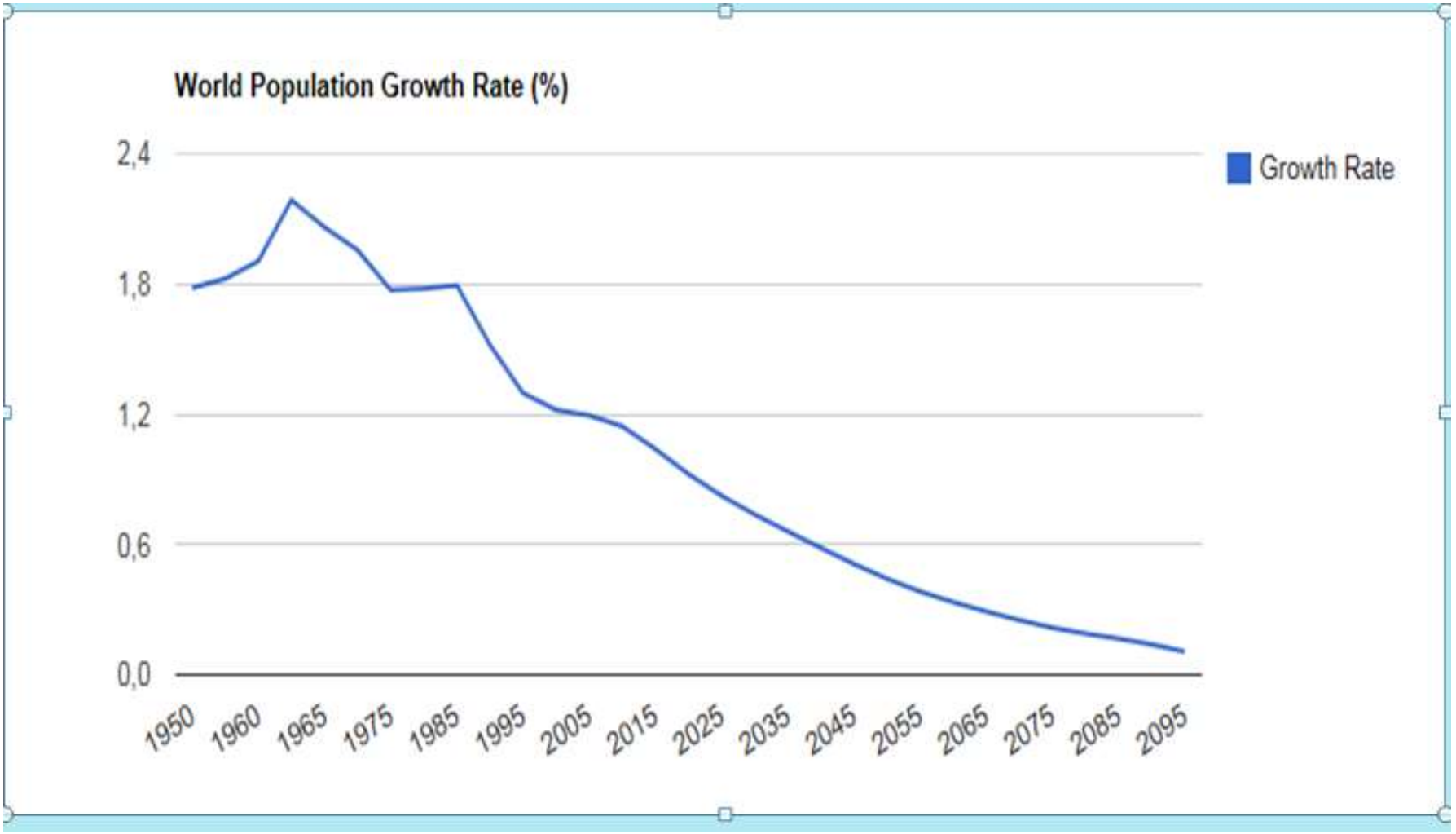
% of population aged 65+ years

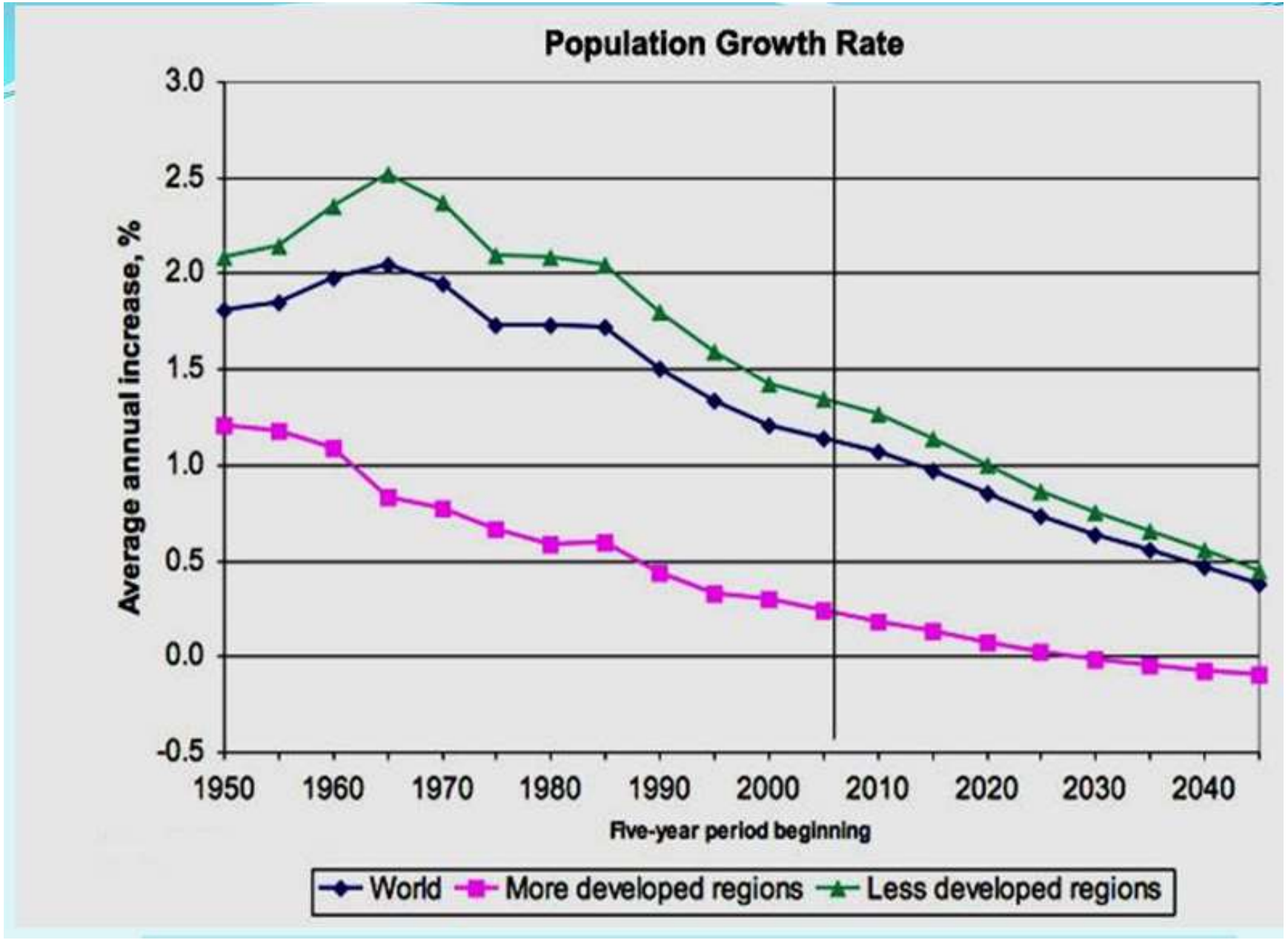


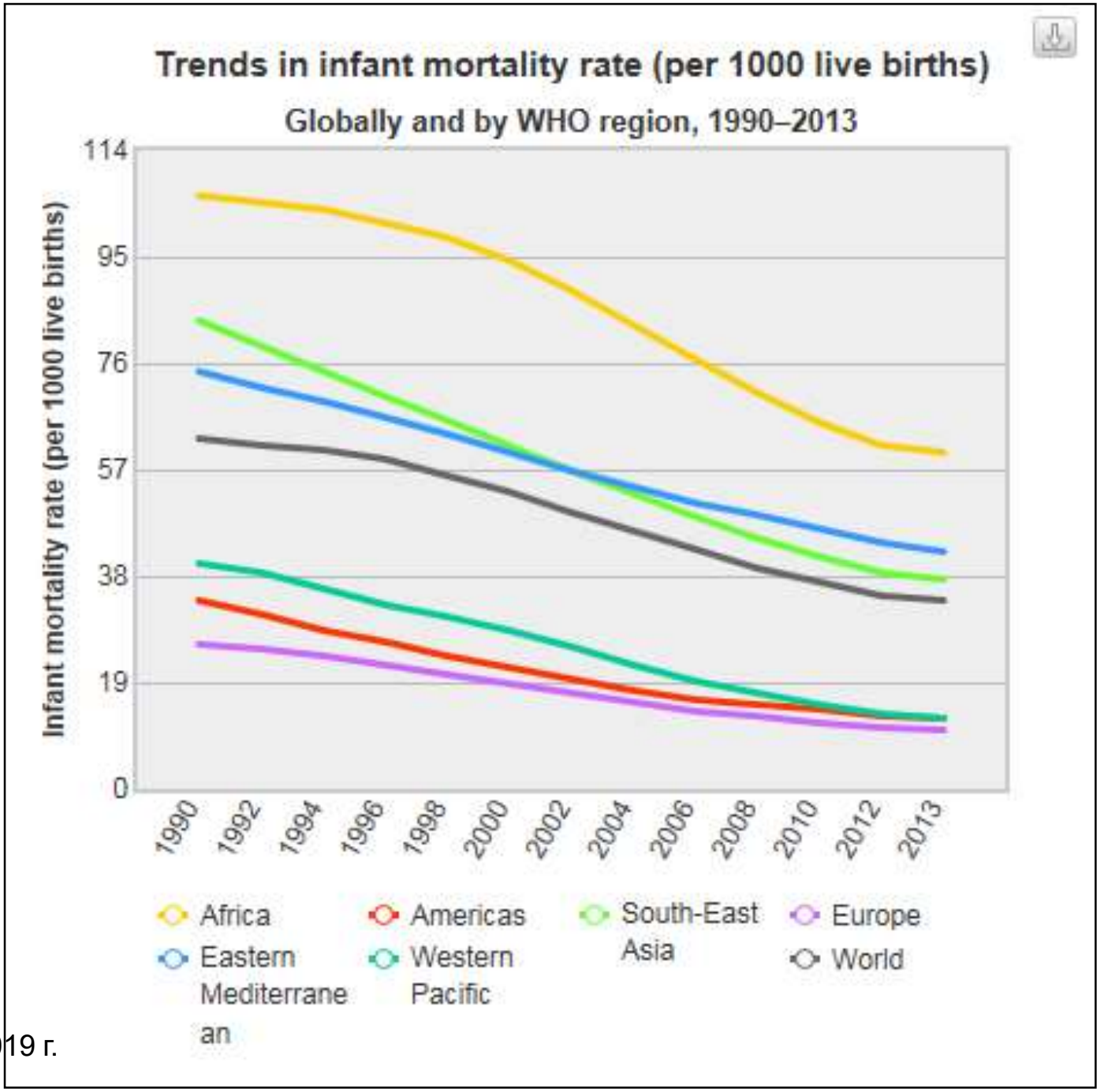
Last available

The linear charts are appropriate to express changes over time.


World Population: 1950-2050
Population (billions)
0
1
2
3
4
5
6
7
8
9
10
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
Year
3 Billion
4 Billion
5 Billion
6 Billion
7 Billion
8 Billion
9 Billion
Source: U.S. Census Bureau, International Data Base, August 2016 Update.

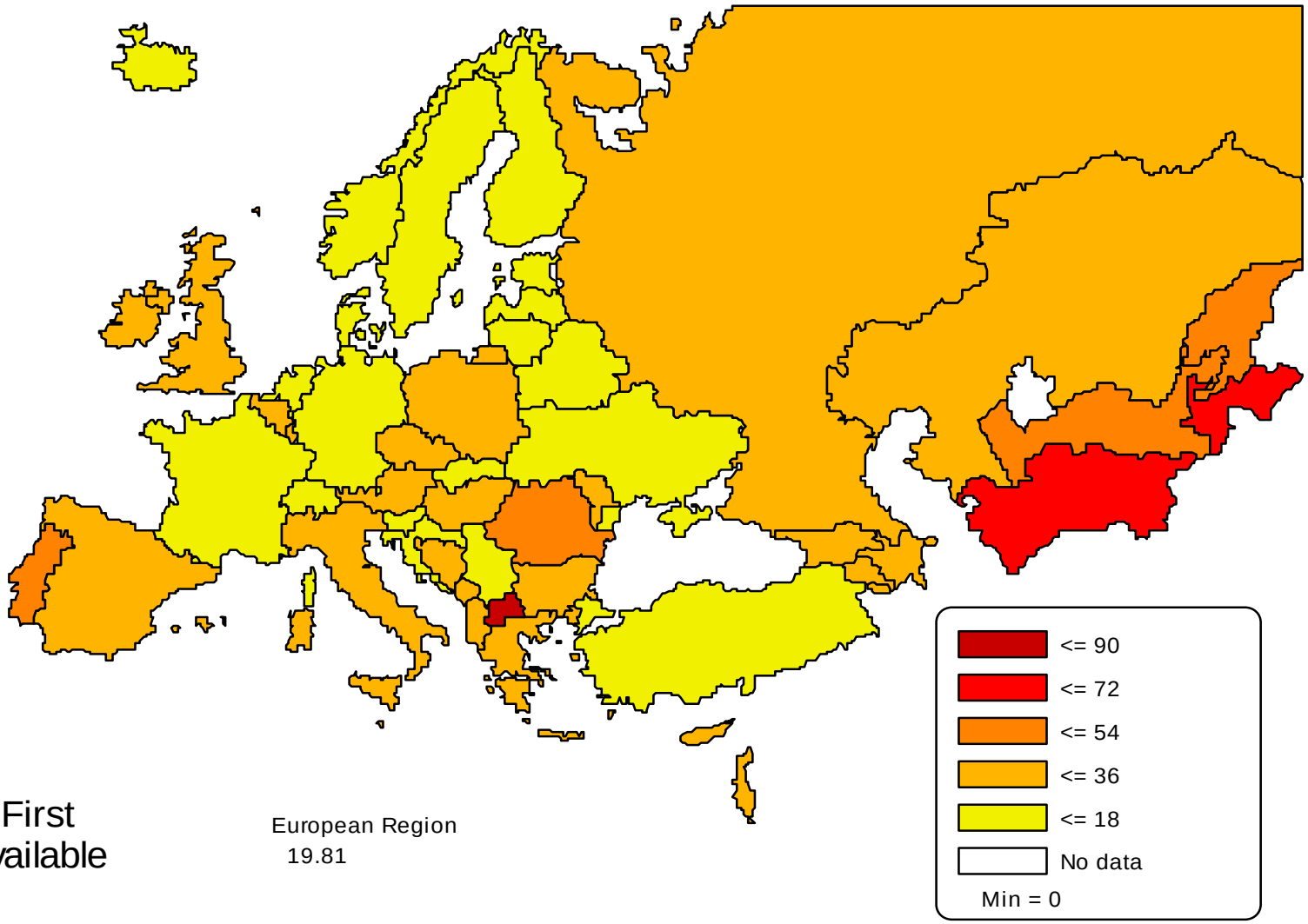

World Population Growth Rate (%)
Growth Rate
2,4
1,8
1,2
0,6
0,0
1950
1960
1965
1975
1985
1995
2005
2015
2025
2035
2045
2055
2065
2075
2085
2095


Population Growth Rate
Average annual increase, %
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
Five-year period beginning
World
More developed regions
Less developed regions


Trends in infant mortality rate (per 1000 live births)
Globally and by WHO region, 1990–2013
Infant mortality rate (per 1000 live births)
114
95
76
57
38
19
0
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2013
Africa
Americas
South-East Asia
Europe
Eastern Mediterranean
Western Pacific
World

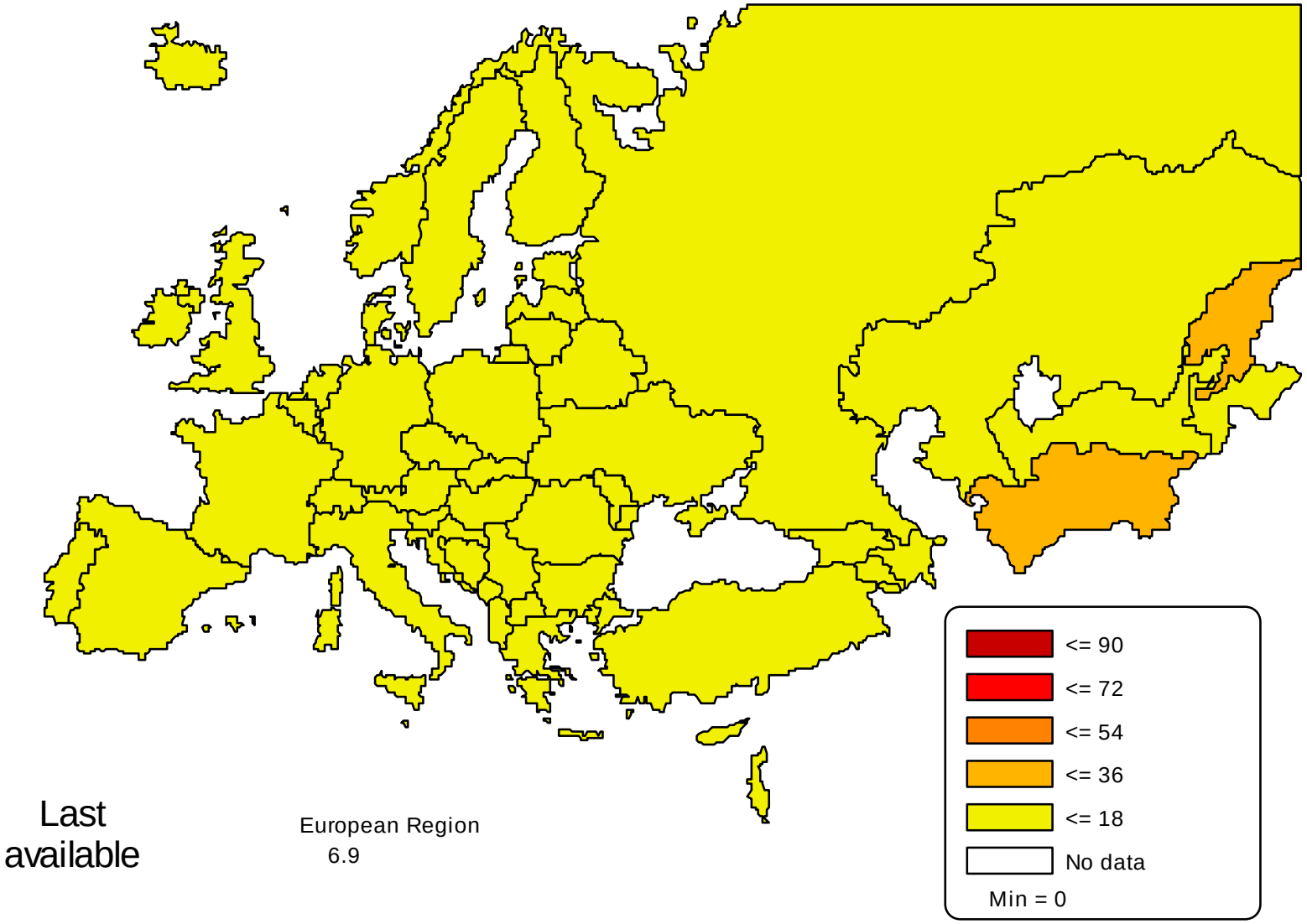
The maps are appropriate to express different levels of rates in different region.

Infant deaths per 1000 live births



Source: European Health for All database

Infant deaths per 1000 live births



Source: European Health for All database

Test examples related to Lecture 1

1. ***The basic idea underlying sampling is to select a representative sample, from which the investigator can make inferences to the population.***

A. True **B.** False

2. ***A sample is said to be random when it is not representative of the population.***

A. True **B.** False

3. ***A random sample is one in which 50% of the elements of a population have equal chances of being sampled.***

A. True **B.** False

4. *A representative sample:*

A. Consist of at least 500 cases

B. Must be a random sample

C. Is defined as the inverse of the square root of the sample size

D. Reflects precisely the crucial dimensions of a population

5. *Stratified random sampling involves the selection of the most accessible elements of the population.*

A. True **B**. False

6. If a sample is representative, then it yields valid data for making generalizations about the population from which it was drawn.

A. True **B.** False

7. ***If a population contains 50% males and 50% females, and our sample 10% males and 90% females, then our sample is said to be biased.***

A. True **B.** False

8. ***If a sample is large (say n>500) then the sample must be representative.***

A. True **B.** False

9. ***A sample size increases:***

A. The sampling error decreases

B. The population become more accessible

C. The sample becomes more biased

10. ***The larger the sample size, the larger the sampling error.***

A. True **B.** False

11. ***If the sample size is halved, the sampling error will be doubled.***

A. True **B.** False

12. ***When you take a patient's blood pressure daily, you are in fact sampling from a population of potential blood pressure readings.***

A. True **B.** False

13. Say that it is known that coronary disease occurs twice as frequently among males as females and three times more commonly among over 50 year-olds than those under 50. Given a stratified sample of 120, how many subjects would you expect to be female and under 50?

A. 60 **B.** 40 **C.** 30 **D.** 10

14. A random sample is one in which:

A. All the elements had an equal chance of selection

B. A chance method was used to select the elements included in a sample

C. Both statements are true

15. ***Within the following statements select one which you believe to be true. An ordinal variable is one for which:***

A. The data are discrete and can take one of many values

B. The data are continuous and follow an ordered sequence

C. The categories of response are ordered

16.*Which of the following measures of the variable 'weight' is nominal?*

A. Weight in kg

B. Weight as obese/overweight/normal/underweight/grossly underweight

C. Weight as 'normal against pathological' (obese or grossly underweight

D. Weight as percentage overweight in relation to "healthy"" weight

17. ***Select one of the following variables measured on a nominal scale.***

A. Height in cm

B. Ethnic group

C. Education categorized as primary school, secondary school, bachelor degree

D. Age in years

18. ***The readings ’64 kilograms” is a value on a(n):***

A. ratio scale

B. interval scale

C. ordinal scale

D. nominal scale

19. A sample by convenience is always representative.

A. True **B.** False

20. ***An interval scale has an absolute zero.***

A. True **B.** False

21. The levels of measurements which have equal intervals are the ordinal and nominal scales.

A. True **B.** False

22. ***Ordinal scales are generally preferable to interval scales.***

A. True **B.** False

23. ***Ordinal measures involve rank-ordering the values of a variable.***

A. True **B.** False

24. ***Nominal scales do not have the characteristic of 'distinctiveness' (categories).***

A. True **B.** False

25. ***The gender of patients is an example of a(n):***

A. ratio scale

B. nominal scale

C. ordinal scale

D. interval scale

26. "The tenth" is a value on a(n):

A. ratio scale **B.** interval scale

C. ordinal scale **D.** nominal scale

27. If a population contains 50% males and 50% females, and a sample 10% males and 90% females, then such sample is said to be biased.

A. True **B.** False

28. In a patient record system, patients are randomly assigned a unique identification number. These numbers represent a (n):

A. nominal scale B. ratio scale

C. interval scale D. ordinal scale

29. An auto analyst is conducting a satisfaction survey, sampling from a list of 10,000 new car buyers. The list includes 2,500 Ford buyers, 2,500 GM buyers, 2,500 Honda buyers, and 2,500 Toyota buyers. The analyst selects a sample of 400 car buyers, by randomly sampling 100 buyers of each brand. What is the sample type?

A. Simple random sample

B. Stratified random sample

C. Systematic random sample

30. Which of the following statements are true? (Check one)

A. Categorical variables are the same as qualitative variables.

B. Quantitative variables can be continuous variables.

C. Both statements are true

31.10.2019 г.

Answers:

1-A	2-B	3-B	4-D	5-B	6-A	7-A	8-B
9-A	10-B	11-B	12-A	13-D	14-C	15-C	16-C
17-B	18-A	19-B	20-B	21-B	22-B	23-A	24-B
25-B	26-C	27-A	28-A	29-B	30-C		