MODULE 8:

ANALYSIS, INTERPRETATION, AND PRESENTATION OF MALARIA DATA

MODULE 8: ANALYSIS, INTERPRETATION, AND PRESENTATION OF MALARIA DATA

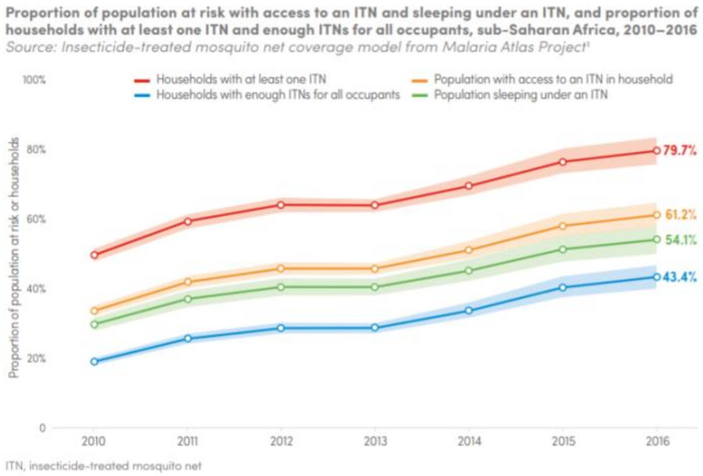
This module provides an introduction to the functions of and common concepts for data analysis and interpretation. It also covers how to present data effectively, walking learners through how to select an appropriate graphic for presenting data and providing tips for good data presentation.

Module Objectives

By the end of this module, you will be able to:

- Identify the functions of data analysis and interpretation.
- State common concepts for data analysis and interpretation.
- Identify appropriate graphics for presenting various types of data.
- Differentiate between the characteristics of a good data presentation and a poor data presentation.

Figure 20. Example of a graphic display of malaria data



Source: World Malaria Report, 2017

What Is Data Analysis and Interpretation?

Data analysis refers to the process of inspecting, cleaning, transforming, and modeling data for the purpose of highlighting useful information. In other words, data analysis is taking the raw ***data*** that you have collected for your program and transforming them into ***information*** that can be used for decision making.

Data analysis helps provide answers to questions being asked about your program or other research studies. It means taking the data that you collect and looking at them in comparison to the questions you would like to answer. For example, if you want to know whether your program is meeting its objectives—or if it is on track—you would look at your program targets and compare them to the actual program performance.

Data interpretation goes a step beyond data analysis. Interpretation is using the analysis to further understand your findings and the implications for your program. It is the process of adding meaning to information by making connections and comparisons, by exploring causes and consequences, and by explaining patterns or trends observed within the data.

Quiz Question

Which question can be answered through data analysis?

a. In which site did the malaria program achieve greater coverage of insecticide-treated nets?
b. Why did the malaria program fail to achieve its target of 80 percent coverage of insecticide-treated nets in Community A?

Correct answer:

a. This question can be answered through data analysis, comparing the coverage of insecticide-treated nets in each of the different program sites.

Analysis of Malaria Data

You can perform a number of different types of analyses at the program level. The type of analysis will depend on the information you need to manage your program effectively and track its progress. Here are three examples of common types of analyses that will help you to track your program's progress:

- **Actual performance compared to program target:** In this example, we are comparing our program's actual performance in terms of the number of persons trained on malaria case management to the target that was set for the program by June 1, 2016. In this case, we see that our progress over the first year of the program (from January 2015 to December 2015) is not progressing at the pace we would expect to meet the program target of 100 persons trained on malaria case management by June 2016. For the program to meet the target, an additional 85 people need to be trained in the next six months of the program.
- **Current performance compared to previous performance:** In this example, we are comparing our program's current performance (2015) in terms of the number of insecticide-treated nets (ITNs) that were distributed during the previous three years (2012–2014). As we can see for 2015, the program is outperforming all previous three years by a substantial number. It is clear that the current performance is better than previous performance for this program.
- **Comparison of performance between different sites or groups:** In this example, we are comparing the performance of the program in District A to the performance of the program in District B in terms of the number of fever cases tested for malaria in the health clinics. We can see that District B has tested 8,000 fever cases for malaria, compared to only 3,500 in District A. It appears that District B is performing better than District A, but it is hard to interpret these results without further information about the two districts. For example, we would also want to know the different target populations in each of the districts, as well as more information on the malaria endemicity in each district to see whether they are comparable.

Common Measures for Analysis

Some basic statistical measures that are important to understand, calculate, and interpret include the following:

- Measures of central tendency
    - Mean
    - Median
    - Mode
- Measures of variation
    - Range
    - Variance and standard deviation
    - Interquartile range (IQR)
- Ratio, rate

The following sections provide explanations and examples for each of the measures.

Measures of Central Tendency

These are some basic statistical measures that are important to understand, calculate, and interpret.

Mean

The most commonly investigated characteristic of a data set is its center, or the point around which the observations tend to cluster. The mean is the most frequently used measure to look at the central values of a data set.

Definition: The sum of the values divided by the number of cases or observations. It is also referred to as the average.

Calculation: Mean = sum of values / number of observations

Example: What was the mean number of malaria cases per month during the past year?

- Sum of malaria cases (January–December): 1,110
- Number of observations: 12
- Mean: 1,110 / 12 = 92.5 average number of malaria cases per month in the past year

Median

Definition: The median is the middle value in an ordered set of values.

Calculation: The first step is to sort the data from the lowest value to the highest value. The second step is to choose the middle observation within the data set. For data sets with an even number of values, the median is the average (mean) of the two middle values.

Example 1: What is the median number of malaria cases per month during 2013?

1. Sort the observations.
2. Select the two middle numbers because there are an even number of observations in the data set.
   Middle numbers: 45 and 45
3. Add the two numbers and divide by two: (45+45)/2=45.

Example 2: What is the median number of malaria cases per month during 2014?

1. Sort the observations.
2. Select the middle number, because there are an odd number of observations in the data set. This number represents the median, which in this case is 49.

Mode

Definition: The mode is the value that occurs most frequently in your data set.

Calculation: Select the value in your data set that occurs most frequently.

Example: What is the mode for the number of malaria cases in 2013 and 2014?

- Mode for number of malaria cases in 2013: 45
- Mode for number of malaria cases in 2014: 40

Of the three measures of central tendency—mean, median, and mode—the mean is the most frequently used measure to look at the central values of a data set. The mode is the least useful and thus, the least used measure of the three.

Measures of Variation

These are some basic statistical measures of variation that are important to understand, calculate, and interpret.

Range

Definition: The range represents the difference between the highest and lowest values within your distribution (data set).

Example: What is the range for the number of malaria cases in 2013 and 2014?

- Range for number of malaria cases in 2013: 36–69
- Range for number of malaria cases in 2014: 35–64

Variance and Standard Deviation

Definition of variance: Variance is a measure of how far a set of numbers are spread out from each other. It helps describe how far the numbers lie from the mean.

Calculation of variance: Variance (s2) is the sum of the squared deviations from the mean divided by the number of observations minus 1.

Definition of standard deviation: Standard deviation is a measure that shows how much variation there is from the mean. A low standard deviation indicates that the data points tend to be very close to the mean, and a large standard deviation indicates the opposite—that the data are spread out over a large range of values.

Calculation of standard deviation: The standard deviation (s) is the square root of the variance.

Interquartile Range

Definition: The IQR is a measure of statistical dispersion. It is equal to the difference between the third and first quartiles, and thus represents the middle 50 percent of the data. Quartiles divide data into four equal groups, with the lower quartile (Q1) being the 25th percentile, the middle quartile being the 50th percentile, and the upper quartile being the 75th percentile (Q3). Since the IQR uses the middle 50 percent of the data, it is not affected by outliers or extreme values.

Calculation: IQR=Q3-Q1

Example:

- Q3=42
- Q1=18
- IQR=42-18=24

Ratio

Definition: A ratio is a comparison of two numbers, expressed in one of the following ways: "a to b," "a per b," or "a:b."

Examples:

- Two household members per (one) mosquito net
- Women are slightly more likely to sleep under an ITN than men, with a ratio of 1.2:1.

Rate

Definition: A rate is a ratio between two measures. In public health, rate is a measure of the number of cases that occur in a given period, divided by the population at risk during that time period.

The comparison is often expressed as the number of occurrences per 1,000, 10,000 or 100,000 population. Rate is a probability statement, most often used in public health to describe infrequently occurring events, such as maternal mortality, because it is easier to express "8 per 100,000" rather than ".00008 percent." The under-five mortality rate is the probability, expressed as a rate per 1,000 live births, of a child born in a specified year dying before reaching age five at the current age specific mortality rates.

Examples of rates used in malaria:

- **Annual parasite incidence (API)**

$$\frac{\textbf{\# of confirmed malaria cases in 1 year}}{\textbf{Population under surveillance}} \times \textbf{1,000}$$

Proportion

Definition: A ratio in which all individuals in the numerator are also in the denominator.

Example:

- Three staff members per clinic is a ratio expressed numerically as 3:1. It is not the same as saying 1 to 3 or 1:3. The order of the numbers matters.
- A clinic has 12 female clients and 8 male clients, the denominator is total clients, 20, and the ratio of male clients is 8 to 20, or 8:20.

Percentage

Definition: A proportion of the nominator, or part of the whole, multiplied by the denominator, or 100, used to compare data across facilities, regions, and countries.

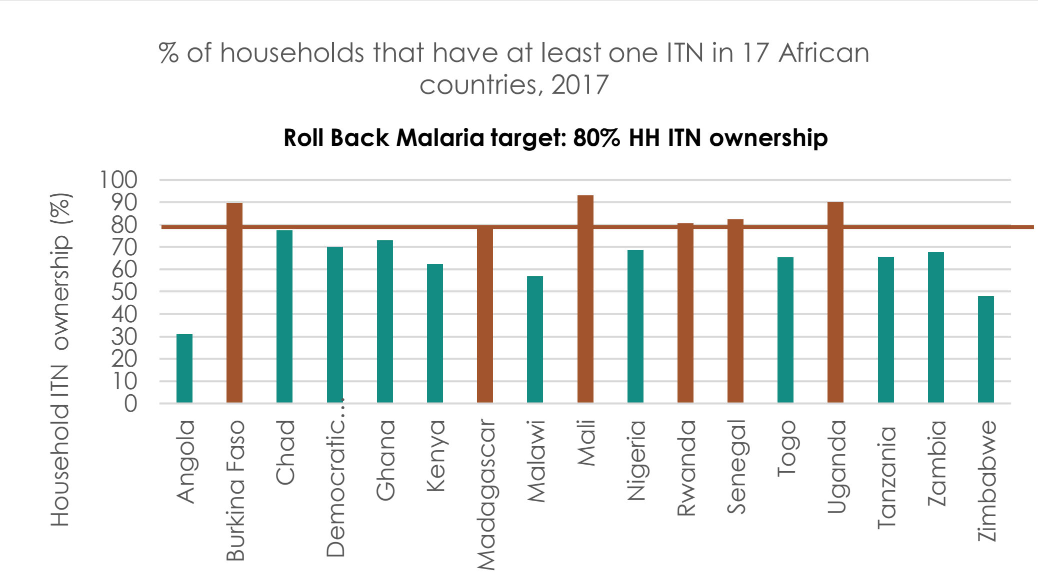
Example:

A clinic has 12 female clients and 8 male clients, which are the numerators, the denominator is 20, the total number of clients. The proportion of male clients is eight-twentieths or two-fifths. To state this as a percentage, convert the fraction to a decimal, 0.4, and then multiply by 100, which equals 40 percent. In this example, the denominator includes all clients, both male and female.

Data Interpretation

After data have been analyzed, the next step is interpreting the information. Interpretation is making sense of your information and understanding the implications for your program. In many cases, additional information is needed for people to be able to understand your findings. In Figure 21, showing ITN ownership for children under five, the target is to have more than 80 percent of children under five sleep under an ITN every night. From the graphic, we observe that this goal was achieved by five countries: Burkina Faso, Mali, Rwanda, Senegal, and Uganda.

Figure 21. Example of a graphic showing the implications of a malaria data set



Interpreting Graphs

In interpreting the graph, there are a few questions we might want to understand. For example:

- Why have some countries not met the target goal of greater than 80 percent ITN usage among children under five?
- What are the reasons behind countries not meeting the target goal? Is it because the program is not distributing enough ITNs? Is it because the health education and behavior change communication interventions are not effective?

We might also want to know why some countries are doing better than others in terms of progress toward the goal. For example, what is different about Country 2 compared to Country 1?

To understand the findings, we need to look at other relevant data that can help us answer our questions. In this case, we could look at household ownership of ITNs or at the coverage and effectiveness of health education and behavior change communication interventions in the countries. Sometimes conducting further analyses is needed to interpret the findings.

Challenges with Interpretation of Common Malaria Indicators

A number of issues need to be taken into consideration when you are interpreting findings for common malaria indicators. These issues can pose a challenge in understanding what your program findings mean, and, in many cases, other data may need to be collected to better understand your findings. Two of the main issues to consider when interpreting malaria-related data are seasonality and malaria endemicity.

Seasonality

In many places, malaria transmission is seasonal, meaning that transmission occurs or occurs more frequently during certain times during the year and not necessarily year-round due to changing climatic conditions. Given that transmission can fluctuate dramatically within a location within the year or across different years, it is important to take into account the season in which your data were collected.

Intervention coverage and usage levels for the four main malaria prevention interventions (ITN/long-lasting insecticide-treated net [LLIN], indoor residual spraying, intermittent preventive treatment in pregnancy [IPTp], and receiving prompt and effective treatment) may differ slightly between seasons. For example, during the rainy season, more people will likely sleep under an ITN/LLIN because they perceive the threat of getting malaria to be higher because there are more mosquitoes.

Seasonality will also affect malaria morbidity and mortality. We expect malaria morbidity and mortality to be higher during and for several weeks after the rainy season because there is higher malaria transmission. Thus, interpretation of your data must take into account the season in which the data were collected.

Malaria Endemicity

Malaria endemicity can affect the interpretation of the core malaria indicators, because each involves the definition of the target population. In other words, each indicator is intended to be measured only among the target population, which is defined as those who are at risk for malaria.

In countries where malaria is endemic or epidemic-prone throughout the country, this issue should not be a particular concern when interpreting the findings. In countries where malaria endemicity varies within the country and thus not all of the population is at risk for malaria, it is important to take this into consideration when you are collecting data and interpreting your findings.

Populations that are not at risk for malaria should not be included in your target population. In these situations, it may be necessary to collect additional information to establish which areas are within or outside of a malaria risk area. This is not always possible, however, and it needs to be taken into consideration when you interpret your findings.

For example, if you collect data at the national level on household ownership of ITNs/LLINs, you need to consider whether you are including data from non-malarious areas. If you do include data from non-malarious areas, then it is likely that your data will underestimate the national coverage level of ITNs/LLINs. This is because you have overestimated your target population (those actually at risk for malaria).

Effective Data Presentation

Regardless of the type of communication format you are using, whether it is an annual progress report or a presentation, the information should be presented in a clear and concise way with key findings and recommendations that are actionable. Presenting data in this way helps facilitate their use for decision making. When data are presented in an unclear manner, or when too much information or irrelevant information is provided, then there is less likelihood that the information will be used for programmatic decision making. It is also important to always remember who your audience is when you are thinking of how to present information. This means that you should tailor the information presented to your audience, so that it is useful, clear, and actionable to them.

Summarizing Data: Table

There are two main ways to summarize and present data: tables and graphs. Both of these forms are useful for conveying a message and for portraying trends, relationships, and comparisons.

A table is the simplest way of summarizing a set of observations. It has rows and columns containing data, which can be in the form of absolute numbers or percentages, or both. In the example table, the number of deaths is listed for years 2010–2016, for countries that are considered to have a low burden of malaria transmission. This table allows you to see the trend in the number of malaria deaths over seven years within regions and worldwide.

Figure 22. Example of presenting malaria data in a table

Estimated number of malaria deaths by WHO region, 2010–2016 *Source: WHO estimates*

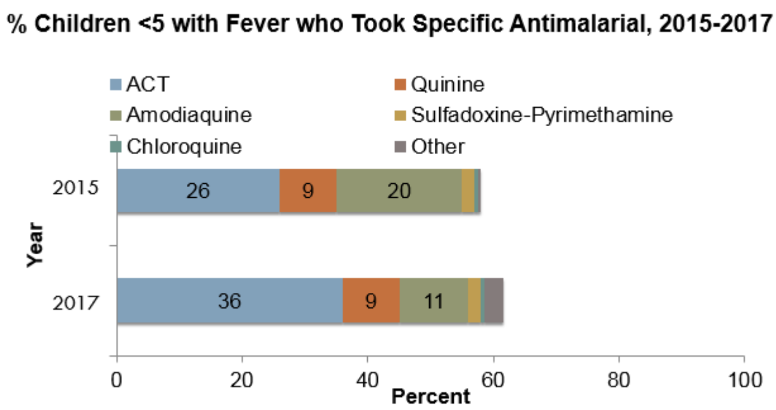
	Number of deaths						
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
African	538 000	484 000	445 000	430 000	423 000	409 000	407 000
Eastern Mediterranean	7 200	7 100	7 700	7 800	7 800	7 600	8 200
European	0	0	0	0	0	0	0
Americas	830	790	630	620	420	450	650
South-East Asia	41 700	34 000	29 000	22 000	25 000	26 000	27 000
Western Pacific	3 800	3 300	4 000	4 300	2 900	2 600	3 300
World	**591 000**	**529 000**	**487 000**	**465 000**	**459 000**	**446 000**	**445 000**

Source: World Malaria Report, 2017

Summarizing Data: Chart

Graphs are pictorial representations of numerical data and should be designed to convey a pattern or trend of the data. For example, in Figure 23—a stacked bar chart—we are able to compare the use rates of different antimalarial treatments among children under five with a fever within Country X over two years to see the changes in uptake of treatment, as well as changes in the type of treatment.

Figure 23. Example of a bar chart showing malaria treatment trends over time

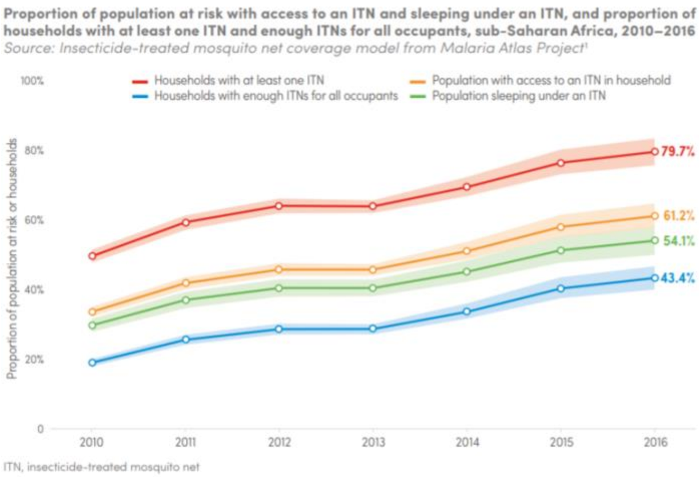


Source: World Malaria Report, 2017

Summarizing Data: Graph

In this line graph, there are two main messages being conveyed. First, we are able to see a trend in the number of people with access to ITNs over the past seven years globally. Overall, the number of people with access to ITNs has increased over the years.

Figure 24. Example of a line graph showing trends in access to ITNs



Source: World Malaria Report, 2017

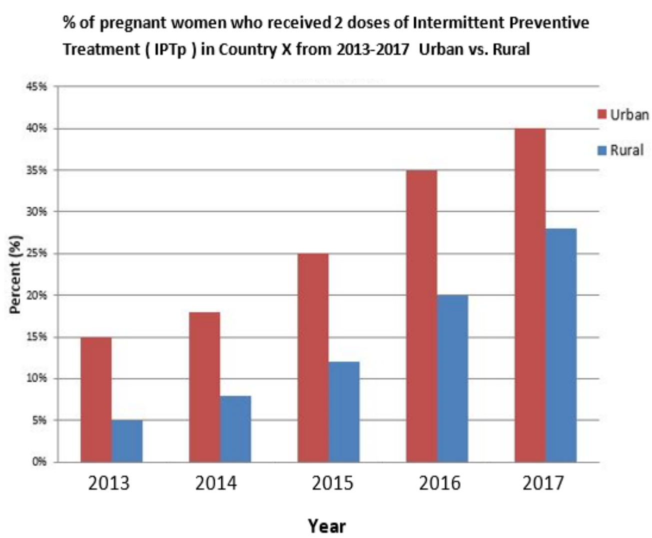
Presenting Data

There are four main charts and graphs used to present data. Each chart and graph has a specific use or function, which is important to keep in mind when you are deciding the best way to present your data.

Bar Chart

A bar chart is used to compare data across categories. The chart has rectangular bars with lengths that are proportional to the values they represent. They are used to plot data that have discrete values and are not continuous. In the example, we are able to compare the percentage of women who received two doses of IPTp in Country X across the five year time span as well as compare the differences between pregnant women in the rural areas versus the urban areas in terms of access to IPTp.

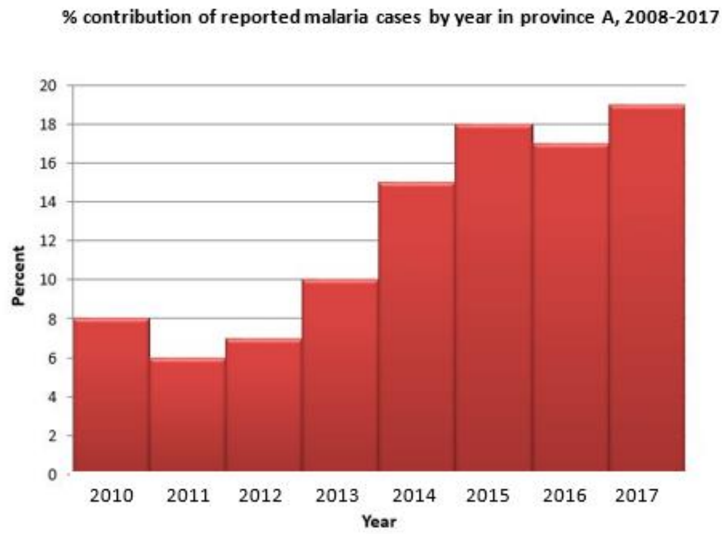
Figure 25. Example of a bar chart for comparison of malaria data by category



Histogram

A histogram represents the relative frequency of continuous data. In other words, it is a graph that shows a visual representation of the distribution of data. It consists of tabular frequencies, shown as adjacent rectangles, that are positioned over discrete intervals, whose area is equal to the frequency of the observations in the interval. The total area of the histogram is equal to the number of data. In the example here, we see the distribution of malaria cases by year from 2010 to 2017 in Province A.

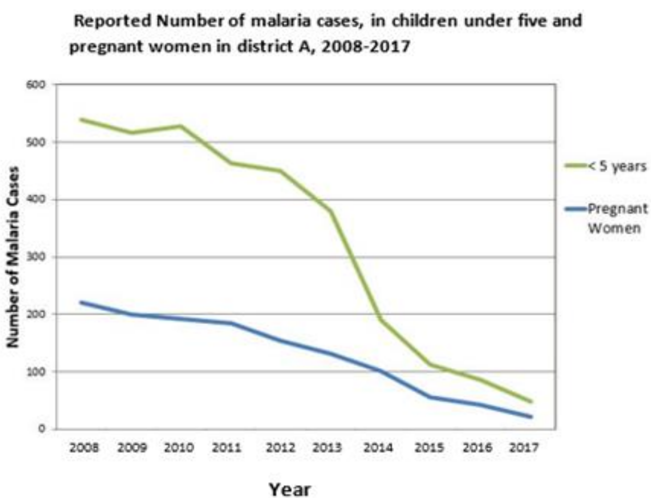
Figure 26. Example of a histogram showing the distribution of malria cases by year



Line Graph

A line graph is a graph that displays a trend or trends over time for continuous data. In the example, we are able to observe the decreasing trend in the number of malaria cases over the past 10 years for both children under five and pregnant women in District A.

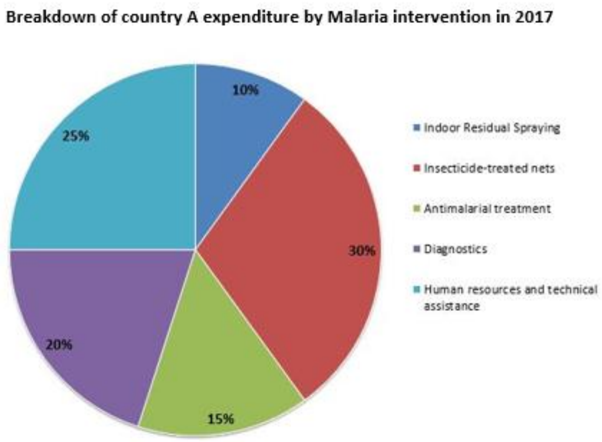
Figure 27. Example of a line graph showing decreasing trends in the number of malaria cases



Pie Chart

A pie chart is a circular chart that is divided into sections that represent the proportion or contribution of each value to a total. The size of the section is proportional to the quantity it represents. In this example, we are able to observe the relative proportion of Country A expenditure by the type of malaria intervention. Thus, we can see that the greatest proportion of the budget in the country is spent on ITNs (30%), and only 10 percent of the country's budget is spent on indoor residual spraying.

Figure 28. Example of a pie chart showing a breakdown of a country's malaria spending

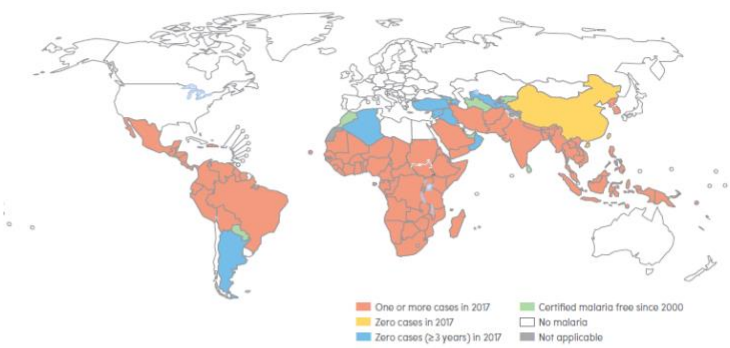


Maps

On a map, the data can be represented according to their geographical distribution. For example, we can create a map showing the type of malaria by zone (endemic or epidemic) or a distribution in percentages of vaccination coverage in a region or a country.

Figure 29. Example of a map showing geographic distribution of malaria worldwide

Malaria Incidence in the World



World Malaria Report, 2018

Presenting Data Quiz

To complete the quiz, select an answer for each statement. *Correct answers are provided on the next page.*

It is important to remember that each of the four main charts and graphs have specific uses for presenting data. To check your understanding of their uses, read each statement and decide which chart or graph would be the most appropriate to use to convey the information. Select from these options: bar chart, histogram, line graph, pie chart.

a. The prevalence of malaria in Ghana over the past 30 years

b. Data that are comparing the prevalence of malaria in 10 different countries during one year in sub-Saharan Africa

c. Data on the reported reasons why individuals do not use insecticide-treated nets (ITNs) among the individuals surveyed who were not currently using ITNs

d. The distribution of patients tested for malaria by parasite density

Correct answers:

a. The prevalence of malaria in Ghana over the past 30 years**—Line graph**

A line graph is the most appropriate way to show the prevalence of malaria in Ghana over the past 30 years, because it can visually convey the trend for the 30-year time span in the country.

b. Data that are comparing the prevalence of malaria in 10 different countries during one year in sub-Saharan Africa**—Bar chart**

A bar chart is the most appropriate chart to use to be able to make the comparisons in prevalence of malaria across the 10 different countries.

c. Data on the reported reasons why individuals do not use insecticide-treated nets (ITNs) among the individuals surveyed that were not currently using ITNs**—Pie chart**

A pie chart is the most appropriate chart to visually show the reported reasons why individuals do not use ITNs, in relative proportion to one another.

d. The distribution of patients tested for malaria by parasite density**—Histogram**

A histogram is the most appropriate graph to show the distribution of patients tested for malaria by parasite density.

Key Tips for Good Data Presentation

Here are some key tips to remember to ensure good data presentation:

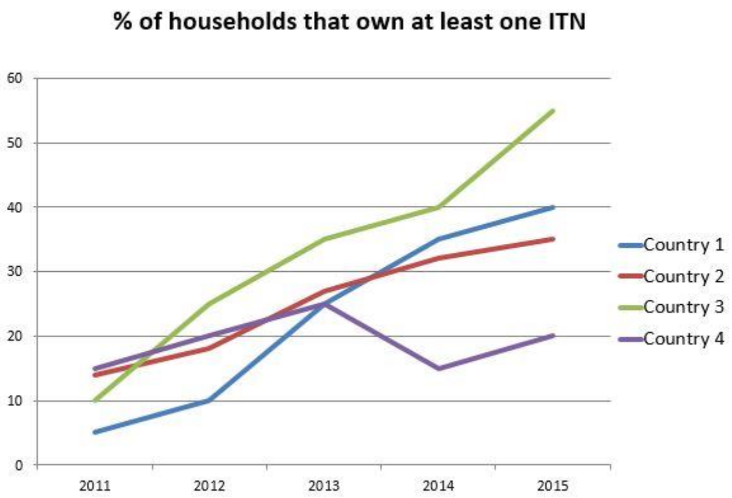
- ***Use the appropriate graphic*—**Make sure to use the right graphic (table, chart, or graph) for your data, which will be based on the message you want to convey.
- ***Know your audience*—**Present your information in a way that is clear, concise, practical, and actionable.
- ***Label all the components of your graphic*—**All graphs and tables should have a title. The title should express the who, what, when, and where. Graphs and charts should have clearly labeled axes, and when appropriate, should include legends.
- ***Provide all relevant information*—**Your graphic should be self-explanatory. This means that you should include the source and date for the data presented in your graphic. If something needs clarification or further explanation, include a footnote that helps bring clarity to the graphic. For good data presentation, it is essential that your audience has all the information needed to understand the message being conveyed in the graphic.

Data Presentation Examples

Keeping in mind the four tips for good data presentation, take a look at the following graphics. Are they presented well? If not, what is wrong with the way in which they are presented?

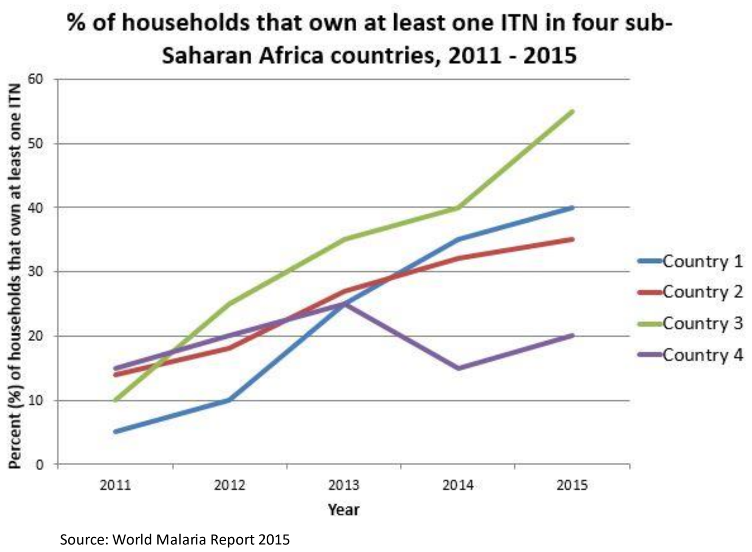
Example 1

What are the ways in which the data on this line graph could be presented more effectively?



A better way to represent the data includes the following elements:

- The title of the graph includes information on "when" and "where."
- The axes are properly labeled.



Example 2

What are the ways in which the data in this table could be presented more effectively?

Table 1

Year	(n)	Relative frequency (%)
2008	4,216,531	8
2009	3,262,931	6
2010	3,319,339	7
2011	5,338,008	10
2012	7,545,541	15
2013	9,181,224	18
2014	8,926,058	17
2015	9,610,691	19
Total	51,400,323	100.0

A better way to represent the data includes the following elements:

- The title explains the data contained in the table.
- The columns are clearly labeled.
- The data source is included.

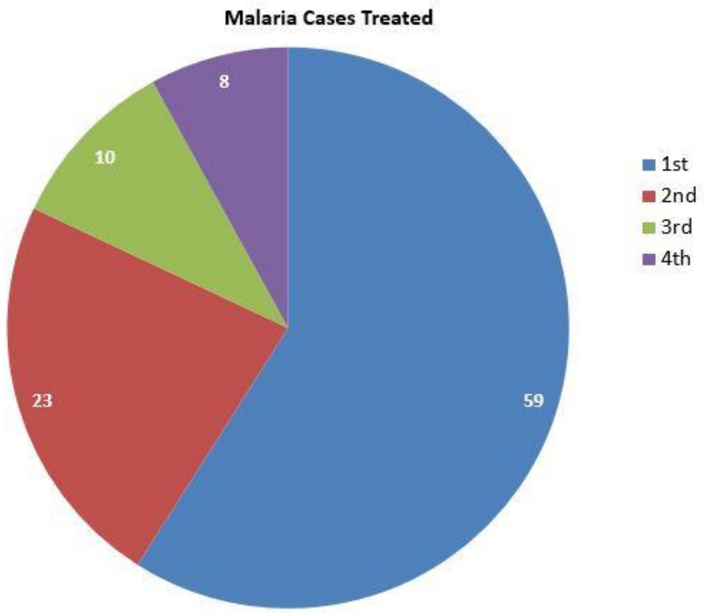
Table 1: Percent contribution of reported malaria cases by year in Country A, 2008 – 2015

Year	Number of malaria cases (n)	Relative frequency (%)
2008	4,216,531	8
2009	3,262,931	6
2010	3,319,339	7
2011	5,338,008	10
2012	7,545,541	15
2013	9,181,224	18
2014	8,926,058	17
2015	9,610,691	19
Total	51,400,323	100.0

Source: World Health Organization

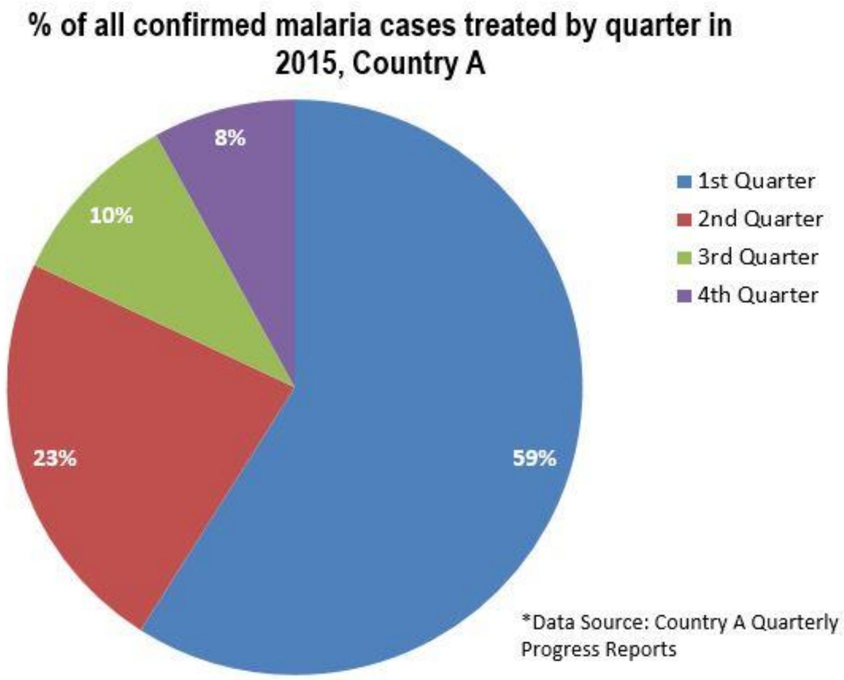
Example 3

What are the ways in which the data in the pie chart could be presented more effectively?



A better way to represent the data includes the following elements:

- The title of the chart is clear and specific.
- The numbers on the chart are clearly labeled.
- The chart legend is descriptive.
- The data source is included.



Module 8 Assessment

Questions

Correct answers are provided on the next page.

1. Which of the following questions cannot be answered through simple data analysis, but would need further information and interpretation to be able to answer?
   a. What was the annual number of malaria admissions in District A health facilities in 2015?
   b. Which country has the highest coverage (proportion) of households that own at least one insecticide-treated net?
   c. Why did the proportion of people sleeping under insecticide-treated nets decrease dramatically from last year?
   d. Which health facility in District B provided greater coverage of intermittent preventive treatment for pregnant women?

2. Based on the data presented in the table for annual number of insecticide-treated nets distributed by the National Malaria Control Program in Country X from 2002 to 2010, what is the mean, the median, and mode for the data set (in that order)?
   a. Mean: 140, Median: 174.67, Mode: 135
   b. Mean: 174.67, Median: 140, Mode: 135
   c. Mean: 135, Median: 140, Mode: 174.67
   d. Mean: 175, Median: 135, Mode: 140

Annual number of insecticide-treated nets distributed by the national malaria control program in Country X, 2007–2015

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015
# of ITNs Distributed (thousands)	125	135	140	132	135	150	155	250	350

3. If you wanted to present a graphic comparing the proportion of households that have been sprayed with insecticides within the past 12 months across 15 high-burden countries, which would you use?
   a. Table
   b. Pie chart
   c. Histogram
   d. Bar chart

4. Annual parasite incidence, which is defined as the total number of confirmed malaria cases during one year (X 1,000) divided by the total population under surveillance, represents which common statistical measure?
   a. Rate
   b. Proportion
   c. Mode
   d. Percentage

Correct Answers

Correct answers are noted in bold.

1. Which of the following questions cannot be answered through simple data analysis, but would need further information and interpretation to be able to answer?

   c. Why did the proportion of people sleeping under insecticide-treated nets decrease dramatically from last year?

   A simple data analysis will only tell you that the proportion of people sleeping under insecticide-treated nets decreased this year compared to last year. To understand why, further information would need to be collected in order to determine why the proportion decreased.

2. Based on the data presented in the table for annual number of insecticide-treated nets distributed by the national malaria control program in Country X from 2002 to 2010, what is the mean, the median, and mode for the data set (in that order)?

   b. Mean: 174.67, Median: 140, Mode: 135

   Mean: 1,572/9 (observation)=174.67; Median: 125, 132, 135, 135, 140, 150, 155, 250, 350 = 140; Mode: 135

3. If you wanted to present a graphic comparing the proportion of households that have been sprayed with insecticides within the past 12 months across 15 high-burden countries, which would you use?

   d. Bar chart

   A bar chart is used for comparing data across different categories. In this case, you are comparing the difference in proportion of households that have sprayed insecticides across the 15 high-burden countries.

4. Annual parasite incidence, which is defined as the total number of confirmed malaria cases during one year (X 1,000) divided by the total population under surveillance, represents which common statistical measure?

   a. Rate

   Annual parasite incidence represents a rate, because a rate is ratio between two measurements. In this case, it is a ratio of the total number of confirmed malaria cases during one year (X 1,000) divided by the total population under surveillance.