

## Twelve communes, one ranking, eleven overlapping intervals

Sorted by prevalence the communes look like a league table. Every one of the eleven adjacent pairs has overlapping intervals, and only three clear the district median. A chart without intervals invites a ranking the data cannot support.

---

Pierre Robentz Cassion

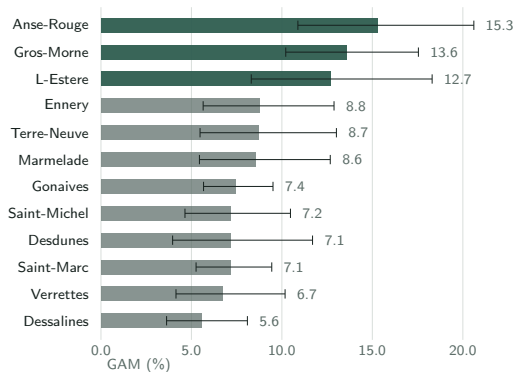
Lesson 2 of 8

Data Visualisation for Programmes — The mark and the comparison

## What this lesson covers

- The ranking the chart invites
- Drawing the interval
- When intervals are not the right encoding
- The reference line does more work than the intervals
- Reading order, and the three that clear
- Report it whole
- What comes next

# The ranking the chart invites



Global acute malnutrition by commune with 95% confidence intervals. The three highlighted communes are the only ones whose interval clears the district median.

# The ranking the chart invites — In Python

```
import pandas as pd

communes = pd.DataFrame({
    "commune": ["Anse-Rouge", "Gros-Morne", "L-Estere", "Ennery", "Terre-Neuve",
               "Marmelade", "Gonaives", "Saint-Michel", "Desdunes",
               "Saint-Marc", "Verrettes", "Dessalines"],
    "gam":     [.153, .136, .127, .088, .087, .086, .074, .072, .071, .071, .067, .056],
    "low":     [.109, .102, .083, .056, .055, .054, .057, .046, .040, .053, .041, .036],
    "high":    [.206, .175, .183, .129, .130, .127, .095, .105, .117, .094, .102, .081],
})

overlaps = sum(communes["high"][i] > communes["low"][i - 1]
               for i in range(1, len(communes)))
print(f"adjacent pairs whose intervals overlap: {overlaps} of {len(communes) - 1}")
```

# The ranking the chart invites — In R

*# Eleven of eleven. The ranking is a ranking of point estimates only.*

## The ranking the chart invites

- Every adjacent pair overlaps — The chart can support "these three are higher than the district median" and cannot. . .
- The interval is what converts a ranking into a finding — Without it, the reader supplies an ordering the data does not. . .

## Drawing the interval — In Python

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(7, 4))
ax.barh(communes["commune"], communes["gam"])
ax.errorbar(communes["gam"], communes["commune"],
            xerr=[communes["gam"] - communes["low"],
                  communes["high"] - communes["gam"]],
            fmt="none", capsize=3)
ax.invert_yaxis()
```

## Drawing the interval — In R

```
ggplot(communes, aes(x = gam, y = reorder(commune, gam))) +  
  geom_col() +  
  geom_errorbarh(aes(xmin = low, xmax = high), height = 0.3)
```

## Drawing the interval

- Say what the interval is — "95% confidence interval" in the caption, every time
- Use the right interval — These are Clopper–Pearson exact binomial intervals, because several communes have small...
- Do not clip a bound to make it fit — An interval that reaches 20.6% sets the axis; truncating the axis to 15% and...
- Whiskers on top of bars, not instead of them — For a proportion the bar carries the magnitude and the whisker carries...

## When intervals are not the right encoding

- A census, not a sample — A count of every water point in a district has no sampling error
- A total, not a rate — "1,850 cases registered" is a count of what happened
- A time series a reader has to see the shape of — Twelve monthly points with whiskers becomes a hedge; a shaded band...

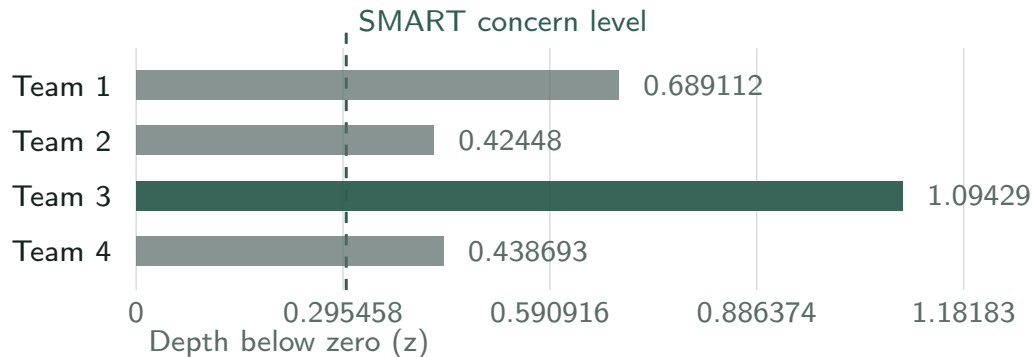
## When intervals are not the right encoding — In Python

```
ax.fill_between(months, lower, upper, alpha=0.2)  
ax.plot(months, values)
```

## When intervals are not the right encoding — In R

```
geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2) + geom_line()
```

## The reference line does more work than the intervals



Weight-for-height mean by survey team, against the SMART concern level. One team's mean sits beyond the threshold at which SMART flags a measurement problem.

## The reference line does more work than the intervals

- Sphere thresholds, SMART concern levels, IPC phase boundaries, national targets, a previous round — Each is a reference. . .
- Label the line with what it is, not with its value — "70% floor" beats "0.70", which the reader can already see on the. . .
- One reference line — Two is a comparison the reader has to perform; three is decoration

## Reading order, and the three that clear — In Python

```
median = communes["gam"].median()
clears = communes[communes["low"] > median]
print(f"median {median:.1%}; clearing it: {list(clears['commune'])}")
```

## Reading order, and the three that clear — In R

```
communes |> filter(low > median(gam)) |> pull(commune)
```

## Reading order, and the three that clear

- Anse-Rouge, Gros-Morne and L'Estère are the only three whose lower bound exceeds the district median of 8.0% — That is. . .
- Emphasis is an argument, so make it the argument the caption makes — Highlighting the top three by value would have. . .

# Report it whole — Example

Global acute malnutrition by commune, Artibonite, 2024

Prevalence by commune with 95% Clopper-Pearson intervals, sorted by point estimate. n ranges from 189 to 768 screenings per commune.

Three communes -- Anse-Rouge, Gros-Morne and L'Estere -- have lower bounds above the district median of 8.0% and are highlighted.

All eleven adjacent pairs have overlapping intervals. The chart supports "these three are above the district median" and does not support a ranking of one commune against its neighbour.

Intervals are exact binomial. Communes with fewer than 300 screenings have intervals roughly twice as wide as Gonaives and Saint-Marc, which is a property of where the screening teams went rather than of their nutrition situation.

## Report it whole

- The third paragraph is what stops the chart being used as a league table — and it is the sentence a reader will not. . .

## What comes next

- Both figures in this lesson use one colour for the ordinary bars and one for the emphasised ones.

Read the full lesson, with runnable code

[https://data-analysis.cassion.dev/courses/data-visualisation/  
viz-02-uncertainty-on-the-page](https://data-analysis.cassion.dev/courses/data-visualisation/viz-02-uncertainty-on-the-page)