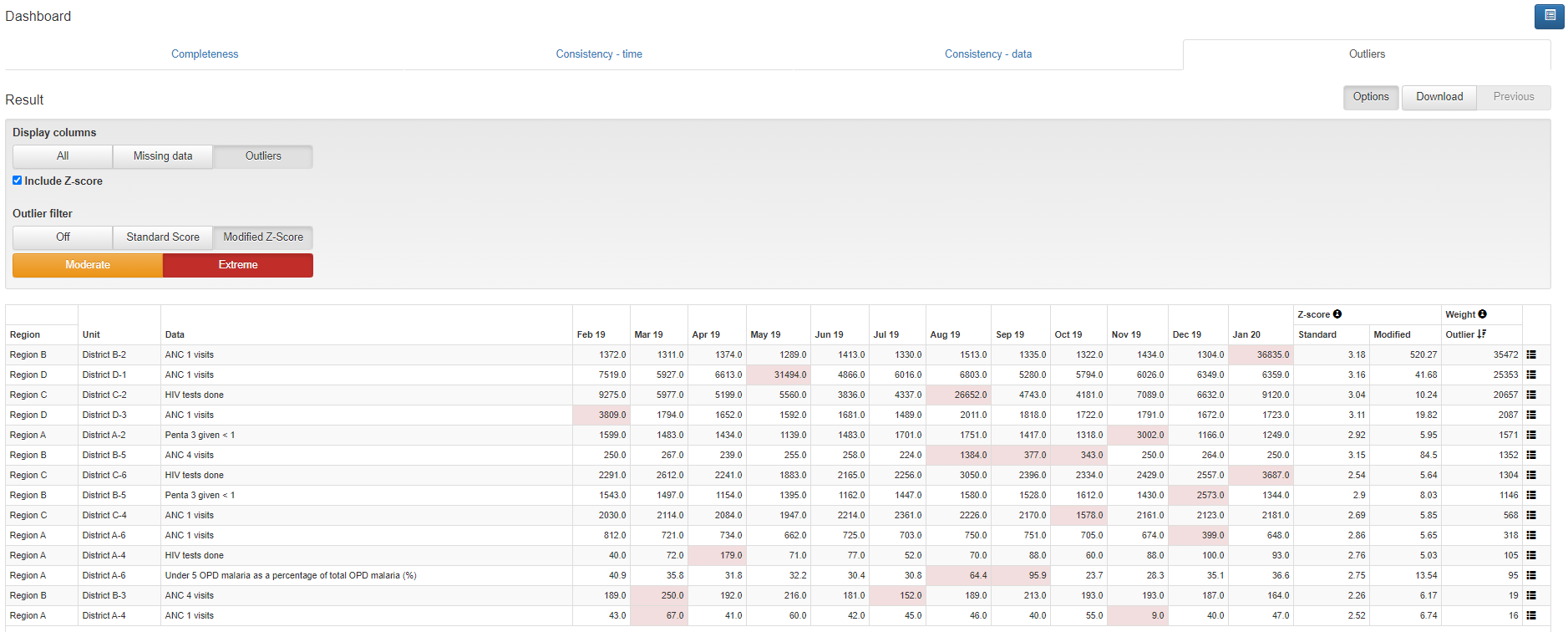
Exercise 2.5: Use the Outliers dashboard

Click on the tab for the “Outliers” dashboard. Review the menu window to confirm that Data is set to District, Period is set to January of the current year and Disaggregation is set to District. Click on the settings icon to hide the settings window and free up space on the screen. Click on the Options button then check the box for “Include Z-score” and click on “Outliers”, “Modified Z-Score” and “Extreme”.

Review the Result table. **Question 11**: Each row shows data for which level of the health system? (National? Regional? etc… ) **Question 12**: Compare the values highlighted in red to the other values in the same row. For the first row, take the value highlighted in red and subtract from this an approximate average of the other values in the same row. How does the result compare to the “weight” for the row? What does the weight represent?

**Question 13**: Visualize the data: Click on the menu icon at the end of the first row of the table and select “Visualize”. Describe the graph. After viewing the graph, click “Close”.

**Question 14**: Drill down to investigate the outlier: Click again on the menu icon at the end of the first row of the table and select “Drill down”. A new “Result” table should appear with each row showing 12 months of data for an individual health facility in the selected district. Review the first row of this new table. The extreme outlier was reported by which health facility? Do you think the outlier is due to an error? Try to drill down further. What happens and why?

## **Discuss the steps to take when an error is suspected**

**Question 15: List the measures which data managers should take** if they identified a value that was very suspicious and so large that it probably has a significant effect on a nationwide statistic.

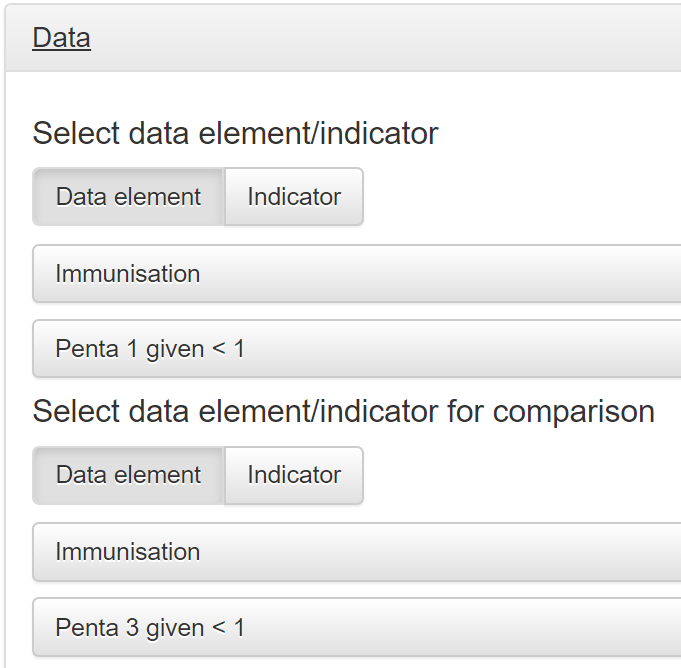
**Question 16:** Who has authority to edit the DHIS2 data? Do the Standard Operating Procedures specify how to document changes that are made to the data?

**Question 17: In your health system how should the DQ Tool be used to identify extreme outliers** – at district level as well as at national level -- each month or quarter as well as each year.

## Exercise 2.6: How to use the Analysis function of the WHO Data Quality Tool

#### Use Analysis-Consistency to assess DPT dropout rates:

Click on the Analysis tab then select “Consistency” from the drop-down menu. Under Analysis Type, click on each of the following: “Between Indicators”, “Expected result” and Dropout”.

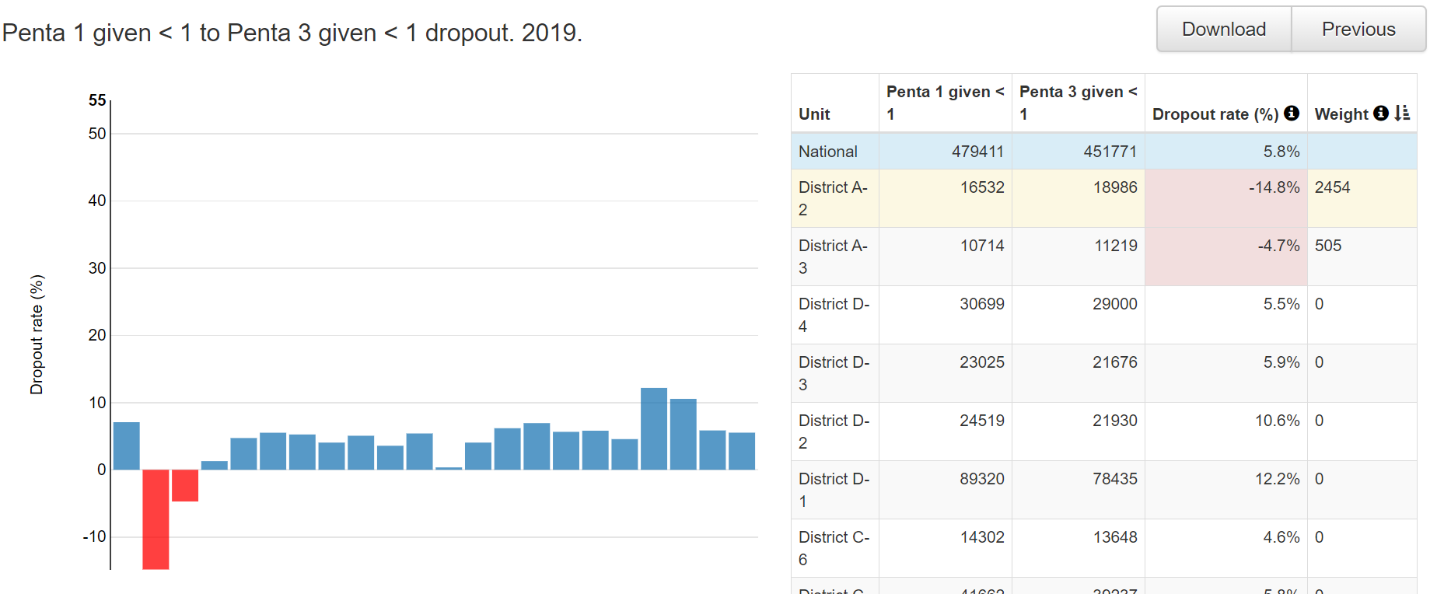
Click on Data to open the Data settings window. Set Data as shown in the figure to the right:

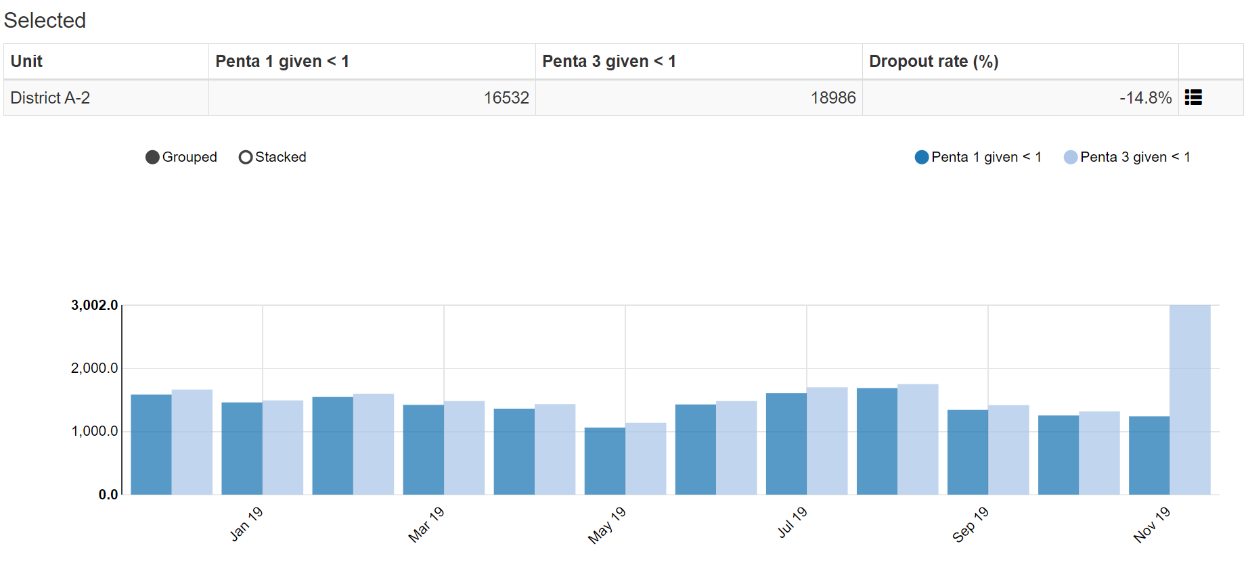
* Under “Select data element/indicators, click on “Data element”;
* Click in the box for “Data element group ….”, then find and click on “Immunisation”;
* Click in the box for “Select data element …”, then find and click on “Penta 1 given < 1”;
* Under “Select data element/indicator for comparison”, click on “Data element”;
* Click in the box for “Data element group ….”, then find and click on “Immunisation”;
* Click in the box for “Select data element …”, then find and click on “Penta 3 given < 1”.

Click on Period to open the Period settings window. Set Period to Year = last year.

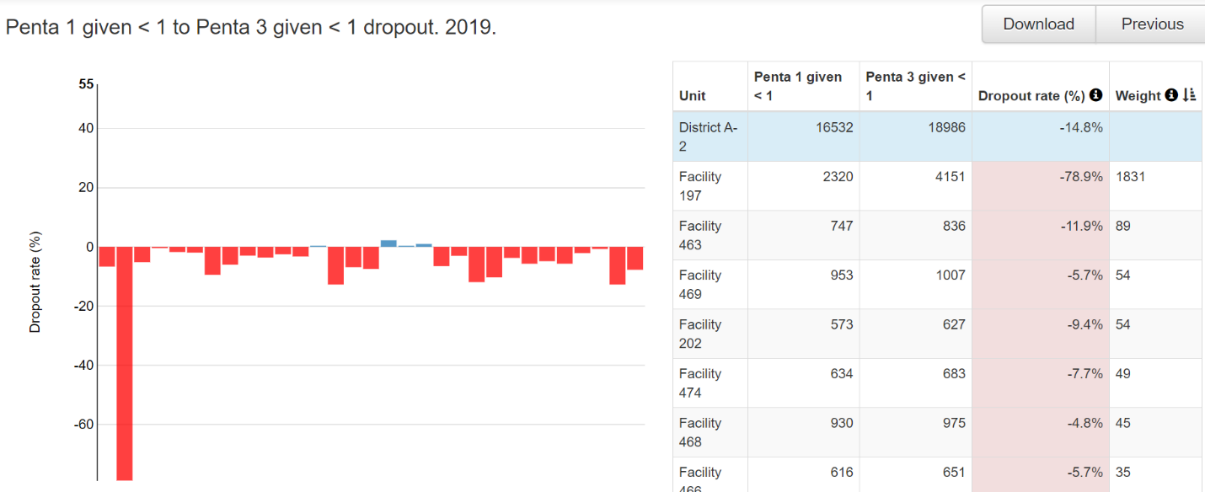
Click on Orgunit to open the Organisation unit settings window. Leave Boundary set to National. Re-set Disaggregation to District.

Click the Analyze button.

1. Review the chart and the table (note: the top row of the table will not be shaded yellow until the next step is completed). Move the cursor over the vertical red bars in the chart to view the details. Compare the names of the districts with vertical red bars to the names of the districts in the table with Dropout rate shaded in red. **Question 18**: Which district has the most negative dropout rate? Find it in the chart and find it in the table. For that district, for last year, what was the value of Penta 1 and what was the value of Penta 3 (see table)? Use these numbers to calculate the dropout rate.
2. Click on the row of the table which shows data from the district with the most negative dropout rate. A new table and chart should appear at the bottom of the page.



**Question 19**: Is the negative dropout rate for District A-2 due to an extremely large Penta 3 value for a single month or is it due to higher values for multiple months?

1. Click on the menu icon at the right end of the small table and select “Drill down”. A new chart and table should appear showing the dropout rates for last year for each health facility in District A-2. **Question 20**: Is the negative dropout rate for the district due to a value from a single facility or is it due to values from multiple facilities? Does it appear that the selected district has a systematic problem with over-reporting of Penta 3 – for multiple months and multiple sub-districts?
2. Click on the row of the table which shows data for the health facility with the most negative dropout rate. A new table and chart should appear at the bottom of the page. **Question 21**: Describe what this table shows. Is the negative dropout rate for Facility 197 due to a single very large outlier value of Penta 3 or is it due to Facility 197 repeatedly reporting more 3rd doses of Penta vaccine than 1st doses?

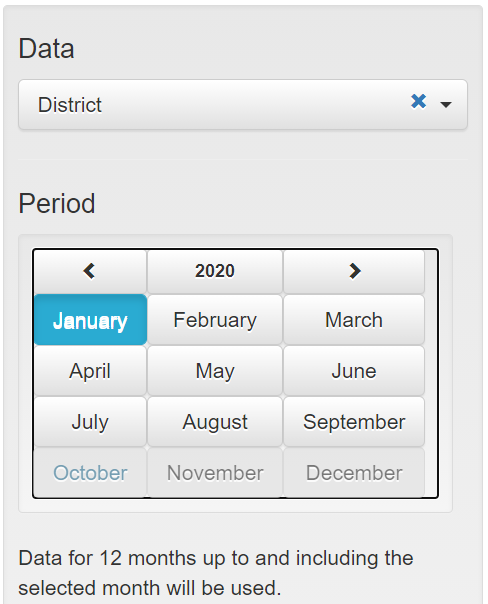
#### Use Analysis-Outliers to identify extreme outliers of additional data elements

Click on the tab for Analysis, then select “Outliers and Missing Data”. Select any data element or indicator. Then set Period to last year and set Orgunit to National with Disaggregation by District. Click on Analyze then proceed to use the Tool to identify extreme outliers using the same steps as with the Outliers dashboard. **Question 22**: What are the advantages and disadvantages of the Analysis-Outliers function compared to the Outliers dashboard?

## Exercise 2.7: How to train staff at district level to use the outlier dashboard

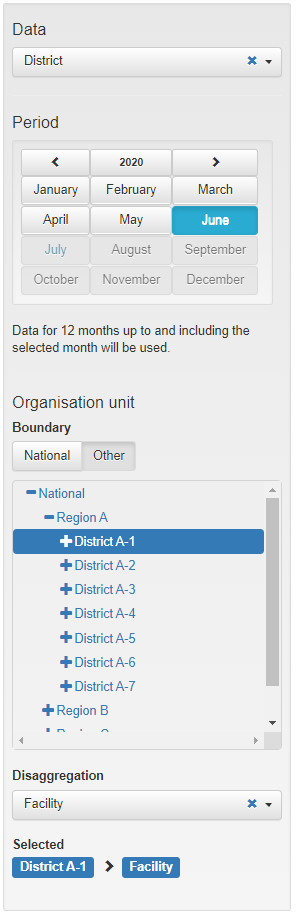
#### Launch the DQ Tool and explore

*Do not spend long reviewing the Completeness dashboard.* The completeness tab of the DQ Tool shows the reporting rate for select data sets. The same information is presented more comprehensively with the charts and tables on the Data Quality dashboard. Thus, do not spend long reviewing this tab.

Explore the settings window of the DQ Tool dashboard

Click on the settings icon ( ) in the upper right of the screen to show a settings window on the right of the screen. If the settings window appears briefly then disappears, click again until the settings window stays on the screen. Under “Data” there is a white space where [Core] appears. Click on the down arrow to the far right of this space and a drop down menu will appear showing the various groups of indicators which can be reviewed on the DQ Tool dashboard: [Core], General Service Statistics, Maternal health, HIV/Aids, Malaria, District and Immunisation.

**Change “[Core]” to “District”**. The reporting period that is shown on the dashboard can be changed by clicking on a different month. After experimenting with changing the **Period**, reset it **January of the current year**.

Change the “Organisation unit” that is being ana**lyzed**

Under Boundary, click on “Other, then click on the “+” sign to the left of “National”. The list of regions will then appear. Click on the “+” sign to the left of “Region A” and the list of districts in Region A will appear. Click on and select District A-1. When staff of District A-1 first launch the DQ Tool, they do not need to reset Organisation unit. It will automatically show the name of their district. By default, Disaggregaton will be set to “Facility”.

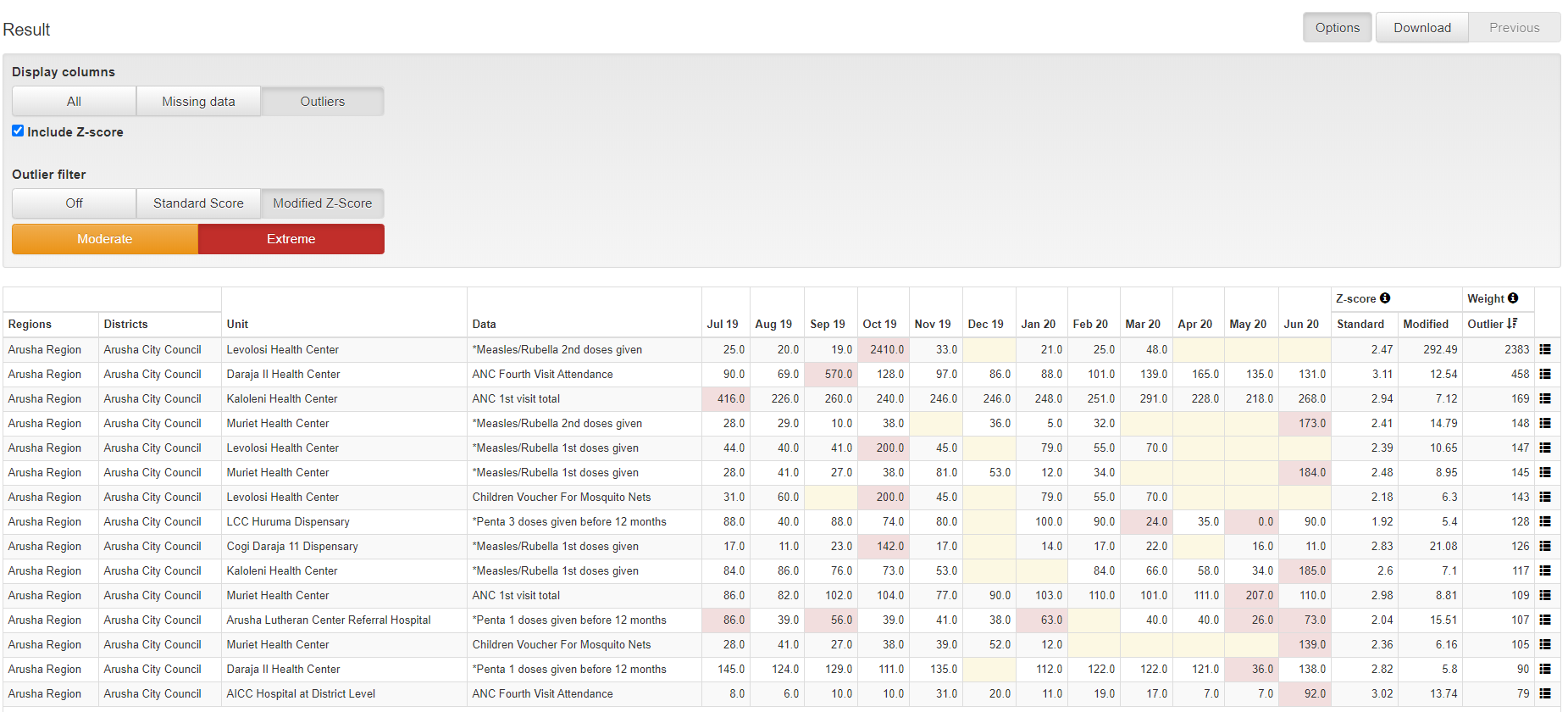
**Use the “Outliers” dashboard to identify suspicious values**:

Click on the tab for “Outliers”.

To free up space on the page, click on the settings icon to make the settings window disappear.

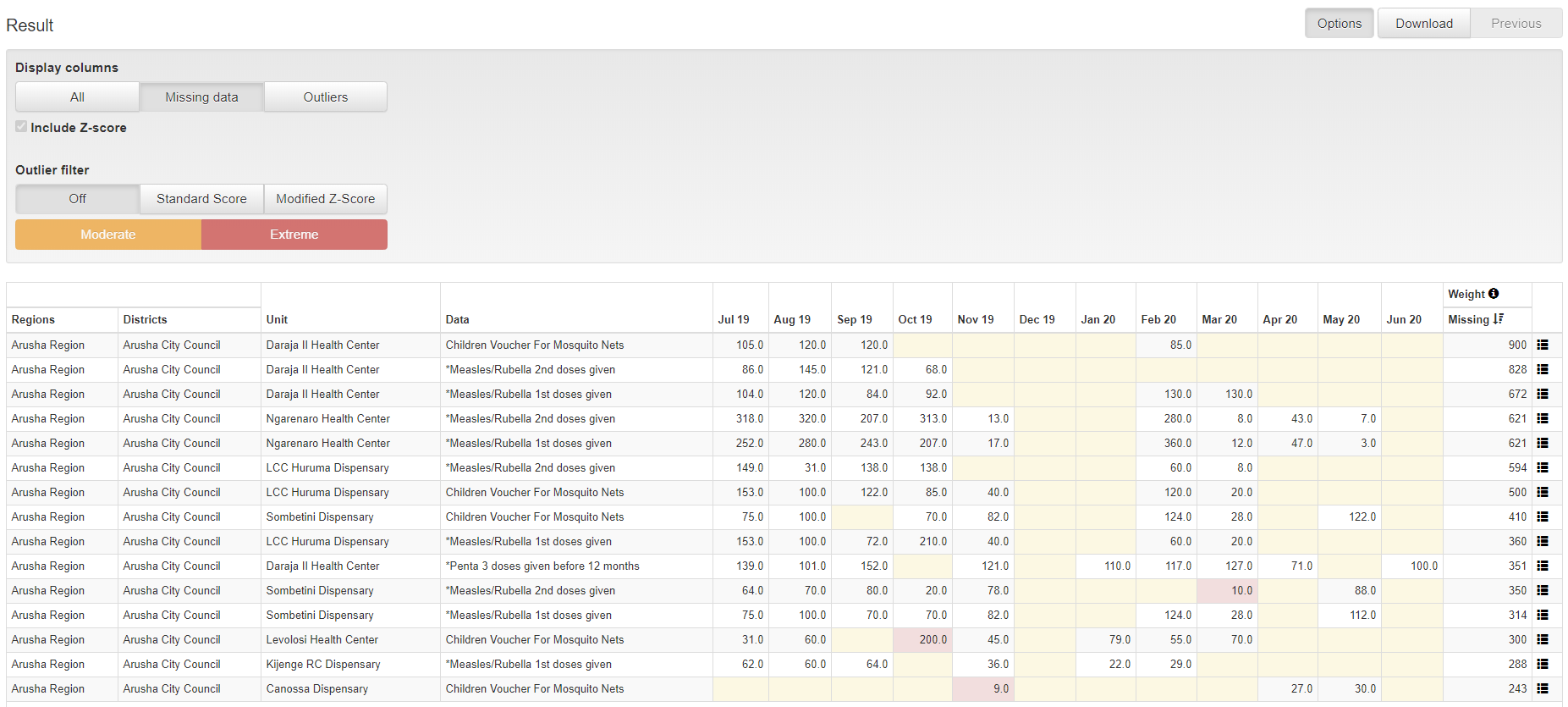
Each row is for one health facility. The name of the health facility is shown in the column headed “Unit”. Each row is for one core indicator reported by that health facility. The name of the indicator is shown in the column headed “Data”. Each row has the 12 values for that indicator reported by that health facility during a 12-month period.

Each row has one or more values highlighted in red and/or one or more cells shaded yellow. **Question 23**: How are the values highlighted in red different from the other values in the same row? **Question 24**: How are the cells highlighted in yellow different from the other cells? **Question 25**: Are the rows of the table sorted in any order? Review the table, then describe how you think the rows are sorted. **Question 26**: How do you think “Outlier weight” is calculated? **Question 27**: How do you think the “Missing weight” is calculated? How is “Total weight” calculated?

**Use the Options window to filter the Result table and see the most important outliers and missing values.** A good way to find the most suspicious values is to use their “Z-score”. To use this method, click on the Options button to open the Options window. To see the Z-score for each row click on the box “Include Z-score”. Review the modified Z-score values in the Result table. They range from zero (not suspicious) to more than 5 (extremely suspicious). Next, click on the following buttons: 1) Outliers; 2) Modified Z-Score; and 3) Extreme. Notice how the Result table reduces to only a few rows. **Question 28** What is the smallest value of Modified Z-score that you now see in the table?

**Inspect each row** –Click on the menu icon at the end of the first row and select “Visualize” from the drop-down menu. **Question 29**: Is the red value suspicious? After you have finished reviewing the bar chart, click on Close.

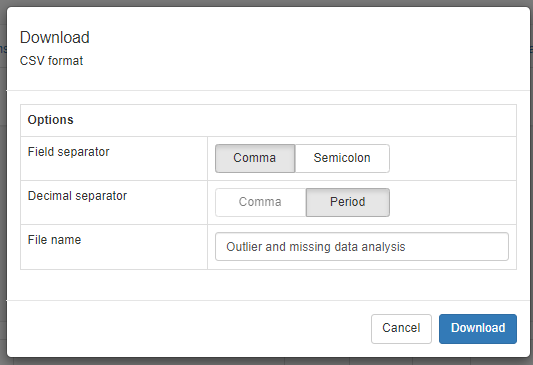
**Use the “Outliers” dashboard to identify important missing values:** To review the missing values, return to the Options window and click on “Missing Data” and “Off” (to turn off the Outlier filter)

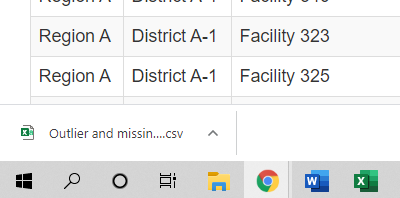


**Question 30**: Are the rows of the table sorted in any order? Review the table, then describe how you think the rows are sorted. **Question 31**: Which rows are most important to focus on? Which rows show the health facilities which have the largest amount of missing data?

Download the data to a spreadsheet and use the spreadsheet to view the rows

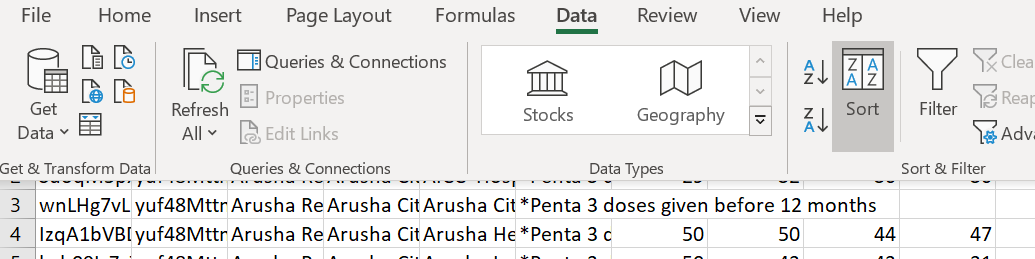
Click on the “Download” button.

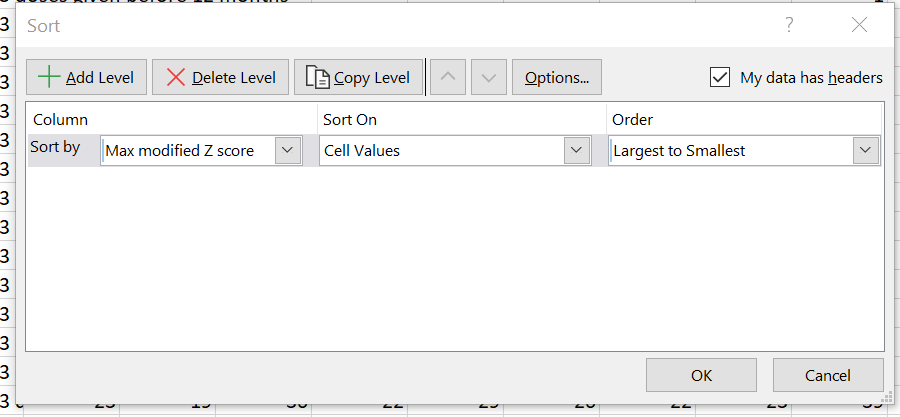
When the Download window appears, leave the default setting as they are and simply click on “Download”.

The file will appear in the bottom left corner of the screen. Click on the file name once to open the file.

Outliers and missing values will **not** be highlighted on this CSV file, so you will have to visually scan the numbers. To identify health facilities with the most suspicious values, sort the rows based on “Max Modified Z-score”:

With Excel, select “Data” then “Sort”



Set the Sort by to “Max modified Z-score” (you have to scroll down to find it). Set Orde to “Largest to Smallest”. Click “OK”.

Review each of the top 3 rows. Identify the suspicious value in each of these rows. **Question 32**: For each of these top 3 rows identify the cell with the most suspicious value and highlight it yellow (Home – Fill color yellow). Ask a facilitator to review your CSV file and confirm that you have followed all steps correctly.

Return to the outlier dashboard and, under Options, click again on Outliers, Modified Z-Score and Extreme. Sort the result table by clicking once on the column header for Modified Z-Score. **Question 33**: How does the Result table of the Outlier dashboard compare to your CSV file?

**Question 34**: How do you sort the table show that the top rows show health facilities with the largest amounts of missing data?