

A gap of half a point, filling half the chart

Boys attend 88.77% and girls 88.21%. On a zero-based axis that difference occupies 0.6% of the chart. Truncate the axis to 88–89 and it occupies 55.7%, a hundred-fold magnification, with every number still correct.

Pierre Robentz Cassion

Lesson 7 of 8

Data Visualisation for Programmes — Charts that cannot drift

What this lesson covers

- The same two numbers, four charts
- The rule, and the exception
- When the truncation is the honest choice
- Five more axis decisions that change the reading
- The missing-data trap, which is an axis problem
- Report it whole
- What comes next

The same two numbers, four charts — In Python

```
import matplotlib.pyplot as plt

rates = {"Boys": 0.8877, "Girls": 0.8821}
gap = (rates["Boys"] - rates["Girls"]) * 100

for lo, hi in [(0, 100), (80, 95), (87.5, 89.5), (88, 89)]:
    share = gap / (hi - lo) * 100
    print(f"axis {lo}-{hi}: the gap fills {share:.1f}% of the chart height")
```

The same two numbers, four charts — In R

Same data, four y-limits, four different stories.

The same two numbers, four charts

Axis	The gap occupies	Magnification
0 to 100	0.6% of the chart	×1
80 to 95	3.7%	×7
87.5 to 89.5	27.8%	×50
88 to 89	55.7%	×100

The same two numbers, four charts

- Every one of those charts is accurate — The bars are in the right places, the axis is labelled, the numbers are correct
- Truncating a bar chart's axis is the single most effective way to mislead with correct data — and it is done far more. . .

The rule, and the exception

- A bar chart's value axis starts at zero. Always — A bar encodes its quantity as a length, and a length measured from a...
- A line chart's axis does not have to — A line encodes change as slope, and the reader is comparing points to each other...

The rule, and the exception — In Python

```
fig, (a, b) = plt.subplots(1, 2, figsize=(9, 3))
a.bar(rates.keys(), rates.values())           # must include zero
a.set_ylim(0, 1)
b.plot(months, completeness)                  # need not include zero
b.set_ylim(0.25, 1.0)
```

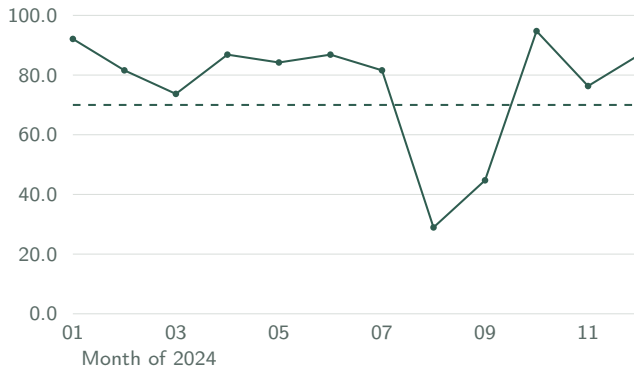
The rule, and the exception — In R

```
# geom_col() + expand_limits(y = 0). geom_line() + coord_cartesian(ylim = ...).
```

The rule, and the exception

- And when you truncate a line's axis, say so on the axis — A break mark, or an axis label that starts visibly above. . .

When the truncation is the honest choice



Monthly reporting completeness against the 70% floor. Completeness collapses to 28.9% in August and recovers to 94.7% in October, which is a reporting failure rather than a change in the epidemic.

When the truncation is the honest choice — In Python

```
values = [.921, .816, .737, .868, .842, .868, .816, .289, .447, .947, .763, .868]  
print(f"range {min(values):.1%} to {max(values):.1%}")
```

When the truncation is the honest choice — In R

```
range(values)
```

When the truncation is the honest choice

- Completeness runs from 28.9% to 94.7% — a range of 65.8 points across a 100-point axis — The data uses two-thirds of...
- The rule to apply is: does the truncation change how large the variation looks? — Here it barely would, so there is no...

Five more axis decisions that change the reading

- Reversed axes — A rate where lower is better — mortality, defaulting, stunting — drawn with the worst value at the top. . .
- Log scales — Legitimate for an outbreak curve spanning three orders of magnitude, and misread by almost every. . .
- Percentage axes above 100 — Gross enrolment reaches 109.2% in the education course's data, and an axis capped at 100. . .
- Time axes with uneven spacing — Twelve monthly points drawn at even spacing when two months are missing draws a. . .
- Aspect ratio — The same line at twice the height looks twice as steep

Five more axis decisions that change the reading — In Python

```
fig.set_size_inches(6.5, 4)      # every figure in the report, the same shape
```

Five more axis decisions that change the reading — In R

ggsave(width = 6.5, height = 4) as a project default.

The missing-data trap, which is an axis problem

- A month with no reports drawn as zero is a claim that nothing happened — A month with no reports left out of the series. . .

The missing-data trap, which is an axis problem — In Python

```
series = series.reindex(all_months)          # missing months become NaN  
ax.plot(series.index, series.values)         # matplotlib breaks the line at NaN
```

The missing-data trap, which is an axis problem — In R

Do not `tidyr::replace_na(0)`. A gap is data about the reporting system.

The missing-data trap, which is an axis problem

- August's 28.9% is a reporting failure, not an epidemiological one — The epidemiology course spent a lesson on it, and...

Report it whole — Example

Figure axis conventions

Bar charts use zero-based value axes without exception. Where a difference is too small to be visible on a zero-based axis, it is reported as a sentence rather than as a chart.

Line charts may use a truncated axis where the data range justifies it; the axis is labelled so the truncation is visible.

Indicators where lower is better carry "lower is better" in the axis label.

Months with no data are drawn as gaps, never as zero. Completeness is reported alongside any indicator whose denominator varies by month.

All figures in this report are drawn at 6.5 by 4 inches, so vertical changes are comparable between charts.

Report it whole

- The first paragraph is the one worth adopting verbatim — "Report it as a sentence rather than as a chart" is the...

What comes next

- Every rule in this course is a rule someone has to remember on the day they make the chart.

Read the full lesson, with runnable code

[https://data-analysis.cassion.dev/courses/data-visualisation/
viz-07-the-axis-that-changes-the-finding](https://data-analysis.cassion.dev/courses/data-visualisation/viz-07-the-axis-that-changes-the-finding)