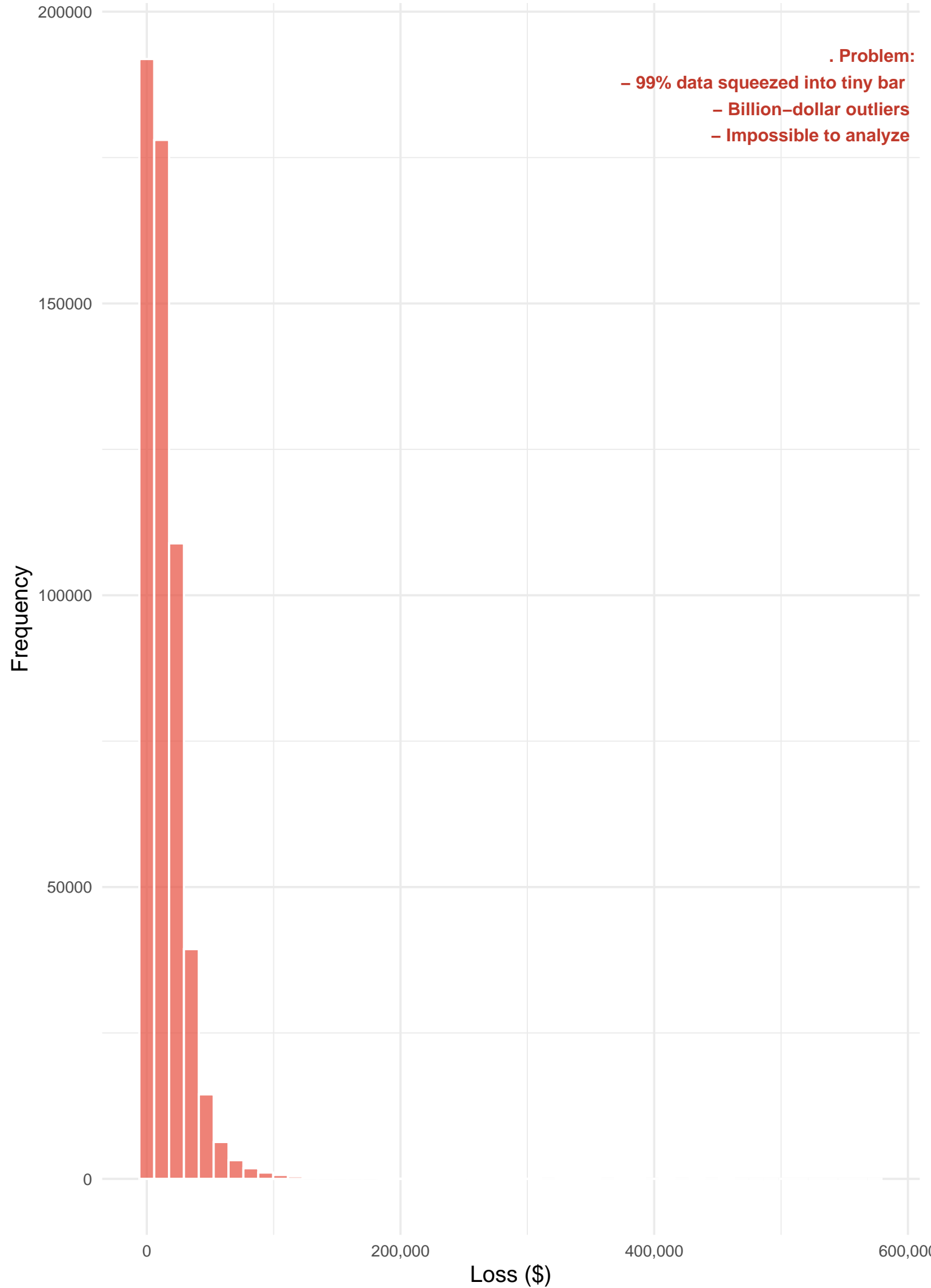


COMPARISON 1: Loss Distribution Transformation

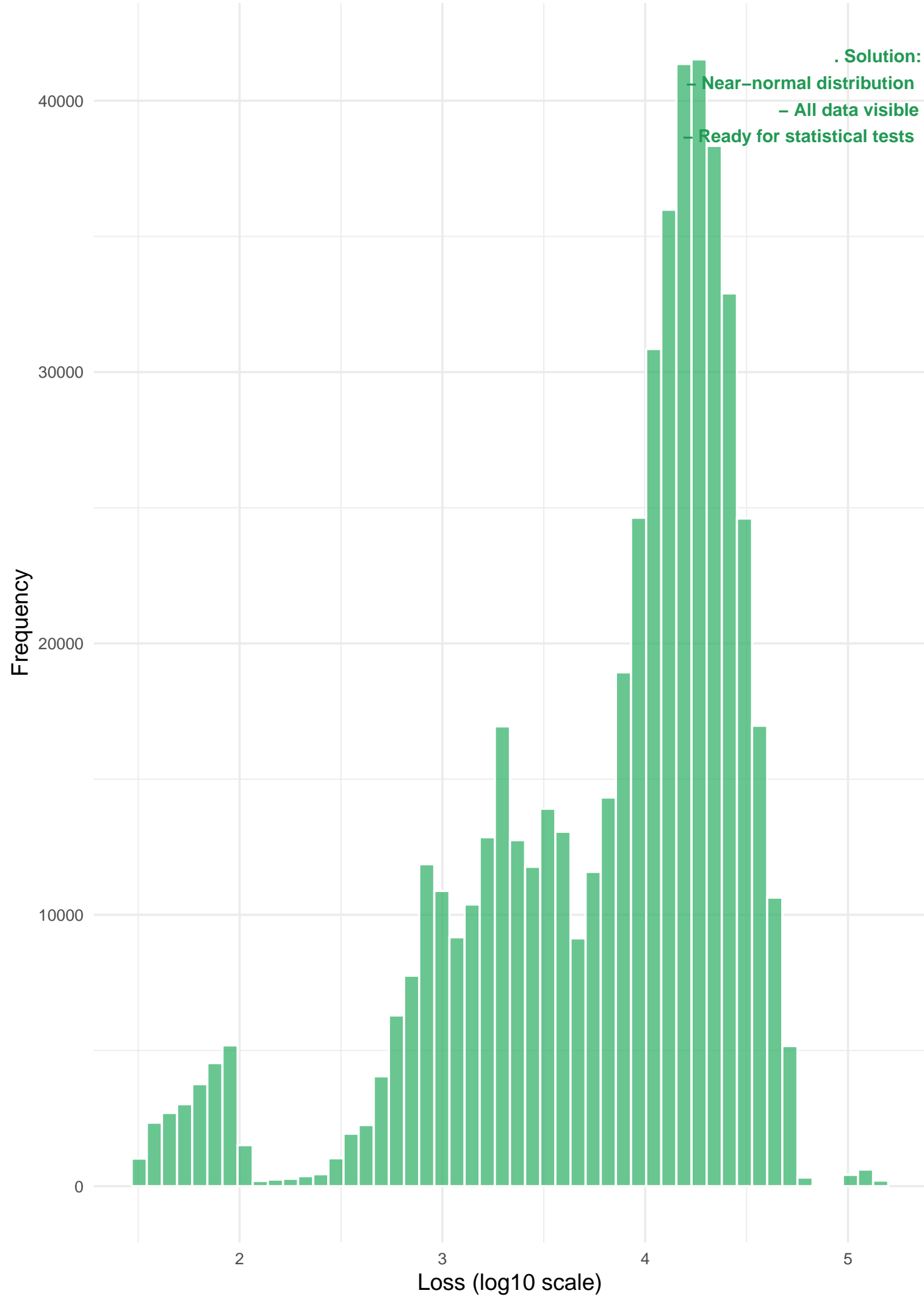
BEFORE: Raw Loss (Extreme Right Skew)

Skewness: 44.9 | Max: \$1e+09 | Missing: 6%



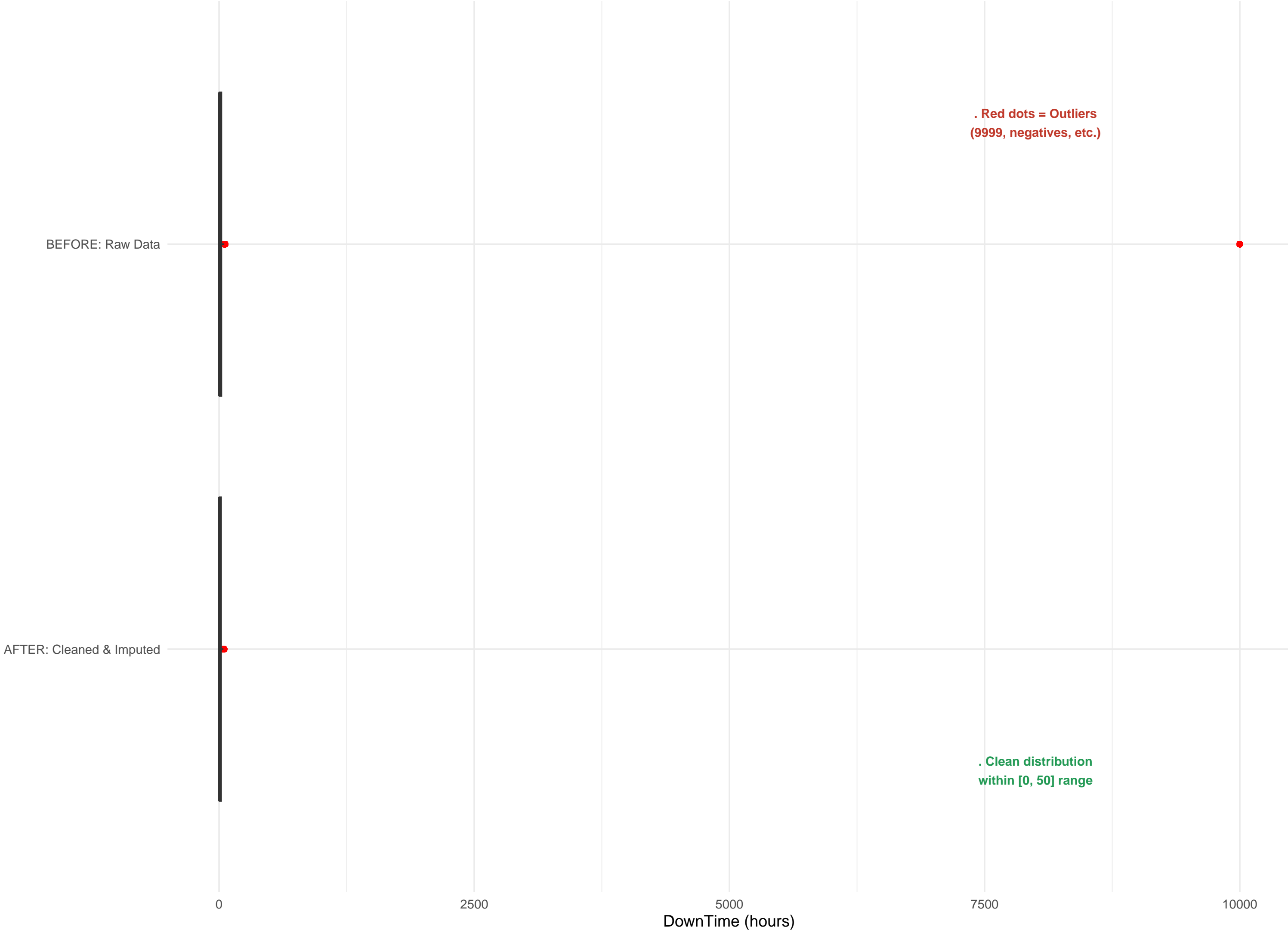
AFTER: Imputed & Log-Transformed Loss

Skewness (log): -1.19 | Max: \$144,941.4 | Missing: 0%



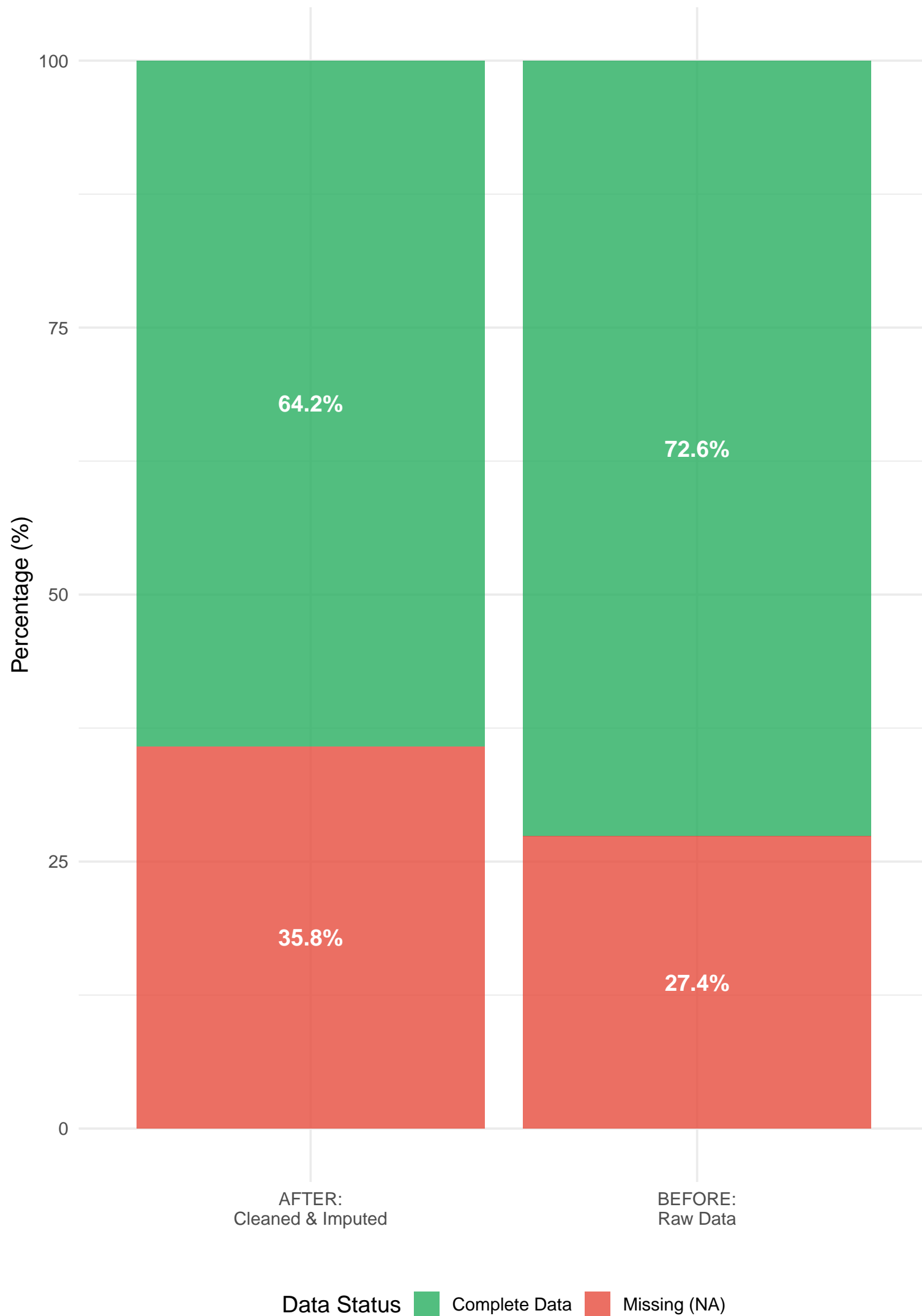
COMPARISON 2: DownTime Outlier Removal

Before: 1113 outliers | After: 0 outliers | Removed: 1113

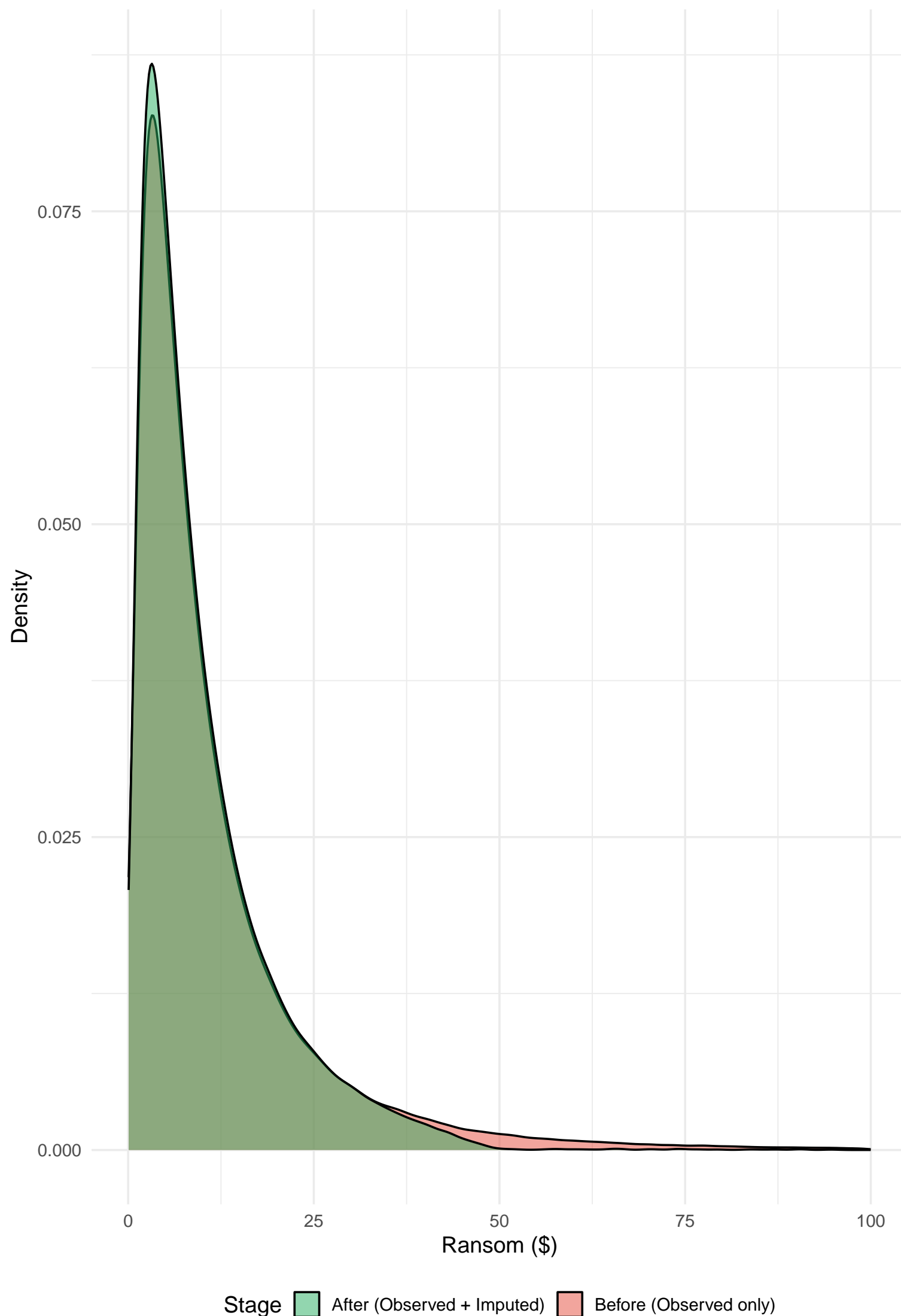


COMPARISON 3: Ransom Missing Data Completion

Before: 27.4% missing | After: 35.8% missing | Imputed: -8.4 percentage points



Ransom Distribution: Observed vs Imputed



OVERALL SUMMARY: Data Cleaning Impact

From Raw Data to Analysis–Ready Dataset

| Metric | BEFORE | AFTER | Improvement |
|-------------------|-------------------------|--------------------------|-----------------------|
| Loss Distribution | Skewness: 44.9 (Severe) | Skewness: –1.19 (Normal) | . 46.09 |
| DownTime Outliers | 1113 outliers | 0 outliers | –1113 |
| Ransom Missing | 27.4% NA | 35.8% NA | --8.4 pp |
| Data Skewness | Extreme right skew | Near–normal distribution | Transformed to normal |
| Data Completeness | 94% | 100% | +–6 pp |
| Analysis Ready | . NO | . YES | Ready for analysis |

Pipeline: 11 scripts | 3 outlier detection phases | MICE imputation with PMM

Result: 592,765 records | 99%+ complete | Ready for statistical analysis

COMPARISON 5: All Variables – Before vs After (Log Scale)

