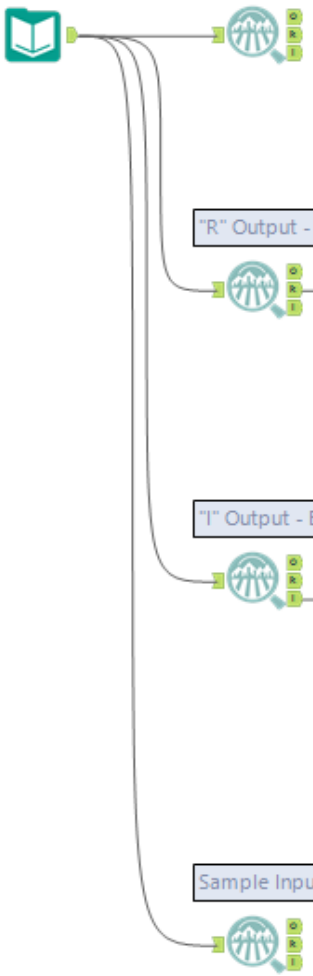
Field Summary Tool

Produce a concise summary report of descriptive statistics for the selected data columns.

1) Run the workflow (Ctrl+R).	2) Select a tool to view its output in the Results window.



"O" Output

The "O" output displays a table in the Results window with detailed information about each column in the incoming data. Use this tool first when inputting any new dataset.

Each column becomes a row in the results, with column type and additional information. 2 particularly useful columns are Percent Missing (nulls) and Unique Values (possible constants).

Note that row 11 (Y4) has 1 unique value, indicating that it is a constant and may need to be removed before being used in modeling.

"R" Output - Browse Tool Required

The "R" output produces a static report that appears in the Configuration window of a Browse tool.

For best results, open the report in a new window and expand it to view the Remarks for each column.

Note that the information in the report is grouped by data types.

The Remark for row 1 points out that the column Counts has a small number of unique values and might work better as a string. However, since these are counts, it is appropriate to have a sm

"I" Output - Browse Tool Required

The "I" output produces an interactive dashboard with expandable panels for each column. Each panel consists of a histogram, along with useful summary statistics.

Open the report in a new window for best results.

- Hover over a panel to display additional icons
- Hover over a bar in the plot to display tooltips
- Select the Expand icon in a panel to open a detailed view
- Use the Select variables to view box to focus on a smaller set of columns
- Sort the panels alphabetically or by percentage of missing values

Note the red section of the quality bar for column Y2. This indicates that 11% of this column is missing values.

Sample Input Data

Select the Sample input data option to take a random sample of rows, which can reduce the run time of a workflow when using a large dataset.

Each time you run the workflow, a different data sample will be displayed.

Note that sampling may exclude important information about the dataset as a whole.