Synthetic Gradebook Reconstruction Report

R-generated public-safe student score dataset

Contents

Purpose

This report summarizes the public-safe synthetic gradebook reconstruction workflow. The source artifacts are generated by R from private reference structure, but this report reads only synthetic outputs.

Load Synthetic Outputs

```
wide_gradebook <- read.csv("data/synthetic/synthetic_gradebook.csv", stringsAsFactors = FALSE, check.nar
score_records <- read.csv("data/synthetic/synthetic_student_scores_long.csv", stringsAsFactors = FALSE)
assignment_metadata <- read.csv("data/synthetic/synthetic_assignment_metadata.csv", stringsAsFactors = F

str(wide_gradebook)
```

```
## 'data.frame':    287 obs. of  39 variables:
##  $ Student             : chr  "Synthetic Student 001" "Synthetic Student 002" "Synthetic Student 003
##  $ ID                  : chr  "SYN-ID-000001" "SYN-ID-000002" "SYN-ID-000003" "SYN-ID-000004" ...
##  $ SIS User ID         : chr  "SYN000001" "SYN000002" "SYN000003" "SYN000004" ...
##  $ SIS Login ID        : chr  "synthetic001" "synthetic002" "synthetic003" "synthetic004" ...
##  $ Section             : chr  "Section A" "Section A" "Section A" "Section A" ...
##  $ Assignment 06       : chr  "Complete" "Exempt" "Submitted" "Submitted" ...
##  $ Current Score       : logi  NA NA NA NA NA NA ...
##  $ Final Score         : num  10 NA NA NA NA 12 26.4 NA NA NA ...
##  $ Unposted Final Score: num  10 NA NA NA NA 12 26.4 NA NA NA ...
##  $ Current Grade       : chr  "F" "" "" "" ...
##  $ Final Grade         : chr  "F" "" "" "" ...
##  $ Assignment 12       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 13       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 14       : int  NA NA NA NA NA NA 100 NA NA NA ...
##  $ Assignment 15       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 16       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 17       : int  NA NA NA NA NA NA 2 NA NA NA ...
##  $ Assignment 18       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 19       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 20       : int  NA NA NA NA NA NA 10 NA NA NA ...
```

```
##  $ Assignment 21       : int  10 NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 22       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 23       : int  NA NA NA NA NA NA 10 NA NA NA ...
##  $ Assignment 24       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 25       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 26       : int  NA NA NA NA NA NA 10 NA NA NA ...
##  $ Assignment 27       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 28       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 29       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 30       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 31       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 32       : int  NA NA NA NA NA 12 NA NA NA NA ...
##  $ Assignment 33       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 34       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 35       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 36       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 37       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 38       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Assignment 39       : int  NA NA NA NA NA NA NA NA NA NA ...
```

```
str(score_records)
```

```
## 'data.frame':    8036 obs. of  17 variables:
##  $ synthetic_student_id  : chr  "synthetic_student_001" "synthetic_student_002" "synthetic_student_0
##  $ synthetic_section     : chr  "Section A" "Section A" "Section A" "Section A" ...
##  $ assignment_id         : chr  "assignment_01" "assignment_01" "assignment_01" "assignment_01" ...
##  $ assignment_label      : chr  "Assignment 12" "Assignment 12" "Assignment 12" "Assignment 12" ...
##  $ assignment_sequence   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ assignment_family     : chr  "Diagnostic" "Diagnostic" "Diagnostic" "Diagnostic" ...
##  $ skill_domain          : chr  "Conceptual Fluency" "Conceptual Fluency" "Conceptual Fluency" "Conc
##  $ reference_column_index: int  12 12 12 12 12 12 12 12 12 12 ...
##  $ score                 : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ score_min             : int  100 100 100 100 100 100 100 100 100 100 ...
##  $ score_max             : int  100 100 100 100 100 100 100 100 100 100 ...
##  $ score_percent         : logi  NA NA NA NA NA NA ...
##  $ completed             : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ missingness_reason    : chr  "missing_submission" "missing_submission" "missing_submission" "not_a
##  $ ability_band          : chr  "approaching" "developing" "developing" "developing" ...
##  $ engagement_band       : chr  "steady" "steady" "low" "emerging" ...
##  $ risk_band             : chr  "elevated" "high" "high" "low" ...
```

```
str(assignment_metadata)
```

```
## 'data.frame':    28 obs. of  17 variables:
##  $ source_column_index     : int  12 13 14 15 16 17 18 19 20 21 ...
##  $ assignment_id           : chr  "assignment_01" "assignment_02" "assignment_03" "assignment_04" ..
##  $ assignment_label        : chr  "Assignment 12" "Assignment 13" "Assignment 14" "Assignment 15" ..
##  $ assignment_sequence     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ assignment_family       : chr  "Diagnostic" "Diagnostic" "Diagnostic" "Diagnostic" ...
##  $ skill_domain            : chr  "Conceptual Fluency" "Procedural Accuracy" "Modeling And Applicati
##  $ reference_nonblank_count: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ reference_blank_rate    : num  0.997 0.997 0.997 0.997 0.997 ...
##  $ reference_mean          : int  100 100 100 100 20 2 10 10 10 10 ...
##  $ reference_sd            : int  12 12 12 12 12 12 12 12 12 12 ...
##  $ reference_min           : int  100 100 100 100 20 2 10 10 10 10 ...
```

```
##  $ reference_p25           : int  100 100 100 100 20 2 2 2 2 2 ...
##  $ reference_p50           : int  100 100 100 100 20 2 2 2 2 2 ...
##  $ reference_p75           : int  100 100 100 100 20 2 2 2 2 2 ...
##  $ reference_max           : int  100 100 100 100 20 2 10 10 10 10 ...
##  $ difficulty_index        : num  -6.411 -6.411 -6.411 -6.411 0.256 ...
##  $ discrimination          : int  1 1 1 1 1 1 1 1 1 1 ...
```

Output Audit

```
data.frame(
  Metric = c(
    "Wide gradebook rows",
    "Wide gradebook columns",
    "Long-form score records",
    "Synthetic students",
    "Synthetic sections",
    "Assignment metadata rows",
    "Assignment families",
    "Skill domains"
  ),
  Value = c(
    nrow(wide_gradebook),
    ncol(wide_gradebook),
    nrow(score_records),
    length(unique(score_records$synthetic_student_id)),
    length(unique(score_records$synthetic_section)),
    nrow(assignment_metadata),
    length(unique(score_records$assignment_family)),
    length(unique(score_records$skill_domain))
  )
)
```

```
##                     Metric Value
## 1      Wide gradebook rows   287
## 2   Wide gradebook columns    39
## 3  Long-form score records  8036
## 4       Synthetic students   287
## 5       Synthetic sections     1
## 6 Assignment metadata rows    28
## 7      Assignment families     5
## 8            Skill domains     5
```

Completion And Missingness

```
completion_summary <- aggregate(
  completed ~ assignment_family,
  data = score_records,
  FUN = function(x) round(100 * mean(x), 1)
)
names(completion_summary)[names(completion_summary) == "completed"] <- "completion_rate"
completion_summary
```

```
##   assignment_family completion_rate
## 1     Concept Check             2.5
```

```
## 2 Cumulative Review             3.9
## 3        Diagnostic             3.4
## 4    Skill Practice             3.1
## 5   Unit Assessment             3.4
```

```r
missingness_summary <- as.data.frame(table(score_records$missingness_reason))
names(missingness_summary) <- c("missingness_reason", "records")
missingness_summary[order(-missingness_summary$records), ]
```

```
##   missingness_reason records
## 4 missing_submission    3169
## 5   not_administered    2053
## 3 late_or_incomplete    1410
## 2  excused_or_absent    1151
## 1          completed     253
```

Score Distribution

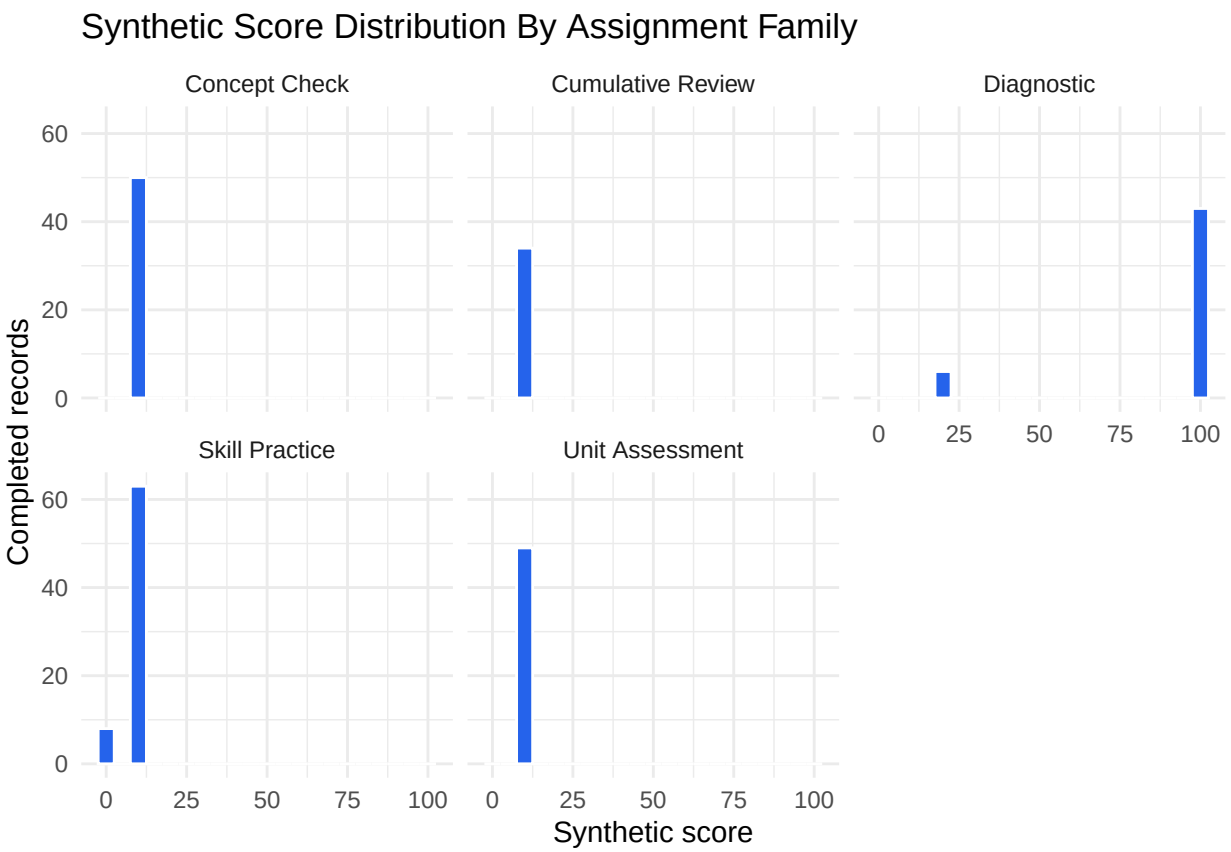
```r
completed_scores <- score_records[score_records$completed & !is.na(score_records$score), ]
if (any(!is.na(completed_scores$score_percent))) {
  completed_scores$score_metric <- completed_scores$score_percent
  score_metric_label <- "Score percent"
} else {
  completed_scores$score_metric <- completed_scores$score
  score_metric_label <- "Synthetic score"
}

data.frame(
  Metric = c("Completed score records", paste("Mean", score_metric_label), paste("Median", score_metric
  Value = c(
    nrow(completed_scores),
    round(mean(completed_scores$score_metric), 1),
    round(median(completed_scores$score_metric), 1),
    round(sd(completed_scores$score_metric), 1)
  )
)
```

```
##                    Metric Value
## 1 Completed score records 253.0
## 2    Mean Synthetic score  25.4
## 3  Median Synthetic score  10.0
## 4      Synthetic score SD  33.9
```

```r
if (requireNamespace("ggplot2", quietly = TRUE)) {
  ggplot2::ggplot(completed_scores, ggplot2::aes(score_metric)) +
    ggplot2::geom_histogram(binwidth = 5, fill = "#2563eb", color = "white") +
    ggplot2::facet_wrap(~ assignment_family) +
    ggplot2::labs(
      title = "Synthetic Score Distribution By Assignment Family",
      x = score_metric_label,
      y = "Completed records"
    ) +
    ggplot2::theme_minimal()
} else {
  hist(completed_scores$score_metric, breaks = 20, main = "Synthetic Score Distribution", xlab = score_
```

```
}
```



Latent Pattern Checks

```
aggregate(
  score_metric ~ ability_band + engagement_band,
  data = completed_scores,
  FUN = function(x) round(mean(x), 1)
)
```

```
##    ability_band engagement_band score_metric
## 1      advanced        emerging         28.0
## 2   approaching        emerging         10.0
## 3    developing        emerging         28.4
## 4    proficient        emerging         46.0
## 5      advanced            high         26.6
## 6   approaching            high         25.7
## 7    developing            high         28.4
## 8    proficient            high         23.9
## 9      advanced             low         10.0
## 10  approaching             low         32.5
## 11   developing             low         28.8
## 12   proficient             low         55.0
## 13     advanced          steady         18.6
## 14  approaching          steady         28.0
## 15   developing          steady         10.0
```

```
## 16   proficient          steady         26.0
```

```
aggregate(
  completed ~ risk_band,
  data = score_records,
  FUN = function(x) round(100 * mean(x), 1)
)
```

```
##   risk_band completed
## 1  elevated       0.7
## 2      high       0.4
## 3       low       9.6
## 4  moderate       2.0
```

Interpretation

The synthetic dataset is designed to support realistic analytics workflows:

- assignment-family comparisons
- completion and missingness monitoring
- score-distribution analysis
- student-risk and engagement segmentation
- public-safe dashboard development

The report is not evidence about a real class or school. It is a reproducible artifact demonstrating how private reference structure can be transformed into a public-safe dataset for analysis and portfolio presentation.