

Douze communes, un classement, onze intervalles qui se recouvrent

Triées par prévalence, les communes ressemblent à un palmarès. Chacune des onze paires voisines a des intervalles qui se recouvrent, et trois seulement dépassent la médiane du district. Un graphique sans intervalles invite à un classement que les données ne peuvent pas étayer.

Pierre Robentz Cassion

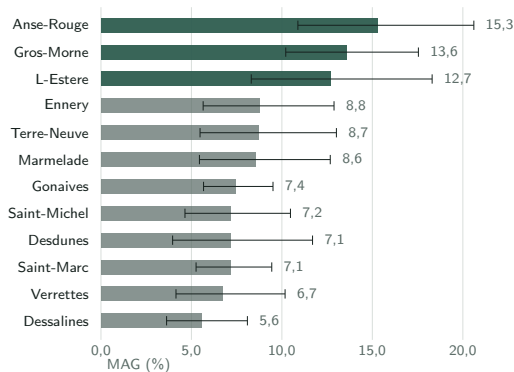
Leçon 2 sur 8

Visualisation de données pour les programmes — La marque et la comparaison

Ce que couvre cette leçon

- Le classement que le graphique invite
- Tracer l'intervalle
- Quand les intervalles ne sont pas le bon encodage
- La ligne de référence travaille plus que les intervalles
- L'ordre de lecture, et les trois qui dépassent
- Rapportez-le en entier
- La suite

Le classement que le graphique invite



Malnutrition aiguë globale par commune avec intervalles de confiance à 95 %. Les trois communes mises en évidence sont les seules dont l'intervalle dépasse la médiane du district.

Le classement que le graphique invite — En 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}")
```

Le classement que le graphique invite — En R

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

Le classement que le graphique invite

- Chaque paire voisine se recouvre — Le graphique peut étayer « ces trois-là sont au-dessus de la médiane du district »...
- L'intervalle est ce qui convertit un classement en constat — Sans lui, le lecteur fournit un ordre que les données ne...

Tracer l'intervalle — En 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()
```

Tracer l'intervalle — En R

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

- Dites ce qu'est l'intervalle — « Intervalle de confiance à 95 % » dans la légende, à chaque fois
- Employez le bon intervalle — Ce sont des intervalles binomiaux exacts de Clopper-Pearson, parce que plusieurs communes. . .
- Ne rognez pas une borne pour qu'elle tienne — Un intervalle qui atteint 20,6 % fixe l'axe ; tronquer l'axe à 15 % et. . .
- Des moustaches par-dessus les barres, non à leur place — Pour une proportion, la barre porte l'ampleur et la moustache. . .

Quand les intervalles ne sont pas le bon encodage

- Un recensement, non un échantillon — Un décompte de tous les points d'eau d'un district n'a pas d'erreur. . .
- Un total, non un taux — « 1 850 cas enregistrés » est un décompte de ce qui s'est produit
- Une série temporelle dont le lecteur doit voir la forme — Douze points mensuels avec moustaches deviennent une haie ;. . .

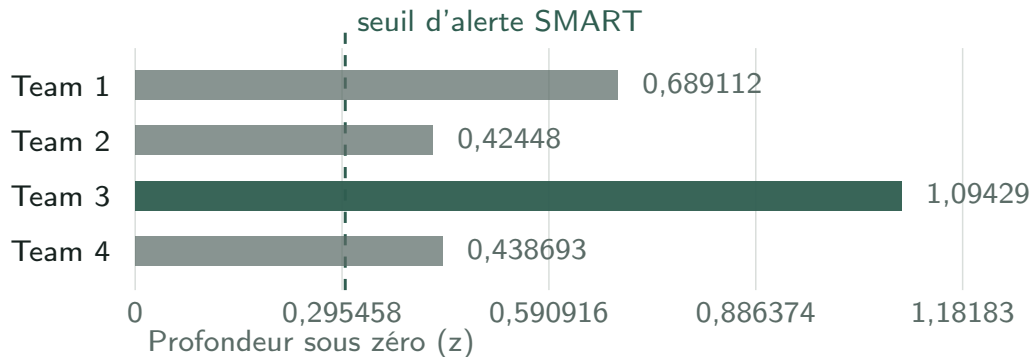
Quand les intervalles ne sont pas le bon encodage — En Python

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

Quand les intervalles ne sont pas le bon encodage — En R

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

La ligne de référence travaille plus que les intervalles



Moyenne du rapport poids-taille par équipe d'enquête, contre le seuil d'alerte SMART. La moyenne d'une équipe dépasse le seuil à partir duquel SMART signale un problème de mesure.

La ligne de référence travaille plus que les intervalles

- Seuils Sphère, seuils d'alerte SMART, frontières de phase IPC, cibles nationales, un passage précédent — Chacun est une. . .
- Étiquetez la ligne par ce qu'elle est, non par sa valeur — « Plancher 70 % » vaut mieux que « 0,70 », que le lecteur. . .
- Une seule ligne de référence — Deux est une comparaison que le lecteur doit effectuer ; trois est une décoration

L'ordre de lecture, et les trois qui dépassent — En Python

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

L'ordre de lecture, et les trois qui dépassent — En R

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

L'ordre de lecture, et les trois qui dépassent

- Anse-Rouge, Gros-Morne et L'Estère sont les trois seules dont la borne inférieure dépasse la médiane du district de 8,0. . .
- La mise en évidence est un argument, faites-en donc l'argument que fait la légende
— Mettre en évidence les trois plus. . .

Rapportez-le en entier — Exemple

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.

- Le troisième paragraphe est ce qui empêche d'employer le graphique comme un palmarès — et c'est la phrase qu'un lecteur. . .

- Les deux figures de cette leçon emploient une couleur pour les barres ordinaires et une pour les barres mises en évidence.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/data-visualisation/
viz-02-uncertainty-on-the-page`