

Behind, and getting further behind

Over-age enrolment runs 17.0% in grade 1 and 50.8% in grade 6. It rises because repetition creates it — 94.5% of the children who repeated in 2023 are over-age in 2024 — and because being over-age quadruples the chance of leaving.

Pierre Robentz Cassion

Lesson 2 of 8

Education Programme Analysis — Enrolment and its denominator

What this lesson covers

- The definition, and the range check that has to come first
- It compounds with grade
- Why it matters: the dropout link
- The disaggregation that does not work
- Report the profile, not the headline
- What comes next

The definition, and the range check that has to come first — In Python

```
import pandas as pd

enrolment = pd.read_csv("school-enrolment-2024.v1.csv")
print(enrolment["age_years"].describe().round(1))
print(f"ages above 20: {(enrolment['age_years'] > 20).sum()}")
```

The definition, and the range check that has to come first — In R

```
library(dplyr)
enrolment |> summarise(min = min(age_years), max = max(age_years),
                      impossible = sum(age_years > 20))
```

The definition, and the range check that has to come first

- Twenty-one rows hold an age above 20 in a primary register — SCH04 records part of its intake in months, so a . . .

The definition, and the range check that has to come first — In Python

```
clean = enrolment[enrolment["age_years"] <= 20].copy()
clean["official_age"] = clean["grade"] + 5
clean["over_age"] = clean["age_years"] > clean["official_age"]
```

The definition, and the range check that has to come first — In R

```
enrolment |> filter(age_years <= 20) |>  
  mutate(official_age = grade + 5, over_age = age_years > official_age)
```

The definition, and the range check that has to come first

- Do the range check before the flag, not after — An age of 112 is over-age for every grade, so a flag computed on the...

It compounds with grade — In Python

```
primary = clean[(clean["school_year"] == 2024) & clean["grade"].between(1, 6)]
by_grade = primary.groupby("grade").agg(
    students=("student_id", "size"),
    over_age=("over_age", "mean"),
)
print((by_grade * [1, 100]).round(1))
```

It compounds with grade — In R

```
primary |> summarise(n = n(), over_age = mean(over_age), .by = grade)
```

It compounds with grade

Grade	Official age	Students	Over-age
1	6	53	17.0%
2	7	153	25.5%
3	8	247	29.6%
4	9	258	36.8%
5	10	209	47.8%
6	11	132	50.8%

It compounds with grade

- Over-age triples between grade 1 and grade 6 — Two mechanisms produce that shape and they are not the same problem
- Late entry — puts a child above age at grade 1 and keeps them there
- Repetition — creates over-age *during* schooling

It compounds with grade — In Python

```
previous = clean[clean["school_year"] == 2023].set_index("student_id")
current = clean[clean["school_year"] == 2024].copy()
current["last_year"] = current["student_id"].map(previous["end_of_year_status"])

print(current.groupby("last_year")["over_age"].agg(["mean", "size"]).round(3))
```

It compounds with grade — In R

```
enrolment |>
  filter(age_years <= 20) |>
  select(student_id, school_year, grade, age_years, end_of_year_status) |>
  tidyr::pivot_wider(names_from = school_year,
                     values_from = c(grade, age_years, end_of_year_status))
```

It compounds with grade

- 94.5% of the students who repeated in 2023 are over-age in 2024, against 30.6% of those who were promoted — Repetition. . .

Why it matters: the dropout link — In Python

```
year23 = clean[clean["school_year"] == 2023]
dropout = year23.groupby("over_age")["end_of_year_status"].apply(
    lambda s: (s == "dropped-out").mean()
)
counts = year23.groupby("over_age").size()
print(pd.DataFrame({"dropout": (dropout * 100).round(1), "n": counts}))
```

Why it matters: the dropout link — In R

```
enrolment |> filter(school_year == 2023) |>  
  summarise(dropout = mean(end_of_year_status == "dropped-out"),  
            n = n(), .by = over_age)
```

Why it matters: the dropout link

	Students	Dropped out
Over-age	478	19.5%
In-age	814	4.8%

Why it matters: the dropout link

- Being behind quadruples the chance of leaving — and that is what closes the loop: a child repeats, becomes over-age, . . .
- So repetition is not a neutral intervention — It is offered as a second chance and it measurably raises the probability. . .

The disaggregation that does not work — In Python

```
for column in ("sex", "disability_reported", "displacement_status"):
    print(primary.groupby(column, dropna=False)["over_age"].agg(
        ["mean", "size"]).round(3), "\n")
```

The disaggregation that does not work — In R

```
primary |> summarise(over_age = mean(over_age), n = n(), .by = sex)
```

The disaggregation that does not work

Cut	Over-age	n
Boys	39.2%	510
Girls	33.8%	542
Disability reported	33.7%	89
No disability reported	36.7%	963

The disaggregation that does not work

- Five points between boys and girls, on 510 and 542 students — That is around the edge of what sampling variation. . .
- Disability shows three points the other way on 89 students — which is noise on a denominator that small
- A disaggregation that mostly shows nothing is a result — Over-age here is a system-level phenomenon driven by. . .

Report the profile, not the headline — Example

Over-age enrolment, primary, 2024

Overall	36.4%	383 of 1,052 primary enrollees
Grade 1	17.0%	the late-entry floor
Grade 6	50.8%	after five years of repetition
Repeated in 2023	94.5%	over-age in 2024, against 30.6% promoted
Dropout, over-age	19.5%	against 4.8% in-age (2023 cohort)

21 ages recorded in months at SCH04, corrected by division.

Sex difference is 39.2% (boys) against 33.8% (girls); report with an interval or not at all.

Report the profile, not the headline

- The grade profile is the deliverable — because 36.4% overall could mean uniform late entry or accumulating repetition, . . .

What comes next

- Every child in this lesson is enrolled.

Read the full lesson, with runnable code

[https://data-analysis.cassion.dev/courses/education-programme-analysis/
edu-02-the-over-age-student](https://data-analysis.cassion.dev/courses/education-programme-analysis/edu-02-the-over-age-student)