

WASH KAP – ANALYZING YOUR DATA V9.5

Contents

I. Aims of this tutorial	1
II. Analysis: getting your analysis file ready	2
II.1. Language selection.....	2
II.2. Import your data into the tool	2
II.3. Update the SURVEY and CHOICES tab	5
II.4. Others options you can set.....	6
II.4.A. Grouping models – “Config” tab	7
II.4.B. Using or not the analysis column	8
II.4.C. Changing the Analysis Column	9
II.4.D. Excluding codes	11
II.4.E. Other tips and tricks for your graphs.....	12
III. Pre-set graphs and indicators	12
III.1. Filtering data for core indicators and graphs.....	13
III.2. Question used for core indicators	14
IV. Additional information	15

Credits to the Kobo team for providing a tool setup to work directly with the XLS forms format. The “Kobo Excel Data Analyzer” is also available directly from the Kobo account. It has been imbedded in this analysis tool to allow the use of a single file.

I. Aims of this tutorial

The WASH KAP Analysis tool’s aim is to help you:

Main indicators for the surveyed population									
	1 - Average liters per person per day collected at HH level (averaged on person surveyed)	2 - % of household with at least 10 L/p potable water storage capacity	3 - Average distance to waterpoint for the 2% with longest distances	4 - % HHS collecting drinking water from protected/treated sources	5 - % HHS with family or shared-family latrine/toilet	6 - % HHS reporting defecation in a toilet/latrine	7 - % HH with access to soap	8 - % HHS with access to solid waste disposal facility	Extra - Number of HH using home tap water or watersellers
Emergency Standards	15	≥ 70%	≤ 500m	≥ 70%	-	0%	≥ 70%	≥ 70%	
Post Emergency Standards	20	≥ 90%	≤ 200m	100%	≥ 90%	0%	≥ 90%	≥ 90%	
Population surveyed	10.1	21.66%	1300	67%	99.64%	64%	69%	42%	0%

- ➔ Calculate the standard HCR WASH indicators in a couple of mouse clicks
- ➔ Have pre-defined graphs to help analyze the data from a WASH KAP in more details depending on local needs

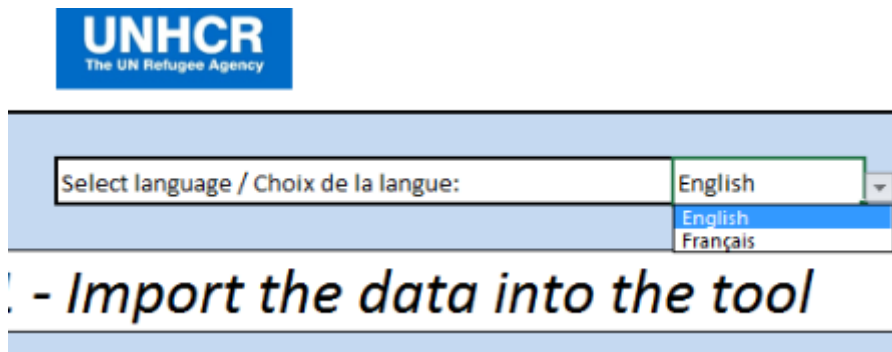
This tutorial’s aim is to explain how to use the WASH KAP Analysis Tool features and how to adapt it to your context.

II. Analysis: getting your analysis file ready


There are a few things we need to do in order to prepare the Analysis tool to show the data we have collected. The next few sections will guide you through the process.

II.1. Language selection

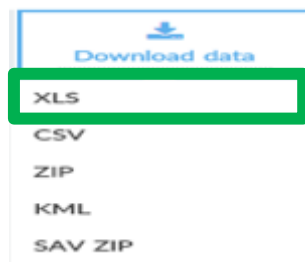
On the “Instruction” sheet, you can select your language (English or French available at the moment):





II.2. Import your data into the tool

 The tool comes with sample data, which you can use to get a feel for how it works and what are the available features. Once you are familiar enough, you can import your own data.

This is covered in details in Step 4 – Setting up the Survey with Kobo Online (“Getting your Data” in section 5). Just make sure to choose the “XLS” format among those proposed on the Kobo website:



 Don’t forget to save it on your computer in a location that you will be able to recover easily!

 You may come across a new option in Kobo to download the data with labels, instead of XML values (as part of the new interface). If you do prefer to use this option, ensure to choose “XML values and headers” or “Labels” and not one of the form’s language, as otherwise the analysis tool will not work.

CAP EAH NIGER v2
Project description

FORM
DATA ▾
SETTINGS

Select export type

XLS

Value and header format

Labels

XML values and headers

Labels

Français

Download

ANALYSER Exports

New Export

#	Filename	Date Created	Delete
1	A-GLO-PHS-WASH-EN-11_ANALYSER_2016_04_04_16_46_06.xlsx	4 Apr 2016, 4:45 p.m.	

Figure 1: Select the new export you have just produced.

- ➔ It opens on the main page, on which you'll find an "Import your Data" button. This will open a window & allow you to navigate to the folder where you have saved the XLS file containing your data:

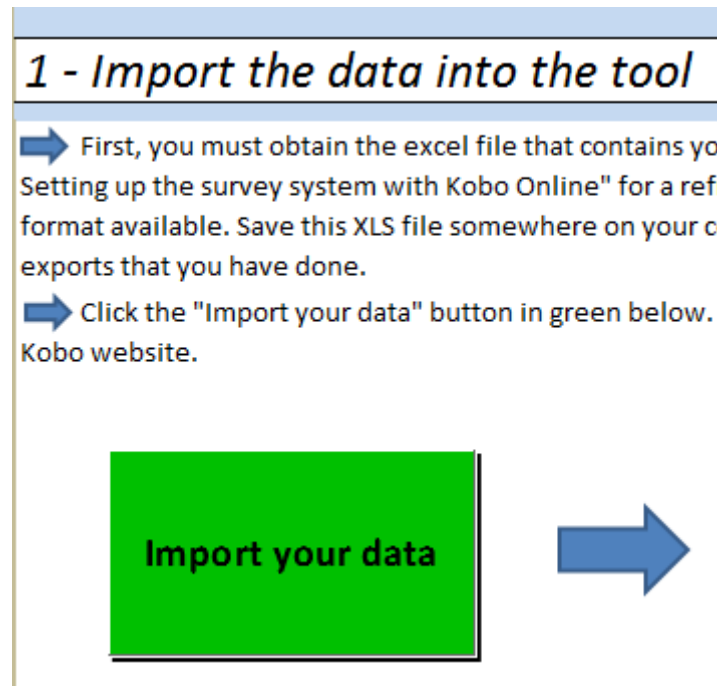


Figure 2: The helper buttons

When you click the button, a file selector opens:

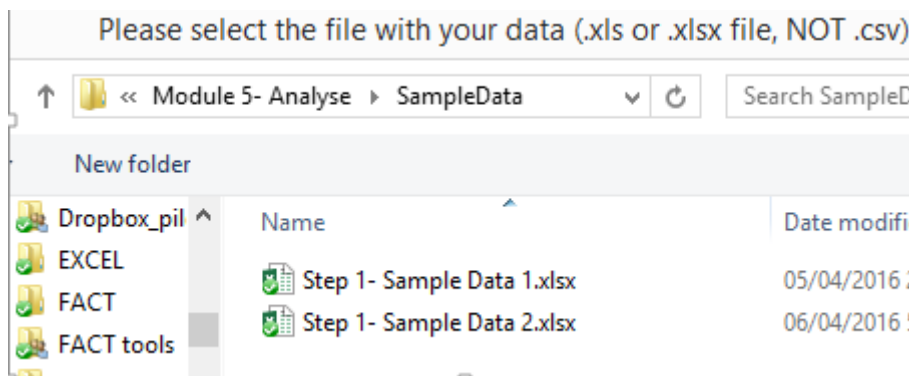


Figure 3: Select the file that contains the data you want to analyze.

The table & the graphs provided in the other pages will update according to your data automatically. These graphs & tables on the "Core Indicators" tab are pre-determined. However, you also have the possibility of generating other graphs to better suit your needs. These graphs can be created in 3 different styles: BarGraph, PieChart and Ranking:

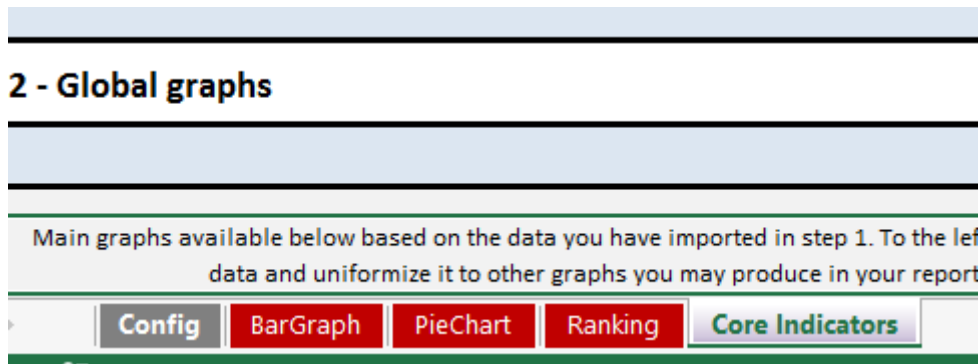


Figure 4: The other sheets allow to generate personalized graphs.

Finally on the “Config” tab, you can set a few useful options – more on this in the later sections of this document. For now we can simply set the language in which the tool will work:

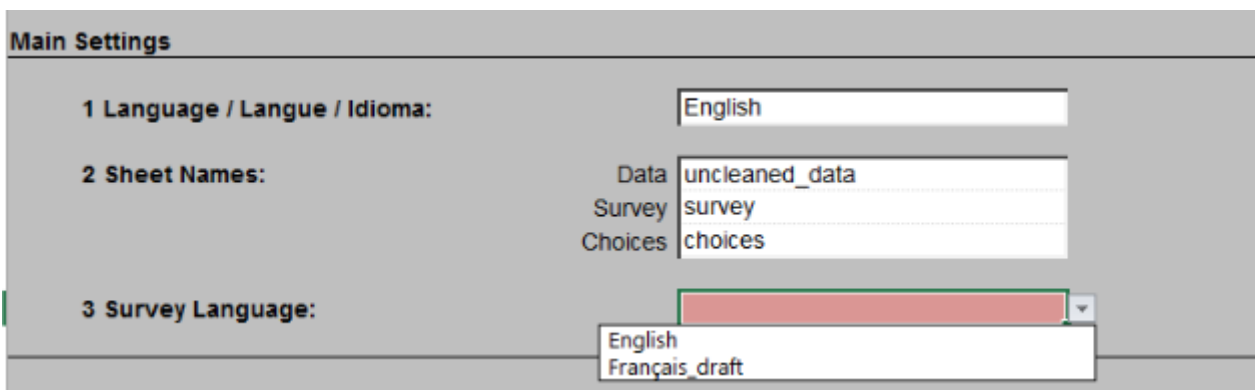


Figure 5: Make sure to select the survey language on the “Config” tab

- ➔ The choice you make here will determine the labels of the graphs that will be produced. It is better to choose the language in which you intend to produce your report. It doesn't have to be the same as the language that the enumerators used in field.

II.3. Update the SURVEY and CHOICES tab

For the tool to work properly, the survey and the choices tabs must be exactly the same as what you have deployed – identical to what you have uploaded on your Kobo account and have used for your data collection (Step 4 in Module 1 – “Setting up the survey system with Kobo Online” under section 3, “Uploading your forms to the server”).

The “Import your survey” button will help for this:

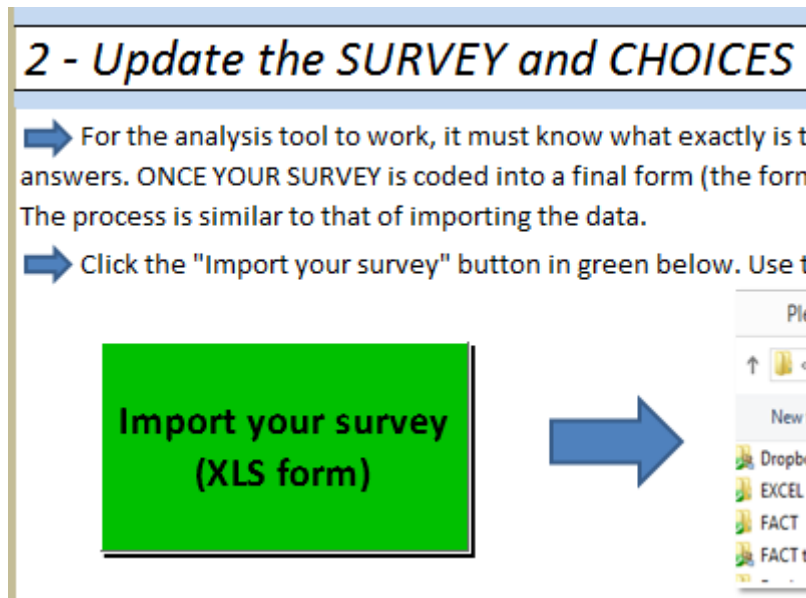


Figure 6: Importing the survey & choices

- ➔ When you click the button, a file selector opens - much like Figure 5 above (when you import the data to analyze using the "Import File" button).
- ➔ Choose the same XLS forms you have used to deploy your survey.
- ➔ This will update both the SURVEY tab and the CHOICES tab. Both are essential to perform the analysis.

⚠ Make sure that the file you specify is the latest version that you have uploaded on Kobo and used for your survey.

II.4. Others options you can set

In summary, to get the analysis tool to work you must:

1. Import the survey data (using the green import button and selecting the "XLS" output from the Kobo account)
2. Import the survey's "SURVEY" and "CHOICES" tab (using again the green button, and then selecting the survey you also uploaded on the Kobo account).
3. Select the survey language on the "config" tab.

We can now look at the features offered.

II.4.A. Grouping models – “Config” tab

Thresholds for number grouping models		
	A	B
0	-1	-1
1	5	0
2	10	2
3	15	4
4	20	6
5	25	
6		

Figure 7: Grouping models

This determines what are the categories used to display integer questions (where enumerators enter a number). There are a few things to note to use them effectively:

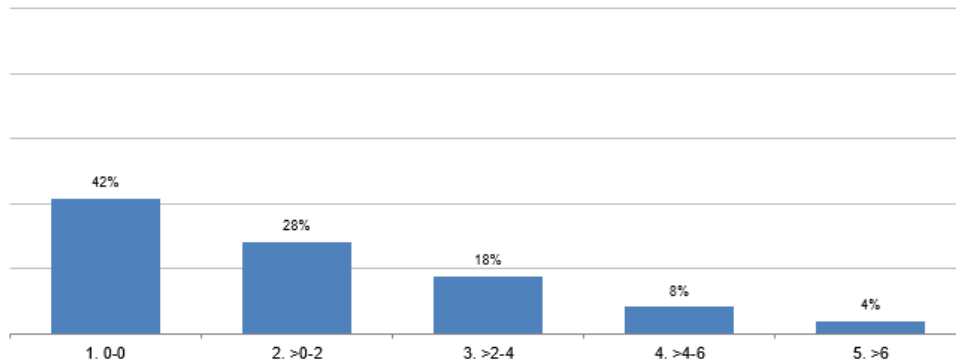
- ➔ In each category, the lower number is excluded, and the higher number is included. For example in column “A” above, the categories would be: 0-25, 0-6, etc.
- ➔ If you want the last category to be open (to include all numbers higher than the last one you have entered), you must leave it empty. In the example above, all numbers “25 and above” will be grouped together in the last category for the model A, while all “6 and above” will be grouped together for the model B.

As an example, we use model “B” to analyze the number of children below 5 in each household:

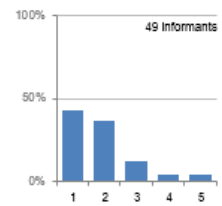
A.3.b. How many children less than 5 years old live and slept in this house last night ?

Percentage selecting each option

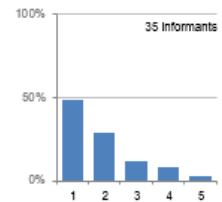
278 informants



Name 5



Name 2



Advanced Settings

A. Options to exclude

All Codes: -1 0 2 4 6

Excluded Codes:

B. Other Settings

Wrap labels

Place options in descending order

C. Number grouping

Model	Code
1	0-0
2	>0-2
3	>2-4
4	>4-6
5	>6
6	

B

Figure 8: Categories have been set according to the grouping model "B"

→ In the above example, we see that 42% of household had no children under 5, while 4% had more than 6 children under 5.

II.4.B. Using or not the analysis column

Still on the CONFIG sheet, you can choose to use or not the "Analysis" column. If you uncheck "Use 'Analysis' Column", then EVERY question that is in the survey will show up in the dropdown for the Excel Analyzer tool. By using the analysis column, you limit the choices available in those menus to what you really need according to your analysis plan.

We have already included in the survey default choices. If you check "Use 'Analysis' Column", you will be using these default choices we have set. If you uncheck it, everything will show.

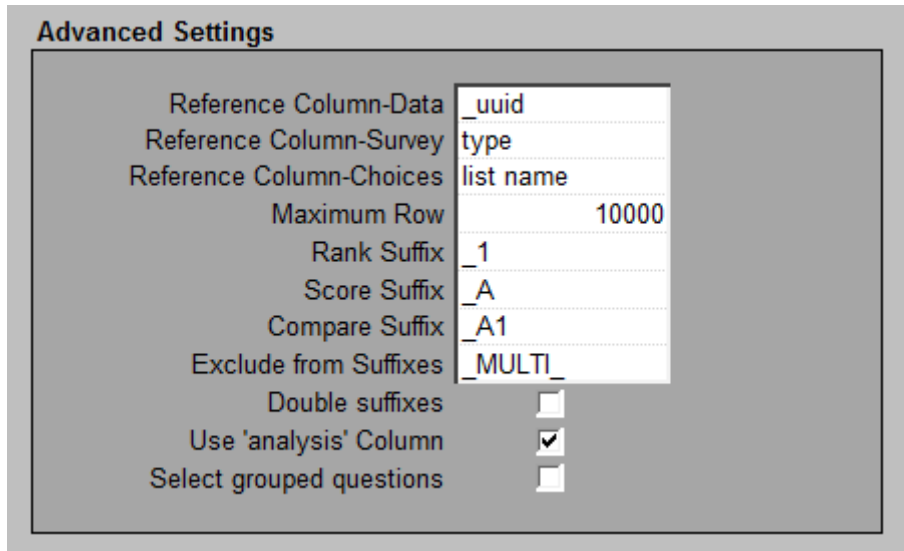


Figure 9: Using analysis column

II.4.C. Changing the Analysis Column

In case you decide to use this column, you may want to change what is available in the graphs. In your XLS form that you have deployed, go to the “analysis” column (one of the last in the “survey” tab). The column “label:English” contains the question – what is actually shown on the phones. (Some columns below between C & R have been hidden for simplicity – simply ignore them).

C	R
label::English	analysis
I8 Team number	D
I9 Who are you?	D
A2a - How many people live in this household and slept here last night?	C U
A2b - How many children less than 5 years old live in this household and slept here last night?	C U

Figure 10: Changing the options in the "Analysis" column

By entering some letters in the “analysis” column, separated by a space, you can decide in which page of the Excel Analyzer the question will be available to graph:

The options:

- ➔ BarGraph: C
- ➔ PieChart: U
- ➔ Ranking: R
- ➔ Disaggregation (by block, cluster, etc): D

⚠ Not all questions can be analyzed with all of the tabs. If a question is not available for a given type of graph, it will not show up in the dropdown, even if you write the letter in the proper cell in the analysis column.

⚠ You can change the analysis column directly within the analysis tool – this is why the survey tab is visible.

In figure 12, we can see that the question *Team Number* is used for disaggregation, while *How many live in this house (...)* and *How many children less than (...)* will be graphed using the CHOICE or UNIQUE graph.

Using options for disaggregation will allow you to analyze the data for sub-groups. For example, if your survey is conducted in different camps:

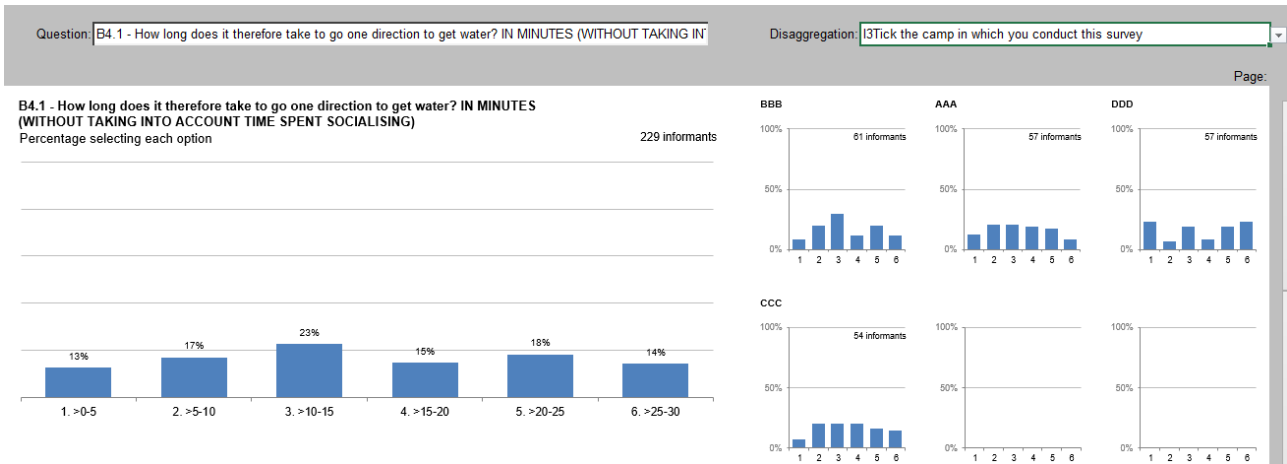


Figure 11: Using disaggregation in our bar graph

- ➔ In this example, we have one of the question selected (about how much time it takes to go to the water point). On the right, the same question is analyzed for each camp in the survey (in that case 4 camps). It helps to see if there are differences in the data.
- ➔ Other questions can be used for data disaggregation, by marking them with “D” in the analysis column.

⚠ Note: you cannot use “integer” type of questions for disaggregation.

The “PieChart” tab works similarly to the “BarGraph”. You select the question you wish to graph, along with any disaggregation you may need, and the resulting graph will appear:

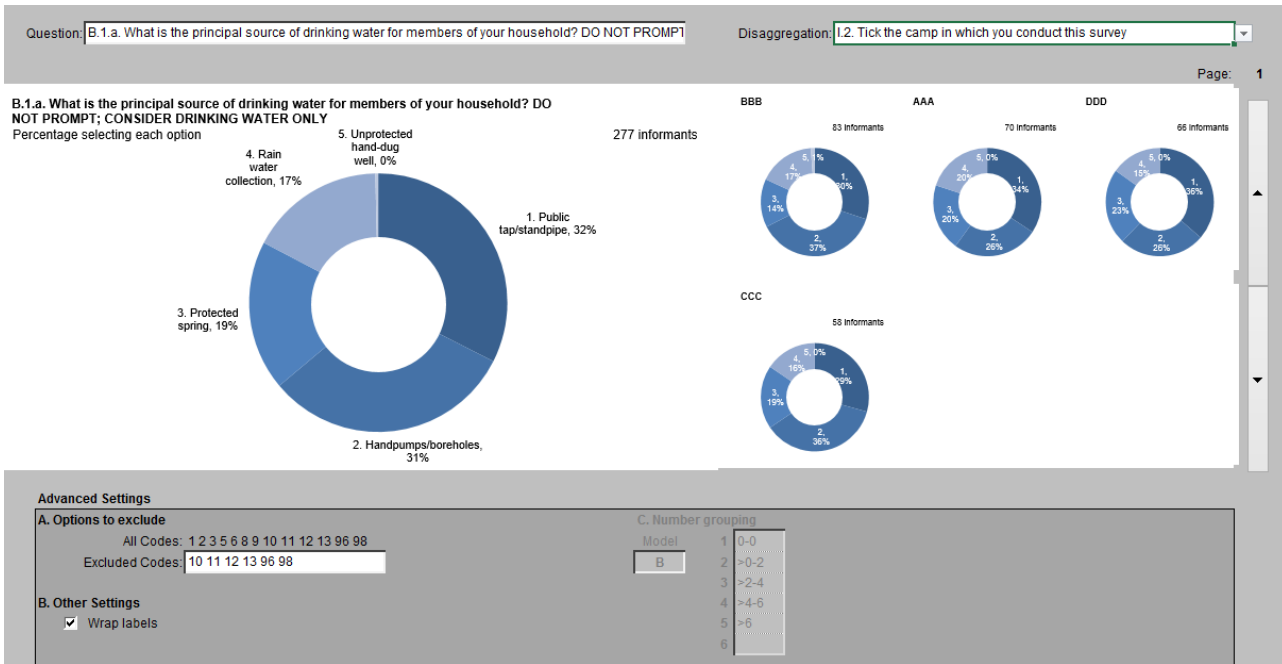


Figure 12: More options available, shown on a pie chart.

II.4.D. Excluding codes

In some cases, you may wish to exclude some of the answers in your analysis – perhaps there is some programmatic reason to do so. Or, as is the case above, there are many choices which have not been selected by any households. If we show them on the graphic, it makes it difficult to read. We can enter the codes under “Excluded codes”, separated by a space if there are more than one to exclude:

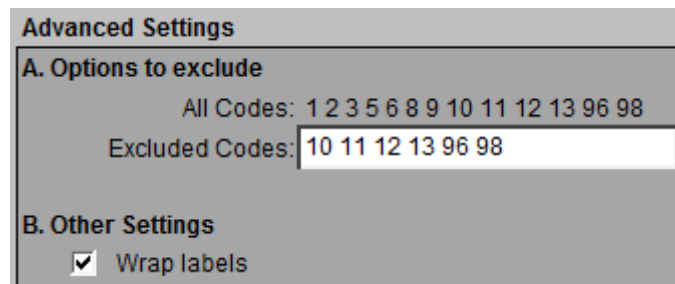


Figure 13: Options to exclude and wrapping of labels on the graph

How to find which value correspond to which answer in your survey? You must look into the “choices” tab of your XLS form (the WASH KAP survey we have provided):

D433		
A	B	C
1	list name	name
		label::English
436	source	1 Piped connection to house (or neighbour's)
437	source	2 Public tap/standpipe
438	source	3 Handpumps/boreholes
439	source	5 Protected spring
440	source	6 Rain water collection
441	source	8 Unprotected spring
442	source	9 Unprotected hand-dug well
443	source	10 Water seller/kiosks
444	source	11 Tanker truck
445	source	12 Bottled water, water sachets
446	source	13 Surface water (lake, pond, dam, river)
447	source	96 Other
448	source	98 Don't know

Figure 14: How to find which code to exclude based on your XLS form survey.

In the above example, we have excluded “Water seller”, “Tanker truck”, “Bottled water”, “Surface water”, “Other” because they have not been selected in our survey. We also decided to exclude “Don’t know” for programmatic reasons – we don’t want to include non-responses in our analysis.

Another useful option in this section is “Wrap labels”. If you have long options in your surveys, wrapping the text of the answer will help the presentation of the graph.

II.4.E. Other tips and tricks for your graphs

- ➔ The graphs are regular Excel graphs – there is nothing special about them. You can easily adjust the style and the presentation to suit your needs, just as you would with any graphs you would have created yourself based on the data from your survey.
- ➔ Some of the data field in the WASH KAP are calculated directly in the phone based on other information (such as the liter/person/day, which has to be calculated based on all the trips to fetch water and the number of members in the HH). These fields cannot be analyzed with the Bar graph or Pie chart tabs of the tool – see the other section below for this.

III. Pre-set graphs and indicators

In addition to the graphs that you can produce according to needs, there is a number of pre-calculated indicators & graphs that are of general interests for WASH KAP surveys – they include 8 indicators of particular interests for HQ, as well as 4 additional ones that are likely to be of interest for many operations. These are available under “Core Indicators” tab:

Main indicators for the surveyed population									
	1 - Average liters per person per day collected at HH level (averaged on person surveyed)	2 - % of household with at least 10 L/p potable water storage capacity	3 - Average distance to waterpoint for the 2% with longest distances	4 - % HHs collecting drinking water from protected/treated sources	5 - % HHs with family or shared-family latrine/toilet	6 - % HHs reporting defecation in a toilet/latrine	7 - % HH with access to soap	8 - % HHs with access to solid waste disposal facility	Extra - Number of HH using home tap water or watersellers
Emergency Standards	15	≥ 70%	≤ 500m	≥ 70%	-	0%	≥ 70%	≥ 70%	
Post Emergency Standards	20	≥ 90%	≤ 200m	100%	≥ 90%	0%	≥ 90%	≥ 90%	
Population surveyed	10.1	21.66%	1300	67%	99.64%	64%	69%	42%	0%

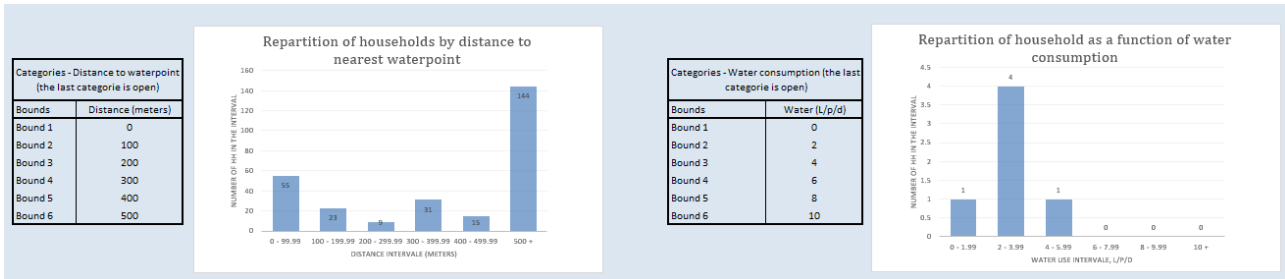


Figure 15: Indicators and graphs already available in the tool

For the graphs, you can determine what boundaries should be used for each of the categories (in the table next to each of the graph):

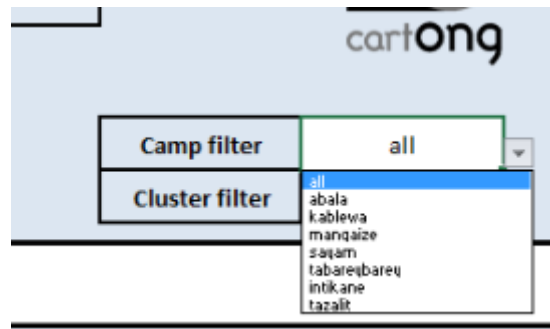
Categories - Distance to waterpoint (the last category is open)	
Bounds	Distance (meters)
Bound 1	0
Bound 2	100
Bound 3	200
Bound 4	300
Bound 5	400
Bound 6	500

Figure 16: Setting the boundaries for categories

This can be useful if you have already some boundaries defined in other reports/internal processes and you would like the KAP survey to match them, or if you data don't fit well with the pre-defined categories.

III.1. Filtering data for core indicators and graphs

As is the case for the graphs mentioned above, it is possible to filter the core indicators and graphs as well – either by camp (if more than once camp surveyed) or by cluster. This can be done with the selectors in the top section of the “Core indicators” tab:



A few important notes:

- ➔ “All” for camps means all of the data will be considered (so the data isn’t filtered by camp)
- ➔ For clusters, if all the clusters should be considered (or if the survey didn’t use clusters for sampling), then “0” should be entered
 - Do not use 0 for cluster number in your survey, otherwise it will not be possible to analyze the data for that cluster separately.

III.2. Question(s) used for each core indicators

This section details to which question in the survey each of the core indicators refers and how they are calculated, so that reports produced based on those indicators can explain precisely what the data actually represents.

All indicators can be filtered by camp and/or cluster based on the selectors available in the “Core Indicators” sheet. Those filters rest on the following questions:

- CAMP
- CLUSTER

Most indicators rely in some way on the total number of household surveyed (since they come in % of respondents). The total number of household always excludes any non-consenting respondents.

In some cases (indicators 1, 2, 3) the indicator requires family which have collected water yesterday (otherwise the data isn’t provided). In those cases, the question CONTAINER must have at least 1 or more for the record to be counted (which means they used at least 1 container to collect water the previous day).


The 3 graphs below the core indicators table are based on the same calculations.

Indicator	WASH KAP questions on which it is based	Notes on calculation
1 - Average liter per person/day	<ul style="list-style-type: none"> • xAvgVolCollectedProtected (calculated field), which is based on: <ul style="list-style-type: none"> ○ LITER ○ CONTAINERPROTECTED ○ NUMTRIPS 	Only water collected with protected containers (CONTAINERPROTECTED = yes) is included
2 - % of HH with at least	<ul style="list-style-type: none"> • xAvgVolStorage (calculated 	Returns the % of HH which have at least 10

10 I/p potable water storage	field), based on: ○ LITER	liters/person of storage capacity (protected or not)
3 - Average distance to water source for the farthest 2%	• XDistToWaterPointMeters (calculated field) based on: ○ TIMEWATER	Returns the average distance from water source for the farthest 2% of respondents (this is meant to ensure that outlying values do not weight too much in the result). This assumes an average walking speed of 80 meters/minute.
4 - % of HH collecting drinking water from protected source	• ASource_1	Returns the % of HH which have as a primary water source one of: tap stand, borehole, water seller, pipe connection, protected spring, bottled water or tanker trucks.
5 - % HHs with household or shared-family latrine/toilet	• FACILITYDEFECATE	Based on family using either private facilities or facilities shared with limited number of families
6 - % HHs reporting defecating in a toilet/latrine	• DEFECATEADULTS	Based on HH which report defecation in a latrine (either public or private). The % only includes respondents which affirm using them ("Others", and "Don't know" aren't counted as positive answer to this question).
7 - % HHs with access to soap	• PRESENCESOAP	
8 - % HHs with access to solid waste disposal facility	• DOMESTICWASTE	Returns the % of HH which report access to one of: "Household pit", "Communal pit", or "Street bin/container for garbage collection"
9 - % HHs with access to a specific hand-washing device	• WASHDEVICE	Returns % of HH reporting access to designated hand washing device
10 - % respondents knowing at least 3 critical moments when to wash hands	• WASHHANDSTIME	Returns the % of HH which have correctly answered at least 3 correct answers as to when they should wash their hands
11 - % HHs practicing open defecation **If adult, or children report OD, or if they have answered yes to defecate in bush at night.	• DEFECATEBUSH • DEFECATEADULTS • DEFECATECHILDREN	Returns the % of HH which are admitting at least some OD based on any of those 3 questions
12 - % HHs having access to a bathing facility	• BATHINGFACILITY	Returns the % of HH which report access to a designated, familial bathing facility

IV. Additional information

You should keep a blank version of this tool (without any modifications to it) and work off a copy.

 If you import data into a tool that already contains data, the old data will be lost and only the latest data you have imported will be shown. In other words, this is meant to work as a one-off importation (it does not cumulatively store data).