Welcome to the final guide in the Mobile Data Solutions introductory resources. In the previous guides, you learned about the key elements needed for assessing whether a mobile data solution is appropriate and how to select devices and tools fitting the specific task, and conducting surveys using mobile devices. This guide will focus on data analysis, visualization and sharing.

Before continuing, it’s important to define a few key terms which will be discussed in each section of this guide:

**Data analysis** refers to the process of transforming raw data into information that is easier to understand using descriptive and inferential statistics, and other data analysis techniques.

**Data visualization** is the process of representing data in a schematic form with the goal of communicating information clearly and effectively through graphical means.

**Data sharing** is the process and practice of making data available to other users.

There are many data analysis and visualization software applications and numerous approaches to data sharing. This overview highlights some of these and raises issues to consider when evaluating options; however, it does not provide recommendations for which application or approach to select in a given situation.

**DATA ANALYSIS**

Datasets do not answer questions or “speak for themselves.” Until they are analyzed, data are “raw” and, for most people, indigestible. Information useful for decision making is extracted from the raw data through the process of data analysis. Depending on the type of data collection technique used, you may have produced datasets consisting entirely of numerical values (quantitative data), entirely words and narratives (qualitative data), or a mix. The nature of your data, whether it is quantitative or qualitative, is an important consideration in choosing the right data analysis application.

There are numerous qualitative data analysis software applications to choose from, both free and open-source products such as the **Coding Analysis Toolkit (CAT)** and commercial proprietary applications such as **Atlas.ti** and **NVivo**. A list of some of the additional software options designed for analysis of qualitative data can be found on the [American Evaluation Association](https://www.aea-web.org) website.

Quantitative data analysis includes descriptive statistics and inferential statistics. There are a number of software solutions designed to facilitate quantitative data analysis. These range from simple non-relational applications, like Excel and Google Spreadsheet, to complex quantitative statistical packages, such as SAS, SPSS and STATA. Some applications are cloud-based; some are free, while others are proprietary. Some software applications are used for the analysis of data for a specific field of study or purpose (such as health), while others are designed for general purpose. An excellent example for epidemiological data analysis and visualization is **Epi Info**, a free application developed by the Center for Disease Control (CDC).

Brief descriptions of the two statistical analyses for quantitative data—descriptive and inferential—and when to use the methods are provided below.
Descriptive statistics quantitatively describe the main features of a collection of data. Providing simple summaries about the sample and about the observations that have been made, descriptive statistics may be presented either as quantitative summary statistics, or visually, as easy-to-understand trend lines, bar charts or other graph types. Descriptive statistical tools allow us to calculate and present frequencies, percentages, proportions, measures of distribution (such as range, mean deviation, standard deviation), and measures of central tendency (such as mean, mode and median). Such statistical tools summarize the data collected using mobile solutions in small and easy to understand numbers. While these summaries may serve as an initial description of the data within a more extensive statistical analysis, they may be sufficient in and of themselves for a particular project.

Cross-tabulation (or crosstab) is an important descriptive statistical tool that displays the distribution of variables by tabulating the results in a table. Used for analyzing the relationship between two or more variables in a given data set, crosstabs are simple to compute, easy to understand, and provide comparative data. The following table provides an example of a cross-tabulation table for a hypothetical survey question asking the relevance of mobile solutions for mapping engendered species in the North Pole.

| Do you think mobile solutions are useful for mapping endangered species in the North Pole? |
|--------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                      | TOTAL | GENDER | AGE            |
|                                      | Adults | Men    | Women          | 18–30 | 31–40 | 41–50 | 51–60 | 61+ |
| YES                                  | 78%   | 76%    | 80%            | 90%   | 85%   | 80%   | 75%   | 60% |
| NO                                   | 9%    | 14%    | 10%            | 5%    | 100%  | 10%   | 11%   | 12% |
| Don't know                           | 15%   | 10%    | 10%            | 8%    | 12%   | 15%   | 18%   | 22% |

Descriptive statistics provide information only about the group from which data are collected. They are used to summarize the sample, rather than provide the basis for making inferences about the population from which the sample of data was drawn. For example, if evaluation data was collected from 100 households, the conclusions we reach based on descriptive statistical tools is valid for the 100 households and should not be extrapolated to the wider population.

Inferential statistics, on the other hand, provide the techniques that allow us to make generalizations about larger populations based on the data collected from a sample of the population. While descriptive statistics are used to describe what is going on in a given data set, we use inferential statistics to make inferences and generalizations from the data set to more general conditions and/or wider populations. Skilled statisticians are usually involved in conducting inferential analysis. It is important make sure that datasets collected using mobile solutions can easily be exported to statistical packages and shared with team members or stakeholders with the background and expertise necessary to conduct complex statistical analysis if this is needed for a particular project.
Using mobile devices for data collection can generate large volumes of data that deal with complex issues. How can we help decision makers better understand such complex data? According to Adam Bly, a strong proponent of making data available to the public, graphical presentation and the ability to interact with the data are important. Bly says, “Data visualization combines data and design to give insight and understanding to complex issues.” Decision makers may need help to be able to “see” the interconnectedness or relationships between complex variables; data visualization tools give you the means to provide that assistance.

Data visualization is the graphical display of abstract information. Since visual stimuli are processed much faster by the brain than cognitive content, we can use visualize data to make sense of it and to communicate it more efficiently. Stephen Few, a highly-regarded visualization thought leader, writes, “Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others.” At first, this may seem trivial, but well-designed data visualization can tell the story in a clear, meaningful, and compelling way, lending credence to the saying, “A picture is worth a thousand words.”

Most data analysis and visualization applications are hybrid in nature, serving multiple purposes including analysis, visualization and sharing. However, the visualization capability of most data analysis packages is rudimentary. Data collected on mobile devices are usually exported into spreadsheets or databases that support graphical presentations such as pie charts and bar charts. Some packages offer less common data presentation choices such as heat maps (two-dimensional maps where the color intensity represents a variable of the underlying data). Choosing software that offers the options most suited for your needs requires both planning and research. For programmatic monitoring and evaluation purposes, data is usually captured at specific intervals to reflect some change in activity or behavior over time and can be visualized in graphical form. Stephen Few provides examples of charts that illustrate quantitative information that are poorly designed for communication and offers recommendations for fixing them. An example of visualized data can be seen in the following chart illustrating the breakdown of asset holdings by sector for a popular exchange traded fund (ETF).

The primary problem with this presentation is the pie chart. To be useful, the graphic should enable us to quickly see the basic allocation breakdown of the ETF. However, in order to understand which slices represent which sectors, we must look to the table. As designed, the pie chart requires us to do unnecessary work, looking back and forth, matching the pie slice colors to the small colored squares in the table to understand which slices belong to which sectors. And because we must look up each slice on the table anyway, the pie chart is superfluous.
The author provides the following graph as an alternative.

![Sector Allocation of Holding](image)

The author explains:

“This simple design shows all the sectors and their corresponding allocations in a sorted table. Although the table data can be easily associated with the corresponding bars to determine each sector’s percentage, discerning individual values is not the purpose of the bars. Rather, they provide a means to rapidly compare the relative magnitudes of the values. People can quickly look at the overall shape of the bar chart and see if it’s relatively steep (as in this case), relatively even, or if it has one or two bars that tower over or are dwarfed by the rest, which provides a useful overview of the fund’s allocation percentages.”

Another approach is the use of “sparklines”—small, data-intense, simply designed, high-resolution graphics embedded within a text that provide a quick temporal understanding of a certain context. Edward Tufte, another thought leader in this field, provides excellent visualization descriptions in his *Sparkline Theory and Practice*.

Online interactive maps providing context and engaging features are a powerful way to share data. Somewhat more complex than rendering charts and graphs from numerical data, creating online interactive maps requires some technology expertise. Online interactive maps have been used effectively in both real time and more static situations for crisis mapping, crowd-sourcing and crowd-seeding, election monitoring, peace building efforts, tracking of agricultural pests, etc. Messages sent via SMS or voice can be incorporated into interactive maps. As with the other software tools discussed in this course, it is important to define your purpose before deciding which interactive map product is the right technology to employ to present your data.

There are multiple data visualization software applications available. *Exhibit*—developed by MIT with interactive mapping capabilities is a free and open source option. Among commercial applications, Tableau, which costs about $1,000 for a single user, provides excellent visualization capabilities. A list of additional free visualization tools can be found in this website and proprietary applications in this site. These lists are not exhaustive and we do not necessarily support the use of any specific tool. As always, it is important to define your needs before making decisions on tools.
DATA SHARING

As discussed in the previous weeks, data collected using mobile devices is transmitted to a server for storage and to facilitate making it available to users for analysis, visualization, report generation, and sharing. The data sets, and ensuing analyses and visualizations, can be shared with other stakeholders by granting them permission to use and/or download all or part of the this information. This includes partners like community health workers and even beneficiaries who may benefit from access to the data collected.

Governments, funding agencies, and implementing partners may have policies regarding data distribution, including rules that forbid sharing of certain datasets to protect the privacy of individuals and institutions. Simultaneously, governments and institutions increasingly encourage transparency and openness in data sharing to improve the translation of research results into knowledge and action. (For example, the “America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act” (America COMPETES Act) requires civilian federal agencies to provide policies and procedures that facilitate open exchange of data and research among agencies, policymakers, and the public.) It is important to understand these policies and regulations prior to designing your systems and to include provisions in your policies and procedures that ensure they are adhered to.

The phrase “data sharing” may define a multitude of arrangements involving the dissemination of data. It may involve one or more institutions exchanging data with each other or with a third party; several institutions may pool data to make it available to each other based on principles of reciprocity; it can be a one-off disclosure during emergency or unusual circumstances; or it can simply be making data available to a different department of the same organization. It is important to clarify who is sharing data with whom and on what terms before designing your systems or developing processes and procedures for data dissemination.

Among the ways in which data may be shared:

- Data is stored in a central server with access granted to users with approved user name and password;
- Datasets are available online via a website or mobile device. The former might be a good option if datasets are meant to be public such as census data while the latter may allow remote access by stakeholders in areas where broadband Internet is not available;
- Data is available on request only with approvals made on a case-by-case basis;
- Peer-to-peer exchange of data conducted informally between researchers and/or staff.

Some of the factors that are useful to consider when sharing data include the following:

- Adhering to your institution's privacy and confidentiality policy.
- Providing the data in a format that is accessible and actionable for each stakeholder group.
- Data sharing mechanisms such as using broadband Internet connection, wireless network connection, CD or flash drives, or hard disks. The findings of your readiness assessment (introduced during the second guide in this series) will help you in determining the best data sharing mechanisms for your area.
- Keeping datasets in a format that can be used by or exported to other applications. Using open standard formats such as Open Document Format (ODF), ASCII, tab-delimited format, comma-separated values (CSV), and XML as opposed to proprietary formats will ensure compatibility with various applications.
- Ensuring data security to forestall unauthorized access to data and/or change or destruction of datasets. This includes physical and network security of mobile and computer systems.

There are various data sharing applications, some of which are free. Examples include Dropbox or Google Drive. Using commercial cloud-based data storage and sharing services may pose possible privacy and security concerns. Carefully review a service's privacy policies before using them to ensure that data security and confidentiality are not compromised. The US government launched data.gov in May 2009 to “to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government.” USAID is working provide data at this site and also through USAID and State Department sites.