AUTHORS

Vy Nguyen
v_nguyen110@u.pacific.edu
Advised by:
Dr. Solomon Berhe

Verifying Data Quality in Dashboards: A Case Study on ReleaseTrain.io

UNIVERSITY OF THE PACIFIC

School of Engineering & Computer Science
Summer 2025 Research Fellowship

01. Research Summary

This research project investigates how to verify data quality in dashboards using **ReleaseTrain.io** as a case study. It is an open-source framework that tracks software updates from CVE reports and social media. Building on a literature review of dashboard design and ISO data quality, we propose a five-step data quality verification process, which is applied to **ReleaseTrain.io**.

We identified 5 data quality invariants within its current dashboard and uncovered 4 inconsistencies across current visuals. From this, we proposed improved visualizations for the current dashboard to present the data more clearly and accurately to users.

02. Background

Dashboards are tools for displaying key insight to users through visualizations. However, traditional dashboards often struggle to present information effectively or communicate insights clearly (Lavelle et al., 2025). Bach et al. (2022) studied 144 dashboards and identified 42 common design patterns for presenting and arranging data to see how different visualization choices can affect clarity, reduce clutter, and make information easier to interpret.

International standards such as ISO 8000 also address the principles, characteristics, frameworks, and processes to ensure data quality across the data lifecycle. However, they do not specifically address how quality should be verified or communicated within dashboards.

This project addresses that gap by exploring verification processes and quality improvements for dashboard data, using ReleaseTrain.io as a case study.

03. Methodology

The research followed a two-step approach: (1) outlining a process for defining data quality invariants for dashboards, and (2) applying this process to evaluate and improve the ReleaseTrain.io dashboard.

The five steps of the process are described below.

1) Identify data sources:
2) Select CDEs:
3) Define invariants
4) Verify process & operations
5) Update dashboards

04. Results

We found 4 main inconsistencies in how data was displayed in the current visualizations.

1. Mismatches between charts:
Solution: Replace both with one stacked bar showing each type's share of the total.

Fig 1. Replacing "Update Type Count" and "CVE vs Other" with a stacked bar chart

2. Clustered component counts:
Solution: Show a sorted horizontal bar chart of top components (e.g., top 5–10) and exclude "Not Typed" to improve readability.

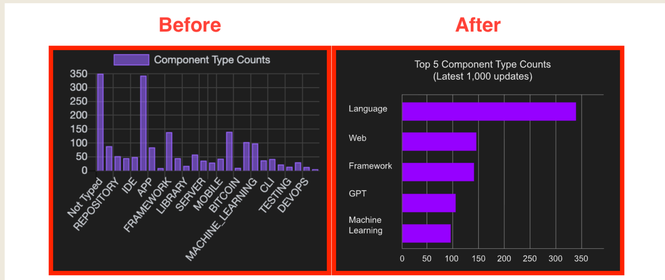

Fig 2. Simplifying clustered component counts by limiting to top 5 categories

3. Missing labels and context.
Solution: Add a clear label for data windows and definitions

4. Overcrowded recent updates
Solution: Add a stacked bar chart summarizing updates by severity (e.g., security, critical, breaking) per component.

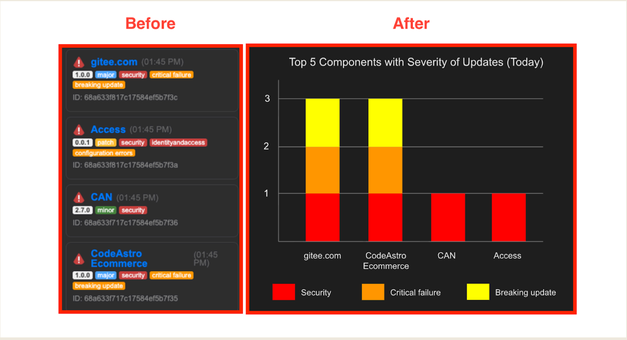

Fig 3. Simplifying Component Updates by Showing Top Components with Severity Levels

05. Conclusion & Takeaways

This research project highlights the importance of consistency, clarity, and transparency in dashboard design. Some key takeaways for designing effective dashboards with quality include:

1. Use one consistent data source for all charts.
2. Show data coverage (e.g., % included) on every view.
3. Define/clarify "Other" and any ambiguous labels.
4. Highlight missing values and anomalies/outliers
5. Keep time scales consistent across related charts.
6. Use clear labels, units, and definitions on every chart
7. Decide how much data to show at once effectively

06. References

Link to full report with reference list: