SECTOR7G

https://www.sector7g.com

Using Spreadsheets to Plot Cumulative Distribution

December 10, 2021 • Rev 1.2

Supplement to Visualizing Performance with Cumulative Distribution

Overview

This documents plotting cumulative distribution using a spreadsheet and, as of this document's last revision, renders properly on desktop versions of Excel and the various flavors of Open Office. Since the steps are identical, this covers only Excel.

Prerequisites

- A flat file with a column containing the dataset you wish to plot (e.g., HTTP round trip times, CPU utilization, etc.).
- Backup of the original flat file.

Steps

1. Open the file in Excel and remove all columns not of interest to focus only on what's of interest.

2. Ensure:
    a. Column A is entirely blank.
    b. Only Column B contains data.

    Then:

    c. Ensure the top row is a header if not already.
    d. Enter "Pct" or "%" or any other value in A:1 indicating it's our cumulative percentages.
    e. Enter the name of our metric (if not already) in B:1; in this example, "Round Trip Time [ms]".

3. Sort column B *ascending*.

4. Enter the following formula in A:2

    ```
    =1/[number of samples]
    ```

    Where *[number of samples]* = number of rows-1 (recall first row is the header).
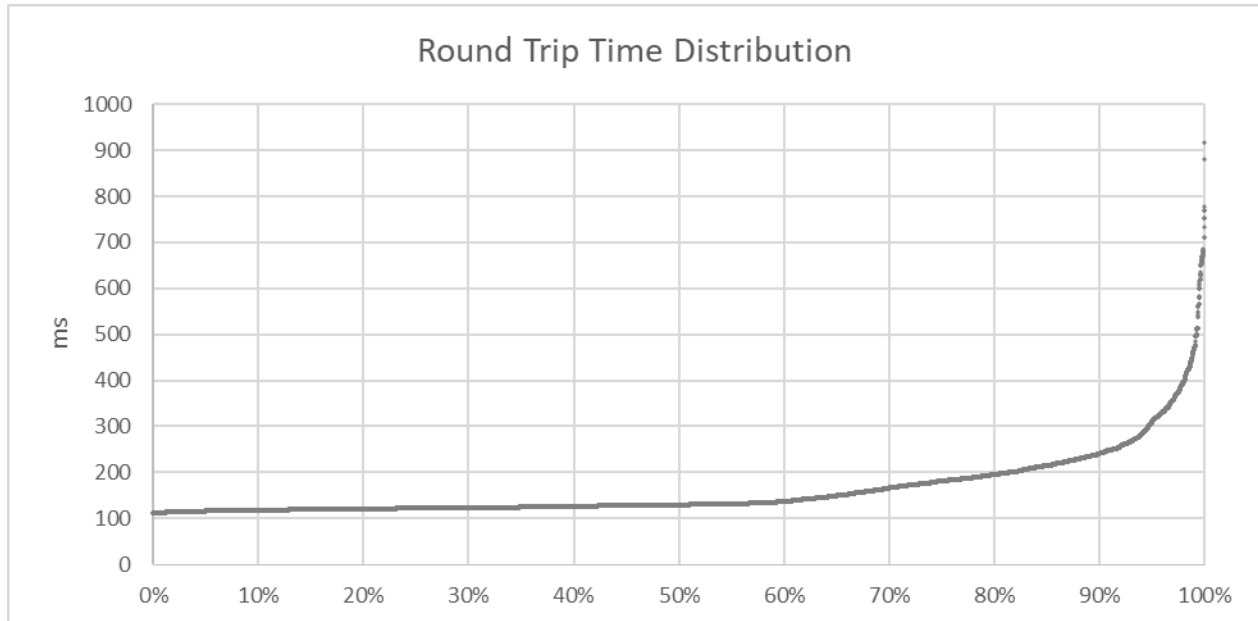   In this example, it's 12,023, so in A:2 we'd enter:

    ```
    =1/12023
    ```

5. Enter the following formula in A:3 (obviously change the number of samples to yours):

    ```
    =(1/12023)+A2
    ```

6. Apply the same formula to column A's remaining cells (in Excel, hover over the cell's lower-right "handle" and double-click to replicate the formula to the end).

7. Check the last cell in column A. If its value is not 1, something was skipped, mistyped, etc. and it's best to retry from scratch. The calculations, and therefore the plot, will be necessarily incorrect unless the final value is 1 since this equates to 100% of the samples (prior values may also be 1 with larger datasets and inherent rounding, but this is fine as long as the last value is exactly 1).

8. Create a **scatter plot** from column A's and B's data and adjust scales, labels, column A's format (e.g., whole number percent), etc.

[END]