The comparison decides the mark

Four food security instruments on the same 1,955 households give 7.4%, 42.9%, 48.2% and 51.0%. Written as a sentence a reader nods; drawn as four bars nobody reads past it. The chart type was decided by the comparison, not by preference.

Pierre Robentz Cassion

What this lesson covers

- Start from the sentence, not the library
- The four-instrument chart
- The chart that is a table
- Where something stops
- Three marks to avoid, and what to use instead
- The stacked bar, which is sometimes right
- Report it whole
- What comes next

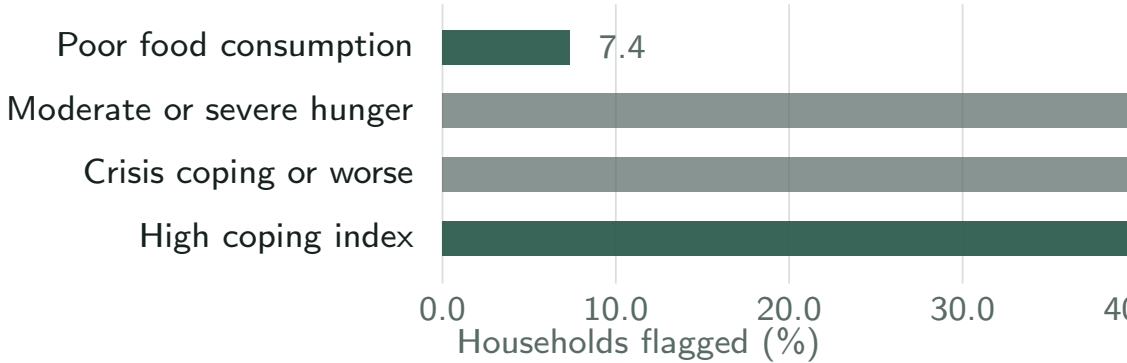
Start from the sentence, not the library

The sentence compares	Mark	Why
Categories against each other	**Sorted horizontal bar**	Length is the most accurately read encoding, and labels fit
One quantity over time	**Line**	Slope is the comparison; a bar chart of months hides it
A part against its whole	**A single bar, or the number**	A pie is two numbers and a legend
Two continuous variables	**Scatter**	Position on both axes
The same comparison across places	**Small multiples**	One shape repeated, read once
Where something stops	**Ordered stages**	The drop between bars is the finding

Start from the sentence, not the library

- Nothing on this platform has yet needed a mark outside that list — which is a fact about programme reporting rather...

The four-instrument chart



Four food security instruments applied to the same 1,955 households. The highest prevalence is seven times the lowest, so the instrument chosen decides the caseload before any analysis begins.

The four-instrument chart — In Python

```
import pandas as pd
import matplotlib.pyplot as plt

prevalence = pd.Series({
    "Poor food consumption": 0.074,
    "Moderate or severe hunger": 0.429,
    "Crisis coping or worse": 0.482,
    "High coping index": 0.510,
})

fig, ax = plt.subplots(figsize=(7, 2.6))
ax.barh(prevalence.index, prevalence.values)
ax.invert_yaxis()                          # first row at the top
ax.set_xlabel("Households flagged (%)")
```

The four-instrument chart — In R

```
library(ggplot2)

ggplot(prevalence, aes(x = share, y = reorder(instrument, share))) +
  geom_col() +
  labs(x = "Households flagged (%)", y = NULL)
```

The four-instrument chart

- Horizontal, because the labels are words — "Moderate or severe hunger" rotated 45 degrees under a vertical bar is a...
- Sorted, unless the order means something — Alphabetical order is an ordering nobody asked about
- One exception worth knowing — When the categories have their own order — IPC phases, MUAC bands, months, education...

The chart that is a table — In Python

```
coverage = pd.Series({"Measles first dose": 0.87, "Measles second dose": 0.61})
```

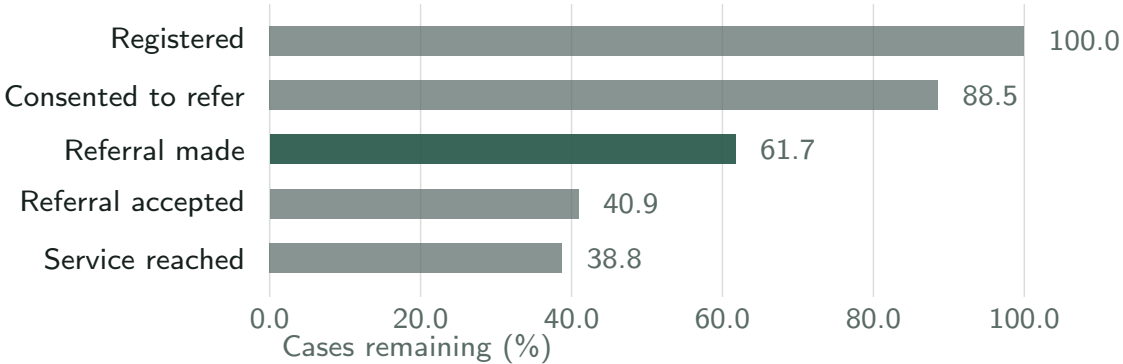
The chart that is a table — In R

```
# Two numbers. A chart of two numbers is a chart of nothing.
```

The chart that is a table

- Two or three numbers belong in a sentence — "Coverage falls from 87% to 61% between the first and second dose" is. . .
- A chart earns its space at about five categories and stops being read at about twenty-five — Below five, write the. . .

Where something stops



Protection referrals surviving each stage, as a share of the 1,850 cases registered. The largest single loss is between consent and a referral being made, which is a caseworker decision rather than a service failure.

Where something stops — In Python

```
stages = pd.Series({
    "Registered": 1.000, "Consented to refer": 0.885, "Referral made": 0.617,
    "Referral accepted": 0.409, "Service reached": 0.388,
})
losses = -stages.diff().dropna()
print(losses.round(3))
print(f"largest drop: {losses.idxmax()} ({losses.max():.1%})")
```

Where something stops — In R

```
# The drop is the finding; the bars are how you show it.
```

Where something stops

- The largest single drop is 26.8 points, between consent and a referral being made — Not at the service, not at...
- Bars in stage order, never sorted — The sequence is the meaning

Three marks to avoid, and what to use instead

- The pie chart — Angles are read less accurately than lengths, and a pie with more than three slices needs a legend to...
- The dual-axis chart — Two series on two scales, and the crossing point is decided by where you put the axes
- The 3-D anything — Depth adds no data and distorts every length it touches

Three marks to avoid, and what to use instead — In Python

```
# The one line of matplotlib configuration this course insists on.
plt.rcParams.update({"axes.spines.top": False, "axes.spines.right": False})
```

Three marks to avoid, and what to use instead — In R

```
theme_set(theme_minimal(base_size = 11))
```

The stacked bar, which is sometimes right

- Stacked bars are readable for the bottom segment and unreadable above it — Every segment except the first starts at a...
- Use stacking when the total matters and the composition is secondary — a caseload split by category, where the reader's...
- Never stack percentages that do not sum to 100 — The four instruments above overlap: a household can be flagged by...

Report it whole — Example

Food security instruments, 1,955 households, 2024

Four instruments were applied to the same households and are shown as a sorted horizontal bar chart. The bars are prevalences, not components of a total: households can be flagged by more than one instrument, so the bars do not sum and are not stacked.

Poor food consumption	7.4%
Moderate or severe hunger	42.9%
Crisis coping or worse	48.2%
High coping index	51.0%

The chart's claim is the spread, not any single bar. 76.3% of households are flagged by at least one instrument and 3.4% by all four.

Report it whole

- The second paragraph is the one that stops the chart being misread — and it is two sentences

What comes next

- Every bar in this lesson is a proportion estimated from a sample, and not one of them carries an interval.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/data-visualisation/viz-01-the-comparison-decides-the-mark