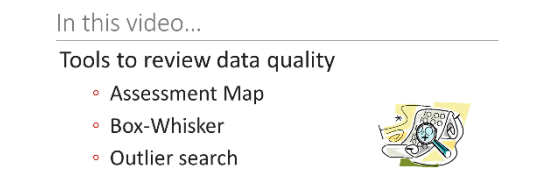
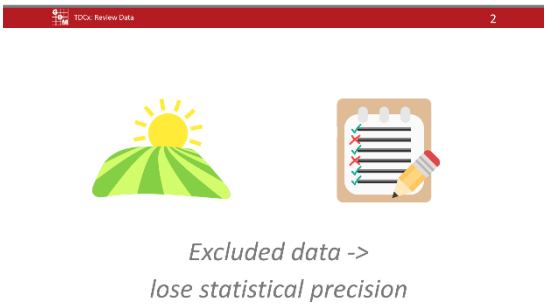
Review Data with TDCx

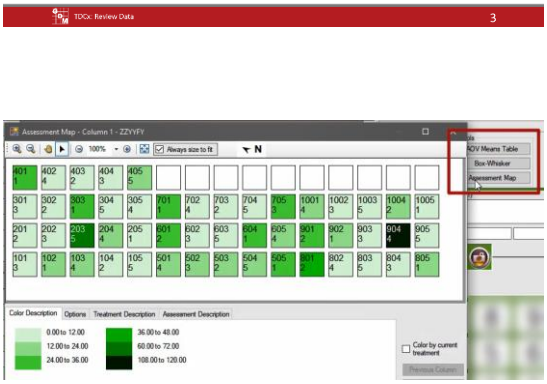


In this video, we will use TDCx to review assessment ratings on-site. We will demonstrate tools to discover any mistakes or unusual data points. These include a heat map of assessment values, a box-whisker graph, and an outliers calculator.



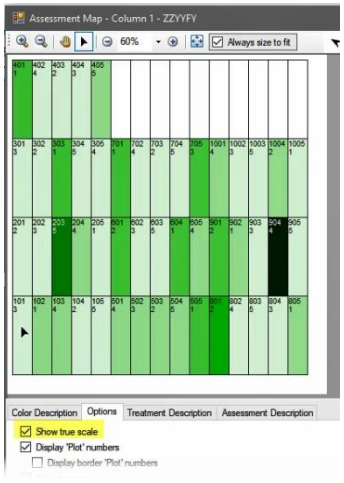
The best time to review data is immediately after making the assessment, while still at the trial site. This way, it is still possible to verify and correct mistakes, or to document irregularities. If these checks are done after leaving the site, the only option may be to *exclude* any questionable measurement from the analysis, resulting in lost precision!

TDCx brings along all of ARM's data review features, to use as soon as the assessments are taken.

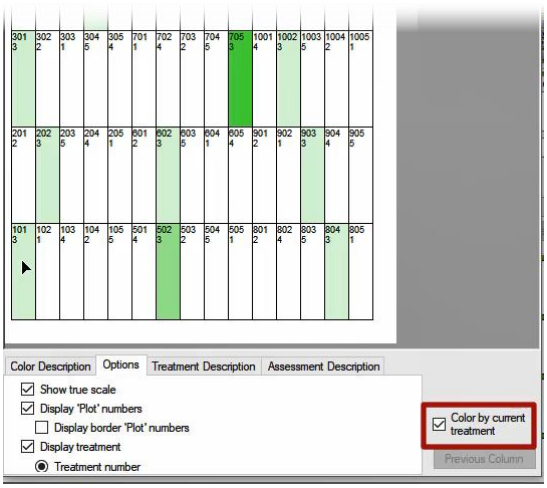


We start in a trial where we just finished entering data. (See our tutorial video for more information about setting up TDCx for data entry.) There are a variety of data review tools available in ARM – and we can take advantage of all of them with TDCx while still at the trial location!

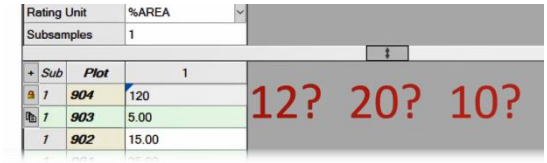
First, create an Assessment Map from the Tools section of the Properties panel, to display a "heat map" of the plot values in the current data column. Low values are colored lighter, and high values are colored darker.



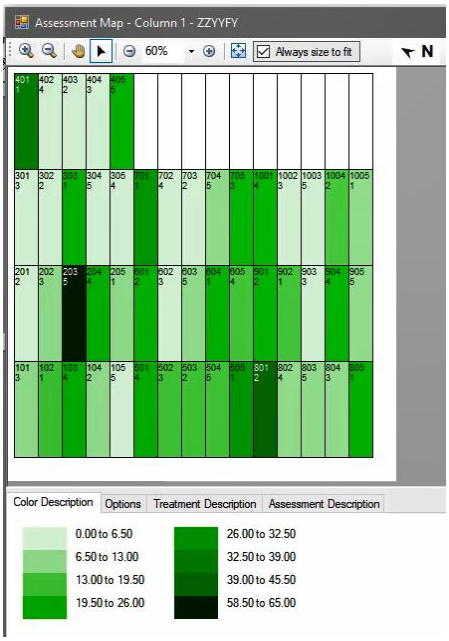
Use the Show True Scale option to view the plots to-scale, based on the entered plot size and layout settings of the trial. Use this to identify spatial effects at the trial location. Is there a pattern in the high and low values? Does this pattern make sense for this trial?



You can also inspect a single treatment, to look for differences amongst replicates. Click on a plot, or select 'Color by current treatment' to color only the chosen treatment on the map. You can visually identify a potential "problem replicate" or extreme value, treatment by treatment.

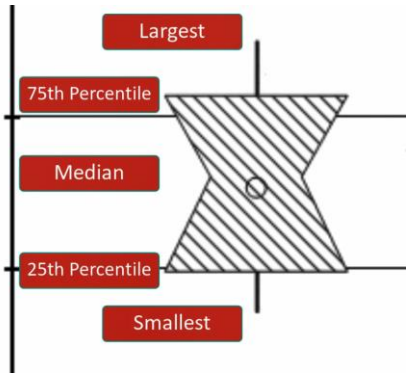


Plot 904 was likely noticeable from the moment we opened the Assessment map – it is so much darker than any nearby plot or any other treatment 4! Double-click on that plot to take a look at the this data point on the Assessment Data editor – ARM moves the cursor right to that value.



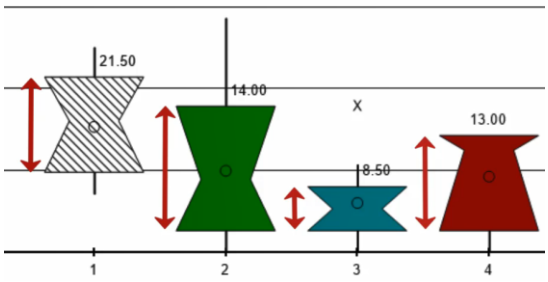
We see that the value is 120 – clearly this is wrong, since 120% isn't even possible! But should it be a 12? 20? Or 10?? If we were reviewing this data from our desk after pasting from Excel or keying in the values from paper, there is not much we can do. But with TDCx, we can reinspect the plot!

Let's say that upon our review we determine that this should be a twenty, so we can set the value right. Now the assessment map shows us much more detail, without the incorrect value skewing everything.

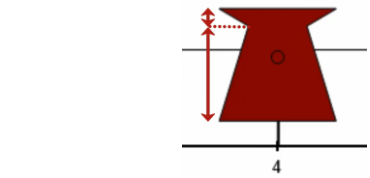


The next tool we can use is the Box-Whisker graph. This graph shows the 'spread' of treatment data around the median, using a 'box' and 'whiskers' to break down each data group by percentiles.

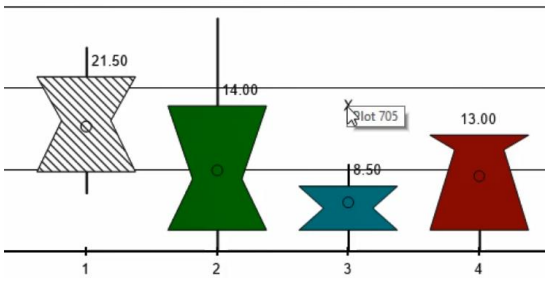
- The box extends from the 25th to the 75th percentile, and is divided by the median.
- The whiskers extend from the ends of the box to the largest and smallest non-outlier values.
- Outliers are values that lie outside the box by more than 1.5 times the height of the box.



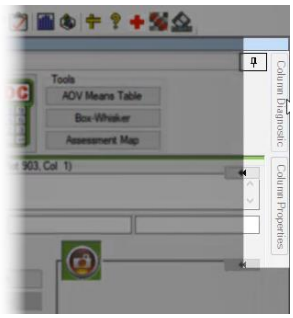
The box height indicates the amount of variability within a treatment, so is a measure of assessment consistency. The heights for all treatments are fairly similar in this case, so no sign of outliers there.



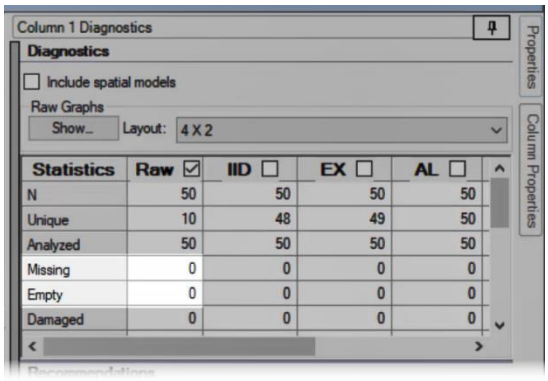
The waist position of each box indicates skewness. Treatment 4 has a "high" waist, which indicates that a value in one of the replicates is somewhat *lower* than in other replicates.



Finally, outliers that are calculated display as x's on the graph. Tap on the x (or hover with the mouse) to display the plot number for that value. Plots 705 and 203 have potential outliers, at least according to the box-whisker definition.



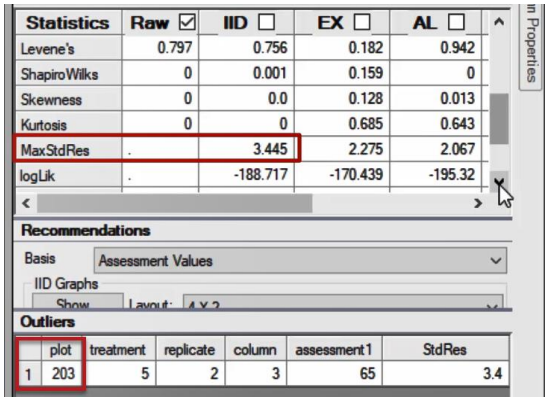
ARM has another tool for statistically finding outliers – Column Diagnostics. Open the panel from the far-right side of the screen. There is a lot of information here, but not all of it is needed at this stage of data review!



Statistics	Raw	IID	EX	AL
N	50	50	50	50
Unique	10	48	49	50
Analyzed	50	50	50	50
Missing	0	0	0	0
Empty	0	0	0	0
Damaged	0	0	0	0

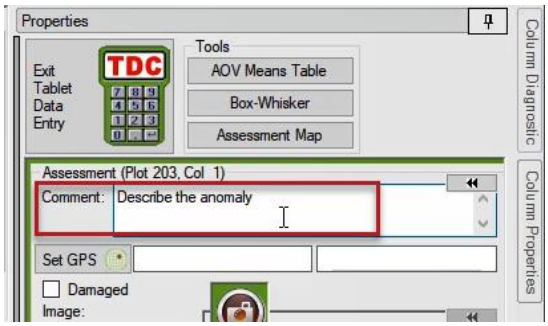
Start with the first few rows in the Statistics table. These describe the data values in the column: how many there are; the number of unique values (so a column of all 0s would have just 1 unique value); if any values are excluded from analysis; the number of values marked as missing, left empty, or marked as damaged.

At this point, it's helpful to see that we didn't mark any as missing, and that none are empty (so we did not skip any values as we entered the data).



Next, look for the Outliers table. ARM uses the standardized residual as the basis for determining outliers in this table. This may differ from the Box-Whisker definition of an outlier, which is true here because only plot 203 is detected. This calculation is more conservative in the fact that it is less likely to find an outlier, because it takes into account the number of observations as well.

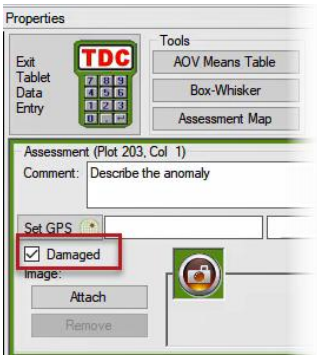
If the Outliers table does not appear, then the data column has no calculated outliers. You can confirm this by reviewing the 'Maximum Standardized Residual' statistic: a value less than 3.3 indicates that no values are considered an outlier.



The table lists the plot location, the assessment value, and the standardized residual value of each calculated outlier. Click on a row to move the cursor to that plot and inspect the value.

If we confirm that this value was not a measurement error or entry mistake, then we should document this anomaly. Back on the Properties panel, use the Comment field to describe the situation.

You can also take a picture for documentation – see our tutorial video for more information on taking plot pictures.



The value can also be thrown out of the analysis, by marking the plot as Damaged. This treats the value as a missing data point when performing an analysis, and draws a strikethrough in the value within ARM. However, this results in a lost of 1 error degree of freedom for each missing data point, resulting in lower statistical precision.

Unusual data does **not** mean *damaged* data

Thus, it is recommended to leave the data value as-is, and document it well. Only in situations where the plot was clearly affected by an outside force (like an irrigation leak or animal damage) is it valid to throw the value out.

Unusual data does **not** always mean damaged data! In the end, the documentation created with TDCx should be used when performing the data summary, to determine if this data should be included or not.