

Source Metadata References

Wrangle supports a set of variables, which can be used to programmatically reference aspects of the dataset or its source data. These **metadata references** allow you to create individual transformations of much greater scope and flexibility.

Tip: Some transformation steps make access to metadata about the original data source impossible to retain. It's best to use these references, where possible, early in your recipe. Additional information is available below.

Tip: You can use the `$filepath` and `$sourcerownumber` to create a primary key to identify source information for any row in your file-based datasets.

`$sourcerownumber`

The `$sourcerownumber` variable is a reference to the row number in which the current row originally appeared in the source of the data.

Tip: If the source row information is still available, you can hover over the left side of a row in the data grid to see the source row number in the original source data.

Limitations:

- The following transforms might make original row information invalid or otherwise unavailable. In these cases, the reference returns null values:
 - `pivot`
 - `flatten`
 - `join`
 - `lookup`
 - `union`
 - `unnest`
 - `unpivot`

Example:

The following example generates a new column containing the source row number for each row in the dataset, if available:

| | |
|-----------------------------------|--------------------------------|
| Transformation Name | New formula |
| Parameter: Formula type | Single row formula |
| Parameter: Formula | <code>\$sourcerownumber</code> |
| Parameter: New column name | <code>'src_rownumber'</code> |

If you have already used the `$filepath` reference, as in the previous example, you can combine these two columns to create a unique key to the source of each row:

| | |
|----------------------------|-------------|
| Transformation Name | New formula |
|----------------------------|-------------|

| | |
|-----------------------------------|---|
| Parameter: Formula type | Single row formula |
| Parameter: Formula | MERGE([src_filename,src_rownumber],'-') |
| Parameter: New column name | 'src_key' |

\$col

The \$col variable is a reference to the column that is currently being evaluated. This variable references the state of the current dataset, instead of the original source.

NOTE: This reference works only for the edit with formula transformation (set transform).

In the following example, all columns in the dataset that are of String type are converted to uppercase:

| | |
|----------------------------|--|
| Transformation Name | Edit column with formula |
| Parameter: Columns | All |
| Parameter: Formula | IF(ISMISMATCHED(\$col, ['String']), \$col, UPPER(\$col)) |

In the above, the wildcard applies the edit to each column. Each column is tested to see if it is mismatched with the String data type. If mismatched, the value in the column (\$col) is written. Otherwise, the value in the column is converted to uppercase (UPPER(\$col)).

Tip: \$col is useful for multi-column transformations.