

# Visualisation de données avec R — retour TP1

*Rémi Pépin (sur base de Arthur Katosky)*

*Janvier 2019*

## Contents

1. Faire un graphique avec <code>ggplot2</code> . . . . .	2
---	---

### La visualisation de données, une représentation *visuelle*

La visualisation repose sur la vision car c’est le sens le plus adapté comme support de mémoire externe:

- parallélisation inconsciente
- perception simultanée
- zone cérébrales développée
- possibilité technique

### Une représentation *efficace*

La visualisation se distingue du design, de la publicité ou de l’art par son intérêt pour l’efficacité de la représentation à communiquer une information ; nous ne poserons donc pas la question de savoir si un graphique est “joli” mais de s’il “fonctionne”.

Malheureusement définir l’efficacité c’est compliqué car elle dépend du but à atteindre. Il n’y a donc pas une solution magique. Cela demande de la réflexion et de la remise en question.

Dans ce cours, nous nous intéresserons surtout à la tâche de comparer des grandeurs dans des graphiques finis, c’est-à-dire publiables. D’un côté parce que la comparaison est une tâche fondamentale (elle conditionne, par exemple, la découverte de régularités). De l’autre parce que c’est l’une de plus étudiées.

### Une représentation *subjective*

Rien n’est objectif dans ce monde, ne vous pensez pas au dessus de tout le monde. Utilisez votre subjectivité pour produire une représentation qui “claque” au lieu de la rendre lisse.

### Les données

Nous travaillerons avec des données Eurostat, principalement des données de population au niveau NUTS 2 (le découpage officiel statistique européen, qui possède 3 niveaux). Le tableau de données `NUTS2_year` possède une ligne par région NUTS 2 et par année d’observation:

```
## # A tibble: 9,019 x 15
##   id_anc année superficie comments population_femm~ population_homm~
##   <chr>  <int>      <dbl> <chr>                <int>      <int>
## 1 AT11   2015      3669 <NA>                147246     141110
## 2 AT12   2015     18917 <NA>                832975     803803
## 3 AT13   2015       395 <NA>                929704     867633
## 4 AT21   2015      9360 <NA>                286371     271270
## 5 AT22   2015     16251 <NA>                621265     600305
## 6 AT31   2015     11717 <NA>                727840     709411
## 7 AT32   2015      7050 <NA>                276378     262197
## 8 AT33   2015     12514 <NA>                370936     357890
## 9 AT34   2015      2534 <NA>                191814     186778
## 10 BE10  2015       161 <NA>                605416     578685
```

```
## # ... with 9,009 more rows, and 9 more variables: population <int>,
## #   population_0_19 <int>, population_20_59 <int>,
## #   population_60_plus <int>, id <chr>, nom_anc <chr>, nom <chr>,
## #   chgt <fct>, anc <list>
```

## 1. Faire un graphique avec ggplot2

### 1.1 Graphique basique

Commençons par un exemple: l'évolution de la population en Champagne-Ardenne et en Picardie entre 1990 et 2015.

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

