

Le graphique et la phrase viennent du même fichier

Chaque règle de ce cours est une règle dont il faut se souvenir le jour venu. Produire la figure depuis le jeu de données supprime le fait de s'en souvenir — et c'est le seul moyen pour qu'une correction de données déplace le graphique et la phrase ensemble, plutôt que l'un des deux.

Pierre Robentz Cassion

Leçon 8 sur 8

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

Ce que couvre cette leçon

- L'échec que cela prévient
- Ce que fait cette plateforme
- Les quatre propriétés à copier
- Le faire sous matplotlib ou ggplot2
- Ce qu'il faut commiter et ce qu'il faut ignorer
- L'habitude en trois lignes pour un rapport sans chaîne
- Rapportez-le en entier
- La suite

L'échec que cela prévient

- Rien dans aucune compilation ne peut le détecter — La figure est une image valide, le rapport est un document valide,...
- Le remède est structurel plutôt que procédural — Si la figure est produite depuis le jeu de données par un script qui...

Ce que fait cette plateforme — En Python

```
def food_security_instruments() -> Figure:
    survey = read("food-security-survey-2024.v1.csv")
    coping = {row["household_id"]: row for row in read("livelihood-coping-2024.v1.csv")}

    # ... compute the four prevalences from the rows ...

    return Figure(
        slug="food-security-instruments",
        kind="bar",
        title={"en": "Four instruments, one set of households", "fr": "..."},
        caption={"en": "Four food security instruments applied to the same 1,955 "
                        "households. The highest prevalence is seven times the lowest...",
                  "fr": "..."},
        x_label={"en": "Households flagged (%)", "fr": "Ménages signalés (%)"},
        bars=bars,
    )
```

Ce que fait cette plateforme — En R

The same shape in R: a function returning data plus its labels, rendered once.

Ce que fait cette plateforme

- Aucun nombre n'est saisi — Les 7,4 % de la légende ne sont pas écrits dans la légende — la légende énonce la relation. . .
- Deux sorties depuis une description — SVG pour le web, TikZ pour le document PDF et le diaporama Beamer
- Bibliothèque standard uniquement — Les générateurs n'ont besoin que d'un Python d'origine et d'aucune installation, si. . .

Les quatre propriétés à copier

- Une description, plusieurs rendus — Le graphique est des données plus des étiquettes, et le moteur de rendu est séparé
- Bilingue par construction — Titre, légende et étiquette d'axe sont des dictionnaires indexés par langue, si bien qu'une. . .
- Sortie commitée — Les fichiers SVG et TikZ sont dans le dépôt, si bien que le site se construit sans Python sur le. . .
- Un test qui échoue quand ils divergent — `figures.test.ts` vérifie que chaque figure référencée par une leçon existe, . . .

Les quatre propriétés à copier — En Python

```
# The check that matters most, in one line:  
assert (OUT / f"{figure.slug}.{locale}.svg").exists()
```

Les quatre propriétés à copier — En R

A test is what turns a convention into a constraint.

Le faire sous matplotlib ou ggplot2 — En Python (suite)

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

FIGURES = pathlib.Path("outputs/figures")

def gam_by_commune() -> None:
    screening = pd.read_csv("data/muac-screening-artibonite-2024.v1.csv")
    summary = (screening.assign(case=screening["muac_mm"] < 125)
               .groupby("commune")["case"].agg(["mean", "size"])
               .sort_values("mean"))

    fig, ax = plt.subplots(figsize=(6.5, 4))
    ax.barh(summary.index, summary["mean"])
    ax.set_xlabel("GAM (%)")
    fig.savefig(FIGURES / "gam-by-commune.svg", bbox_inches="tight")
```

Le faire sous matplotlib ou ggplot2 — En Python (suite)

```
plt.close(fig)

summary.to_csv(FIGURES / "gam-by-commune.csv")      # the numbers, beside the chart

if __name__ == "__main__":
    FIGURES.mkdir(parents=True, exist_ok=True)
    gam_by_commune()
```

Le faire sous matplotlib ou ggplot2 — En R

```
gam_by_commune <- function() {  
  summary <- screening |>  
  mutate(case = muac_mm < 125) |>  
  summarise(gam = mean(case), n = n(), .by = commune)  
  
  ggsave("outputs/figures/gam-by-commune.svg", width = 6.5, height = 4)  
  write_csv(summary, "outputs/figures/gam-by-commune.csv")  
}
```

Le faire sous matplotlib ou ggplot2

- Écrivez les nombres à côté du graphique — Un CSV auprès de chaque figure est ce qui permet à un relecteur de vérifier. . .
- Une fonction par figure, un script qui les lance toutes — Non un carnet où le graphique est la cellule 34 et dépend de. . .
- Enregistrez en SVG, non en PNG — partout où la destination le permet : il reste net quand quelqu'un le redimensionne, . . .

Ce qu'il faut commiter et ce qu'il faut ignorer

Commiter	Ignorer
Le script générateur	Les aperçus PNG exportés
La sortie des figures, si la compilation ne peut pas exécuter Python	Les points de reprise de carnets
Le CSV des nombres auprès de chaque figure	Tout ce qui contient un horodatage

Ce qu'il faut commiter et ce qu'il faut ignorer

- Sortie déterministe ou rien — Une figure qui incorpore la date de production change à chaque exécution, et le diff. . .

L'habitude en trois lignes pour un rapport sans chaîne

- Un script qui produit chaque figure du rapport — lancé avant l'envoi du rapport, en une seule commande
- Aucun graphique fait à la main dans un tableur — parce que c'est celui qui ne sera pas régénéré
- La légende produite avec la figure — ou au moins les nombres qu'elle contient lus depuis le même objet que celui d'où. . .

L'habitude en trois lignes pour un rapport sans chaîne — En Python

```
caption = (f"Global acute malnutrition by commune, {summary['size'].sum():,} children "  
          f"screened. {(summary['mean'] > 0.10).sum()} of {len(summary)} communes "  
          f"are above the 10% emergency threshold.")
```

L'habitude en trois lignes pour un rapport sans chaîne — En R

If the caption is an f-string, it cannot disagree with the chart.

L'habitude en trois lignes pour un rapport sans chaîne

- Cette seule ligne représente l'essentiel du bénéfice — Une légende calculée depuis les données est une légende qui ne. . .

Rapportez-le en entier — Exemple

Figure production for this report

All figures are generated from the committed dataset files by scripts/figures.py, run as part of the report build. No figure is drawn by hand and no number in a caption is typed.

Each figure writes its underlying values to a CSV of the same name, so any figure can be checked without rerunning the analysis.

Figures are produced in English and French from one description, so a figure cannot exist in only one language.

Output is deterministic: rerunning the script on unchanged data produces byte-identical files, so a diff shows only real changes.

- Le dernier paragraphe est ce qui rend le reste vérifiable — Une chaîne dont la sortie bouge à chaque exécution ne peut. . .

- C'est le cours.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/data-visualisation/
viz-08-generate-the-figure`