

Un écart d'un demi-point, remplissant la moitié du graphique

Les garçons sont présents à 88,77 % et les filles à 88,21 %. Sur un axe partant de zéro, cet écart occupe 0,6 % du graphique. Tronquez l'axe à 88-89 et il en occupe 55,7 %, un grossissement de cent fois, sans qu'aucun nombre cesse d'être juste.

Pierre Robentz Cassion

Leçon 7 sur 8

Visualisation de données pour les programmes — Des graphiques qui ne peuvent pas dériver

Ce que couvre cette leçon

- Les deux mêmes nombres, quatre graphiques
- La règle, et l'exception
- Quand la troncature est le choix honnête
- Cinq autres décisions d'axe qui changent la lecture
- Le piège des données manquantes, qui est un problème d'axe
- Rapportez-le en entier
- La suite

Les deux mêmes nombres, quatre graphiques — En 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")
```

Les deux mêmes nombres, quatre graphiques — En R

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

Les deux mêmes nombres, quatre graphiques

Axe	L'écart occupe	Grossissement
0 à 100	0,6 % du graphique	×1
80 à 95	3,7 %	×7
87,5 à 89,5	27,8 %	×50
88 à 89	55,7 %	×100

Les deux mêmes nombres, quatre graphiques

- Chacun de ces graphiques est exact — Les barres sont au bon endroit, l'axe est étiqueté, les nombres sont justes
- Tronquer l'axe d'un diagramme en barres est la façon la plus efficace d'induire en erreur avec des données justes — et. . .

La règle, et l'exception

- L'axe des valeurs d'un diagramme en barres part de zéro. Toujours — Une barre encode sa quantité par une longueur, et. . .
- L'axe d'un graphique en ligne n'y est pas tenu — Une ligne encode l'évolution par la pente, et le lecteur compare des. . .

La règle, et l'exception — En Python

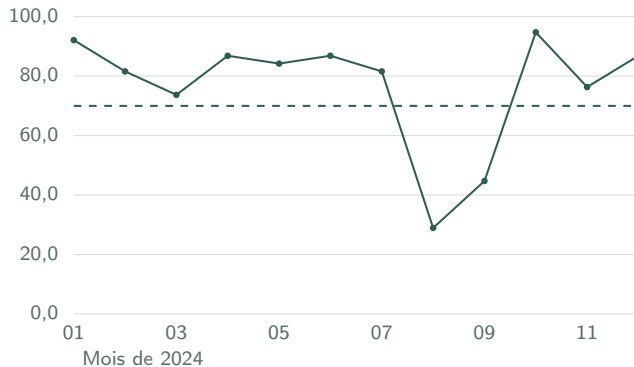
```
fig, (a, b) = plt.subplots(1, 2, figsize=(9, 3))
a.bar(rates.keys(), rates.values())           # must include zero
a.set_ylim(0, 1)
b.plot(months, completeness)                  # need not include zero
b.set_ylim(0.25, 1.0)
```

La règle, et l'exception — En R

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

- Et quand vous tronquez l'axe d'une ligne, dites-le sur l'axe — Une marque de rupture, ou une étiquette d'axe qui. . .

Quand la troncature est le choix honnête



Complétude mensuelle des rapports contre le plancher de 70 %. La complétude s'effondre à 28,9 % en août et remonte à 94,7 % en octobre, ce qui est une défaillance de notification et non un changement de l'épidémie.

Quand la troncature est le choix honnête — En Python

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

Quand la troncature est le choix honnête — En R

```
range(values)
```

Quand la troncature est le choix honnête

- La complétude va de 28,9 % à 94,7 % — une étendue de 65,8 points sur un axe de 100 points — Les données emploient les. . .
- La règle à appliquer est : la troncature change-t-elle l'ampleur apparente de la variation ? — Ici, à peine, il n'y a. . .

Cinq autres décisions d'axe qui changent la lecture

- Les axes inversés — Un taux où plus bas vaut mieux — mortalité, abandon, retard de croissance — tracé avec la pire. . .
- Les échelles logarithmiques — Légitimes pour une courbe épidémique couvrant trois ordres de grandeur, et lues comme. . .
- Les axes de pourcentage au-delà de 100 — Le taux brut de scolarisation atteint 109,2 % dans les données du cours. . .
- Les axes temporels à espacement irrégulier — Douze points mensuels tracés à espacement régulier alors que deux mois. . .
- Le rapport hauteur-largeur — La même ligne à hauteur double paraît deux fois plus pentue

Cinq autres décisions d'axe qui changent la lecture — En Python

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

Cinq autres décisions d'axe qui changent la lecture — En R

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

Le piège des données manquantes, qui est un problème d'axe

- Un mois sans rapports tracé comme zéro est une affirmation que rien ne s'est passé — Un mois sans rapports retiré. . .

Le piège des données manquantes, qui est un problème d'axe — En Python

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

Le piège des données manquantes, qui est un problème d'axe — En R

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

Le piège des données manquantes, qui est un problème d'axe

- Les 28,9 % d'août sont une défaillance de notification, non une défaillance épidémiologique — Le cours d'épidémiologie. . .

Rapportez-le en entier — Exemple

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.

- Le premier paragraphe vaut d'être adopté tel quel — « Rapportez-le comme une phrase plutôt que comme un graphique » est. . .

- Chaque règle de ce cours est une règle que quelqu'un doit se rappeler le jour où il fait le graphique.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/data-visualisation/
viz-07-the-axis-that-changes-the-finding`