

Aide mémoire data viz

Points généraux

- Tout le monde est subjectif. Même un data scientist. Ne vous cachez pas derrière une pseudo objectivité. Vos visualisations doivent porter un message ! Et ce message doit être le **titre** de vos visualisations. Si vous n'arrivez pas à trouver un message alors votre visualisation sert sûrement à rien !
- Par contre vous devez être **honnête** ! Ne faites pas dire à votre graphique ce qu'il ne dit pas en trichant sur les représentations. Les journalistes et les communicants font déjà ça très bien 😊

Points techniques

- %>% (prononcé pipe) : prendre la sortie d'une fonction pour la mettre dans l'entrée d'une autre. Se place sur la même ligne que le membre de gauche

```
x %>% f() <=> f(x)
```

- un graphique ggplot2 doit commencer par ggplot(...). Par contre seul, ggplot ne produit rien
- aes(): permet de préciser les variables qui vont être utilisées dans la production de graphique et quels paramètres (position, forme, couleur, etc.) vont dépendre de quelles variables.
- geom_XXX() : couche graphique. On ajoute une couche au ggplot avec un +

Exemple

```
NUTS2_year %>%  
filter(nom %in% c('Champagne-Ardenne', 'Picardie')) %>%  
ggplot(aes(x=année, y=population)) +  
geom_line(aes(group=nom))
```

- Il est possible de redéfinir les données dans une couche graphique avec l'attribut data

Exemple

```
geom_text(  
  data = . %>% filter(population>5000000 | superficie>50000),  
  aes(label=id_anc)  
)
```

- Pour décaler des textes utiliser nudge_XXX = value. La valeur est dans les unités du graphique (en gros si votre échelle de graph est grand, il faudra une grosse valeur de

nudge)

```
geom_text(  
  aes(label=id_anc),  
  nudge_y=400000  
)
```

- labs : ajouter un titre, sous titre, source au diagramme

Exemple

```
labs(  
y = 'Population', # les titres des axes peuvent être spécifiés ici...  
title = "Istanbul (Turquie) surpeuplée, Oulu (Finlande) isolée",  
subtitle = "Population et superficie en Europe (2015)",  
caption = "Source: Eurostat (niveau NUTS 2)"  
)
```

- theme_XXXX : pour changer de thème
- réordonner des données :

```
mutate(  
# x_min = reorder(id_anc, population, min),  
# x_max = reorder(id_anc, population, function(x) max(x)),  
id_anc = reorder(id_anc, population, function(x) last(x)-first(x))  
)
```

- ajouter de la couleur

```
geom_XXX(aes(color=...))
```