

Séance 7. La visualisation graphique des données

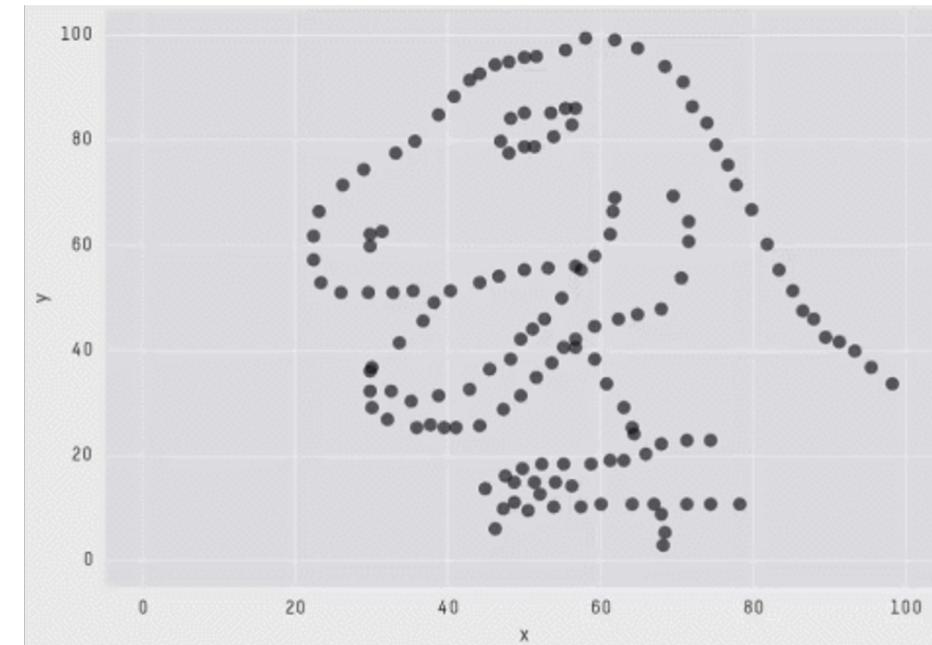
Objectifs de la séance

- Introduire les principes de représentation graphique des données
- Choisir son graphique en fonction du type de données à représenter
- Savoir éviter les pièges pour une représentation fidèle aux données

Pourquoi : exploration et analyse des données

(1/2)

- **Ne pas croire aveuglément les tendances centrales et de dispersion !**
- **Visualiser vos données**



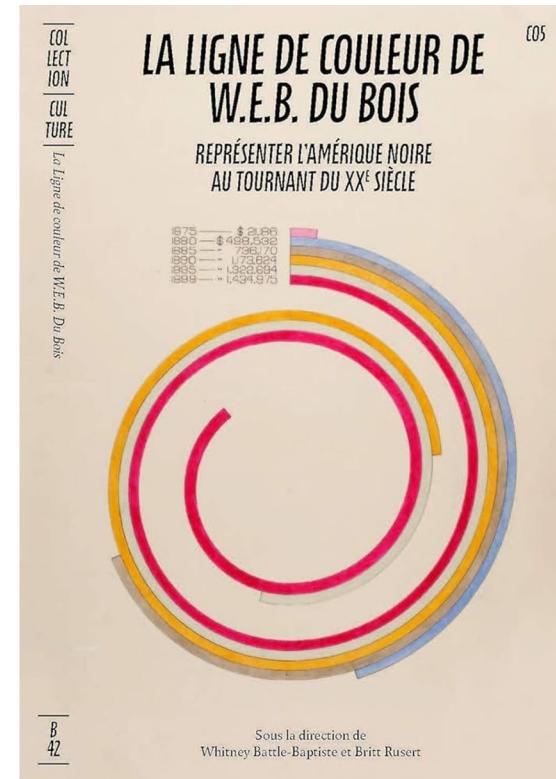
Source : [Datasaurus](#)

Pourquoi : exploration et analyse des données (2/2)

- Diffuser des résultats
- Les rendre compréhensibles
- Mettre en valeur et transmettre l'information

Dataviz everywhere

- Pas nouveau : W.E.B Du Bois, [La ligne de couleur](#)
- Mais se diffuse/démocratise :
 - production plus simple avec numérique (crayon vs. ordi)
 - et plus récemment diffusion en ligne, viralité
 - #dataviz (voir par exemple le site de [Cédric Scherer](#)), etc.



Exemples scientifiques

Source :
 GARGIULO F., FONTAINE S.,
 DUBOIS M. et TUBARO P., 2023. «
 A meso-scale cartography of the
 AI ecosystem », Quantitative
 Science Studies : 1-17.

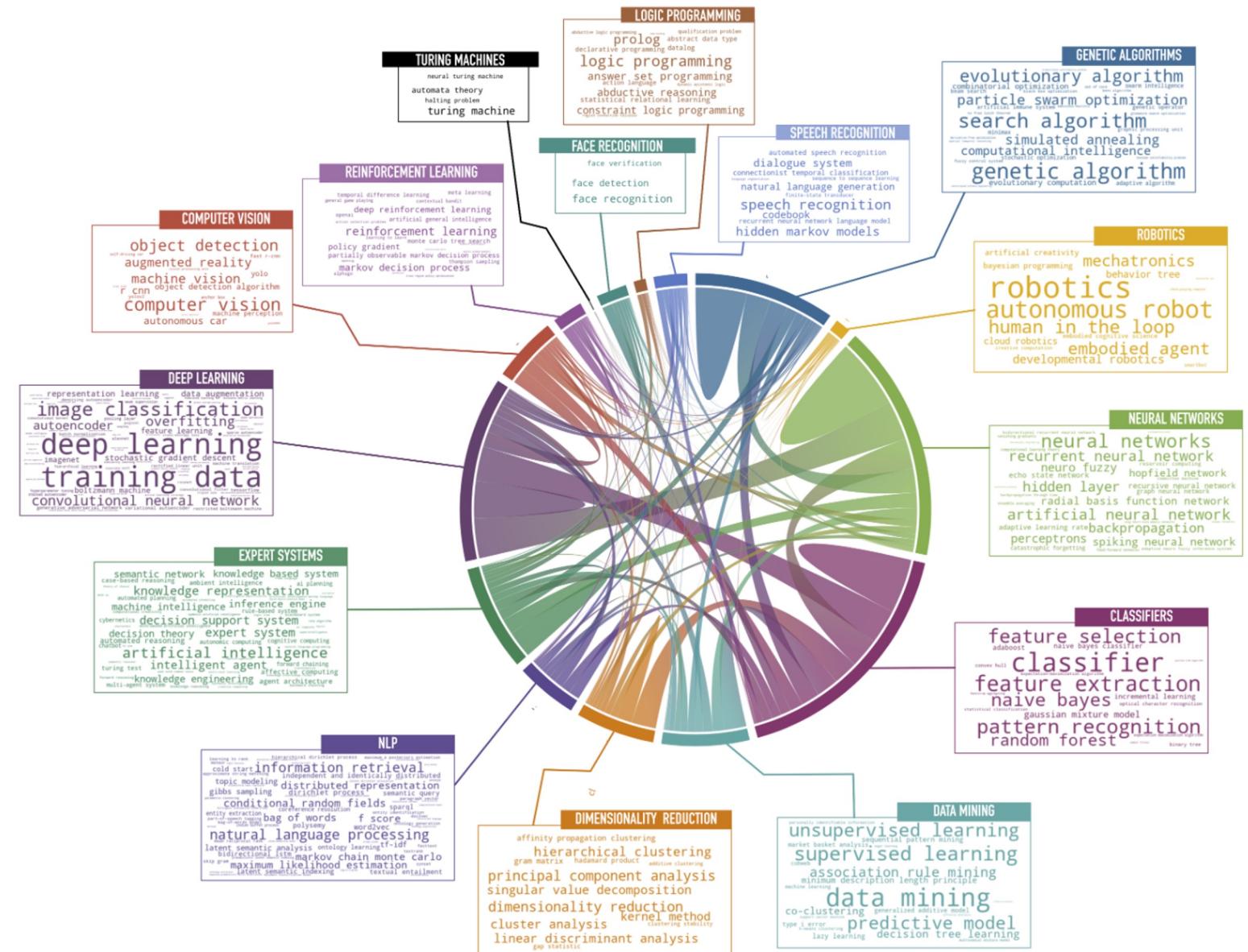


Figure 3. AI specialities and their relationships.

Exemples scientifiques

Source :
 GARGIULO F., FONTAINE S.,
 DUBOIS M. et TUBARO P., 2023. «
 A meso-scale cartography of the
 AI ecosystem », Quantitative
 Science Studies : 1-17.

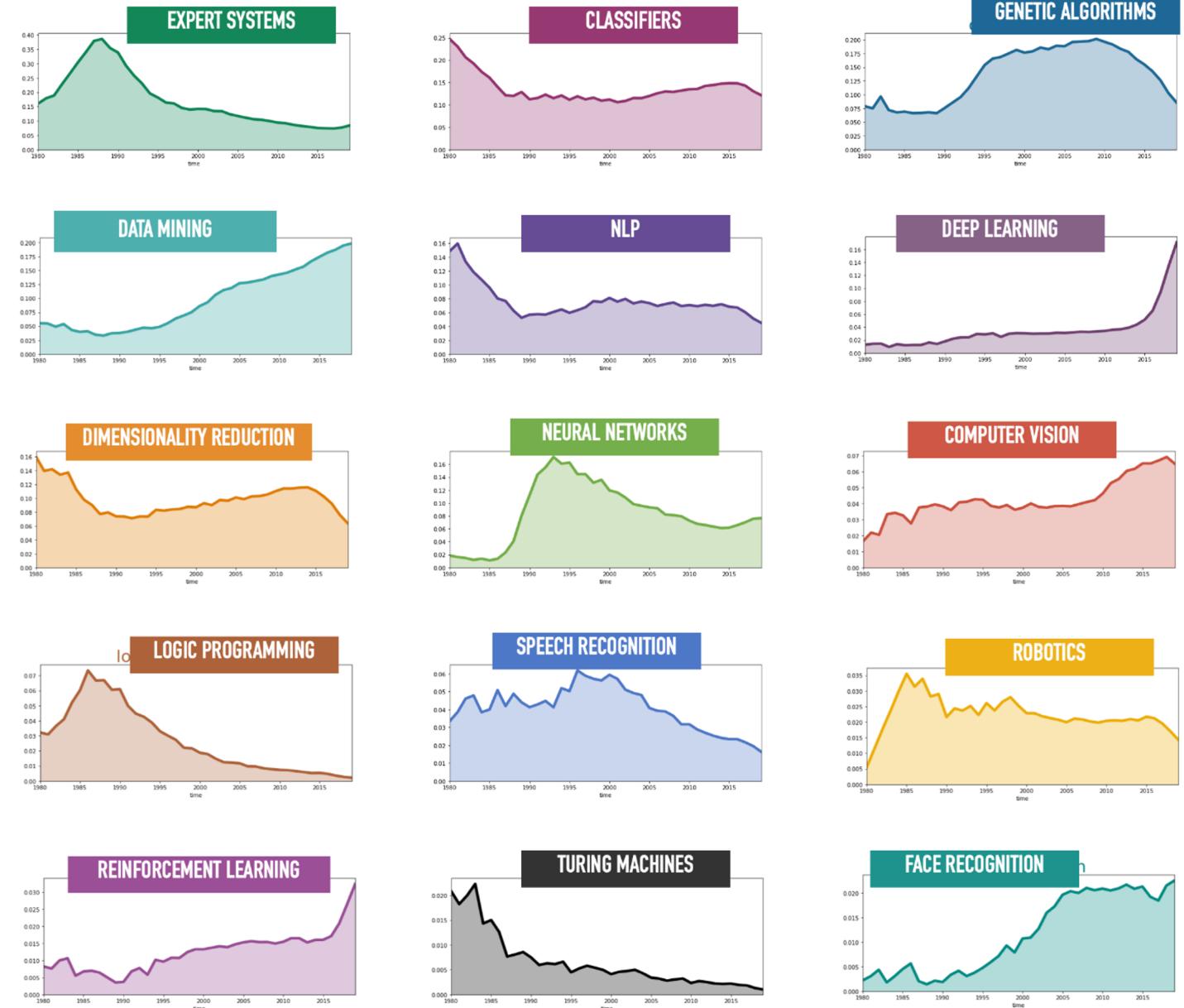
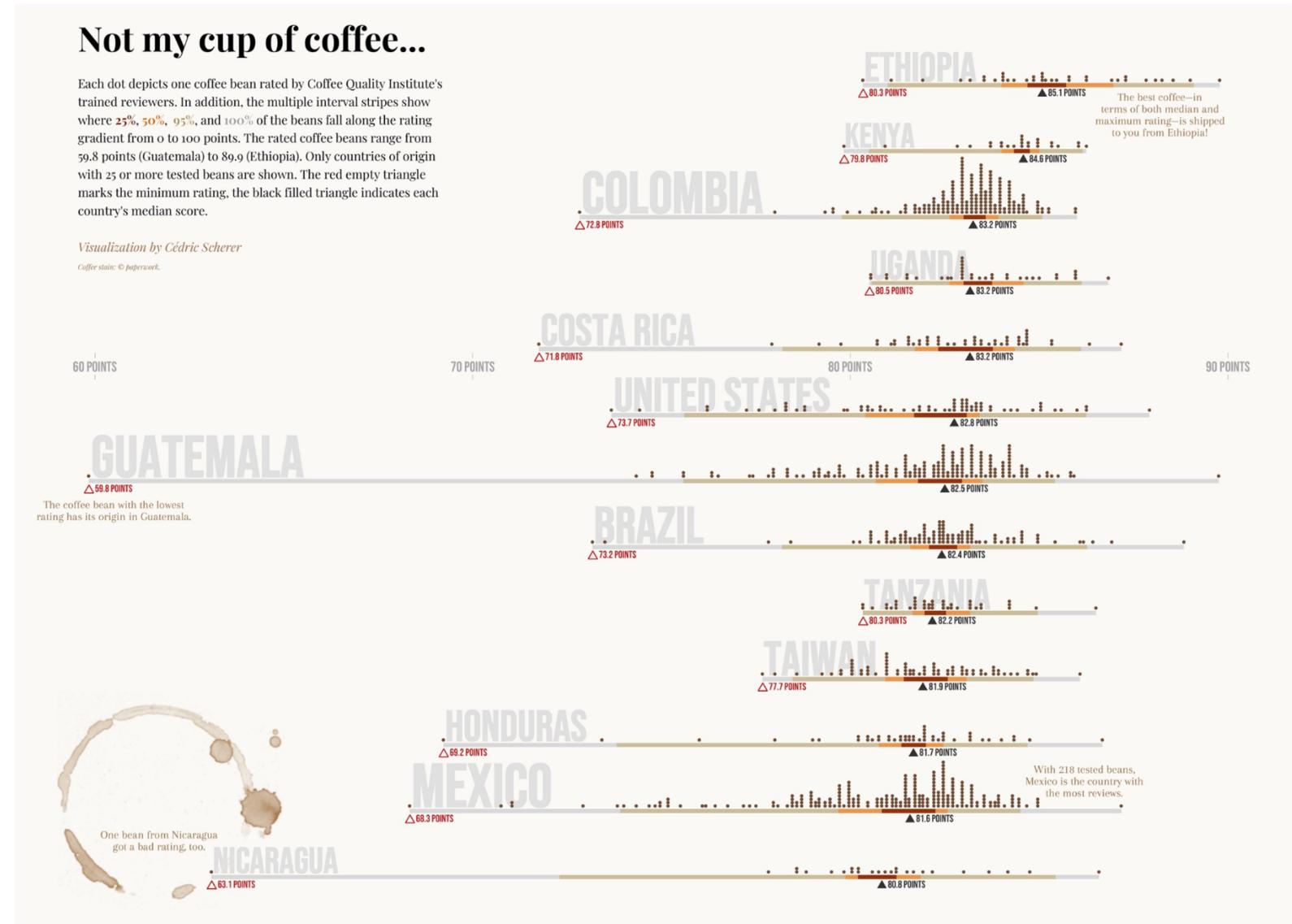


Figure 4. The timeline of AI specialities.

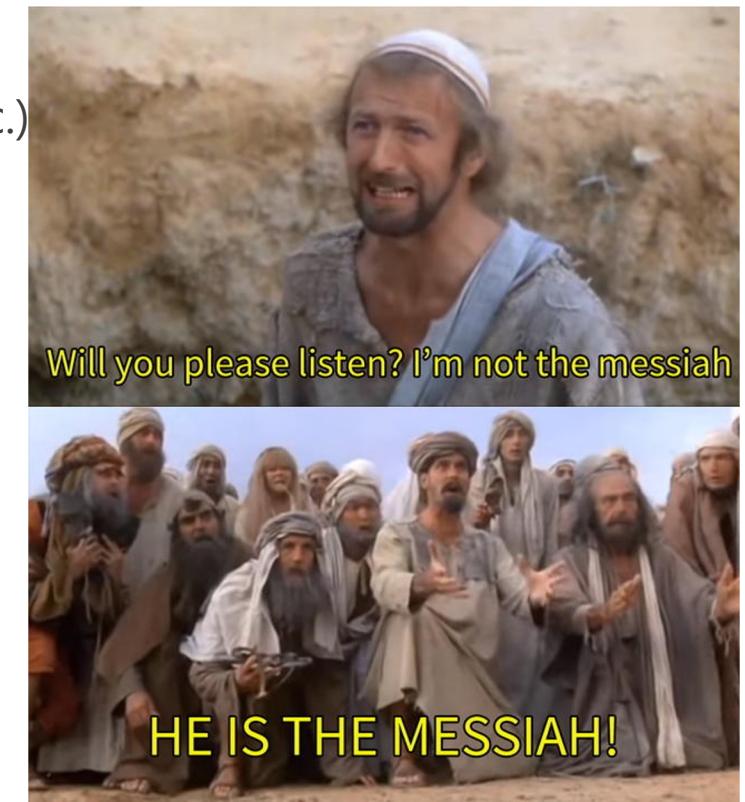
Exemple de Dataviz

Source :
<https://www.behance.net/gallery/100638417/Not-My-Cup-of-Coffee>



Quelle représentation graphique ?

- **Choix d'une visualisation graphique va dépendre**
 - du type de variable : catégorielle vs numérique (+ cartes, réseaux, etc.)
 - de que je veux montrer = qu'est-ce que je raconte/mets en exergue
- **Comment choisir ?**
 - des conventions
 - l'expérience
- **Des ressources multiples disponibles en ligne**
 - « la bible » <https://www.data-to-viz.com/>
 - le site de [Yan Holtz](#)





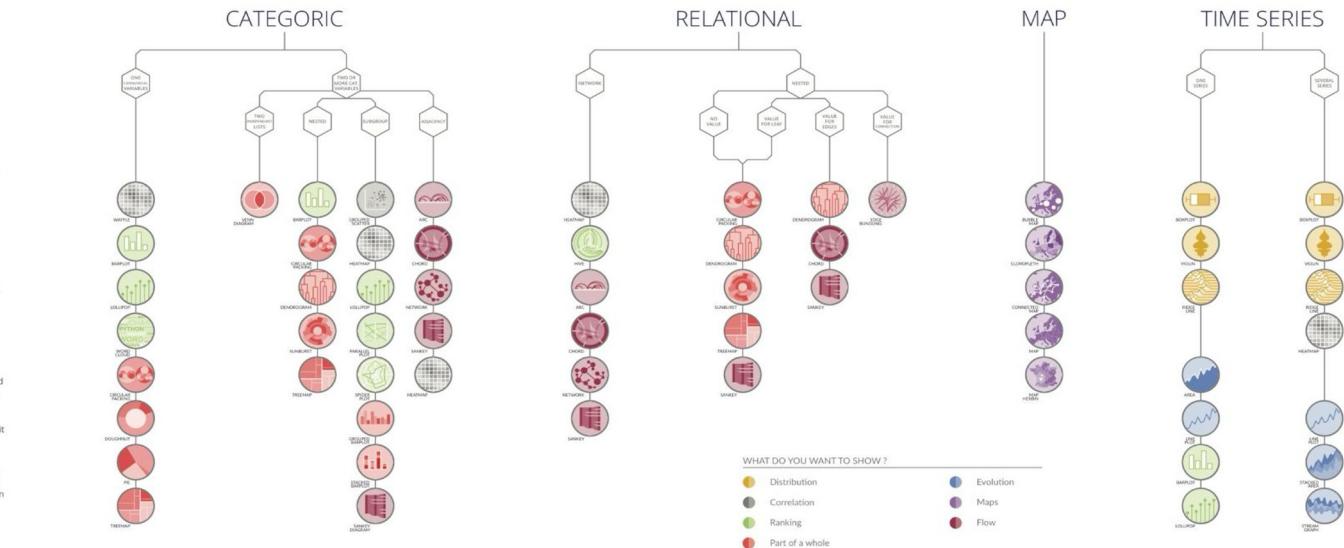
from Data
to Viz

'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- ① Identify what type of data you have.
- ② Go to the corresponding decision tree and follow it down to a set of possible charts.
- ③ Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

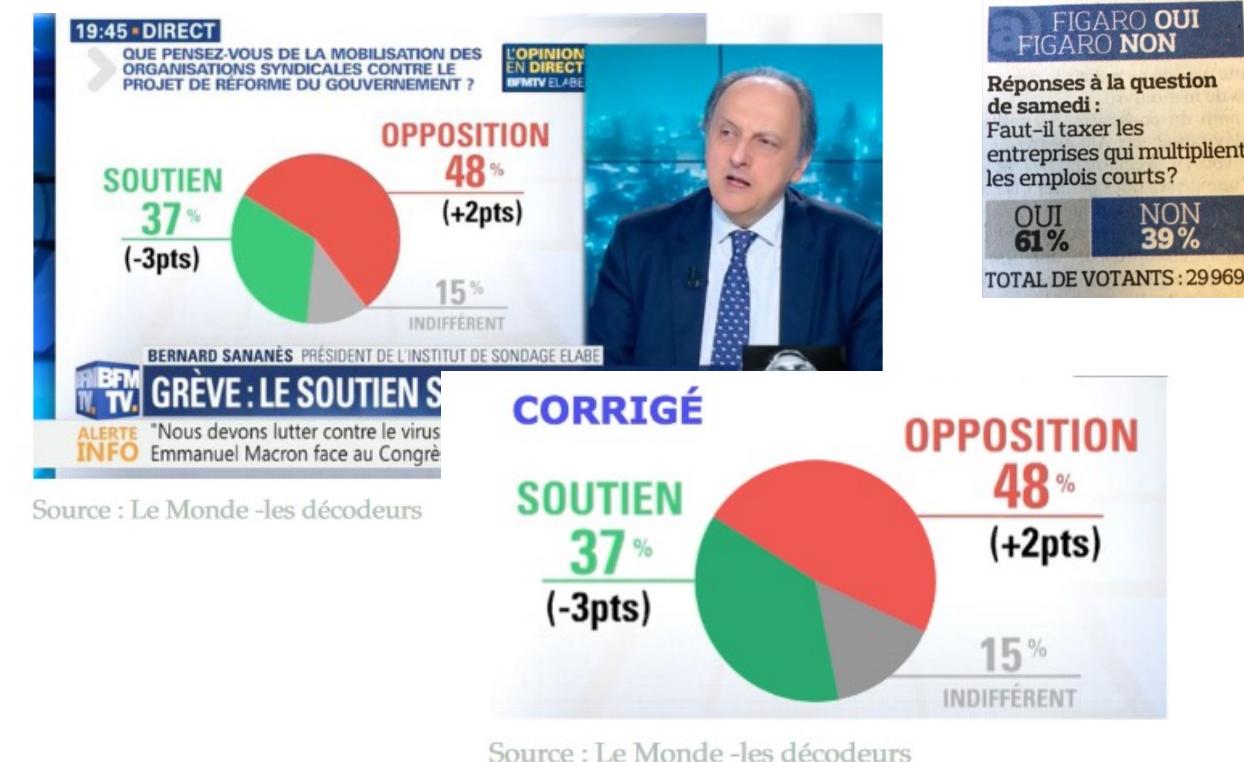


2018 © Yan Holtz & Céline Heudel for www.dataviz.com

Séance 9. La visualisation graphique des données

Des bonnes pratiques à respecter

- Les représentations graphiques peuvent transmettre des informations de manière orientée, de manière involontaire ou volontaire
- Différencier
 - la communication politique ou de certains médias : les chiffres et graphiques peuvent être présentés de manière orientée, pour soutenir un discours ou des idées
 - la communication scientifique : les données font l'objet d'une représentation graphique neutre et mathématiquement juste (en principe !)



Source : « Comment mentir avec des graphiques – Les dessous de nos médias », de Christophe Bontemps
<http://data.visualisation.free.fr/Blog/Mentir6.pdf> (Site internet : <http://data.visualisation.free.fr/>)

Représenter les variables catégorielles

- Diagramme en barres (*bar plot*)
- Camembert, diagramme en secteurs (*pie chart*) / Diagramme donut (*doughnut*)
 - une prudence pour l'utilisation de ce type de graphique (cf. plus loin dans le cours)

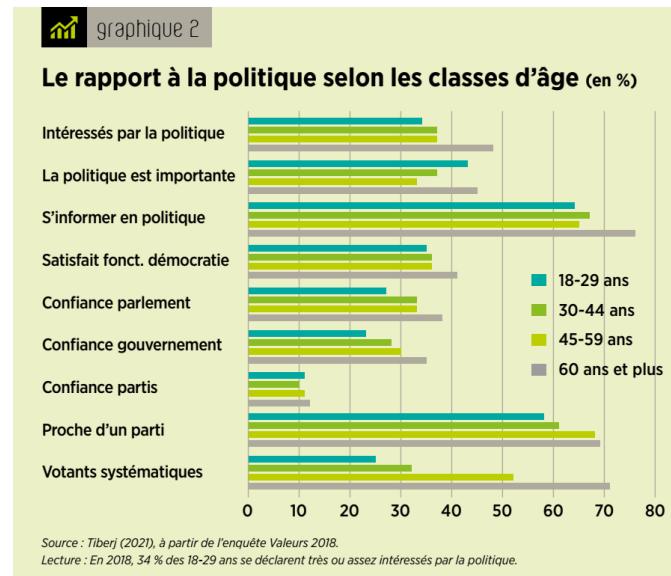
Diagrammes en barres

- Pour représenter **une** variable

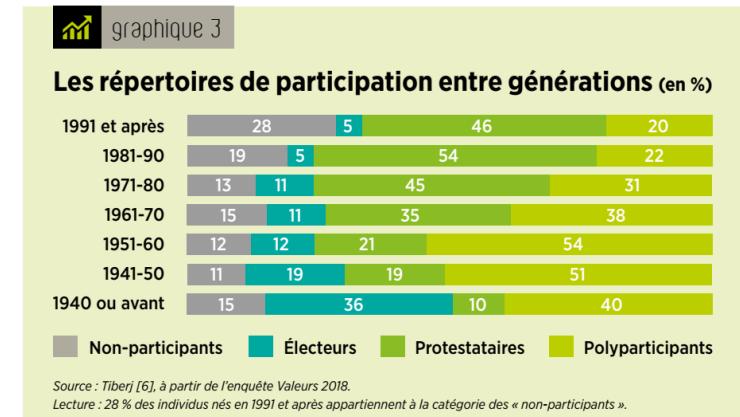


- Pour représenter **plusieurs** variables

barres juxtaposées

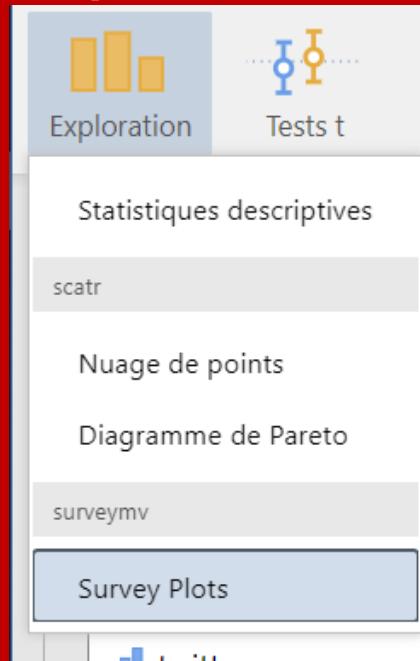


barres superposées

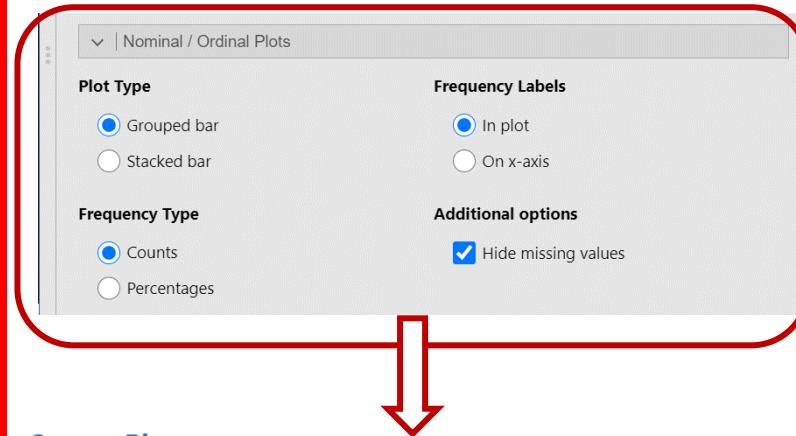
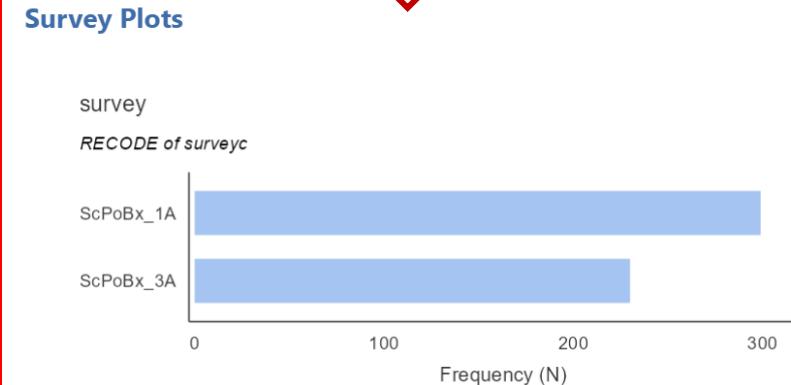
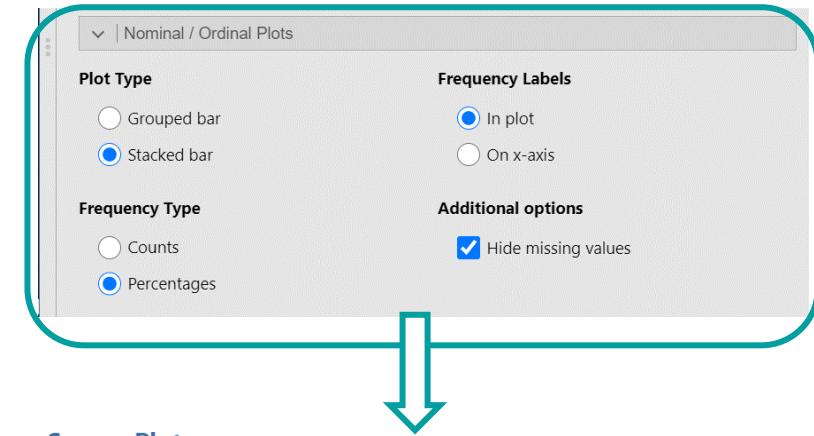
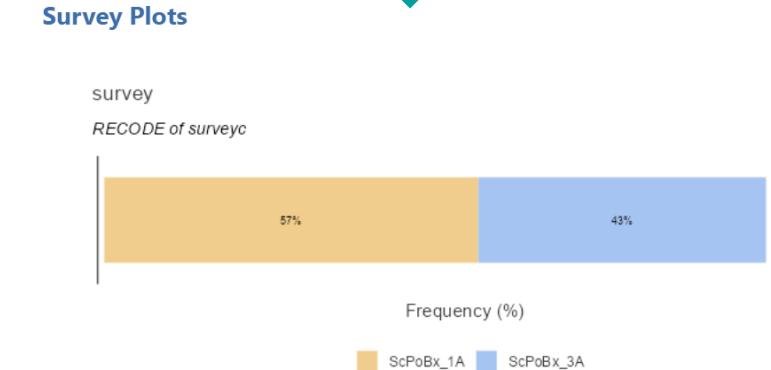


Sous Jamovi

Avec le module **surveymv**
 (menu
Analyses/Exploration)

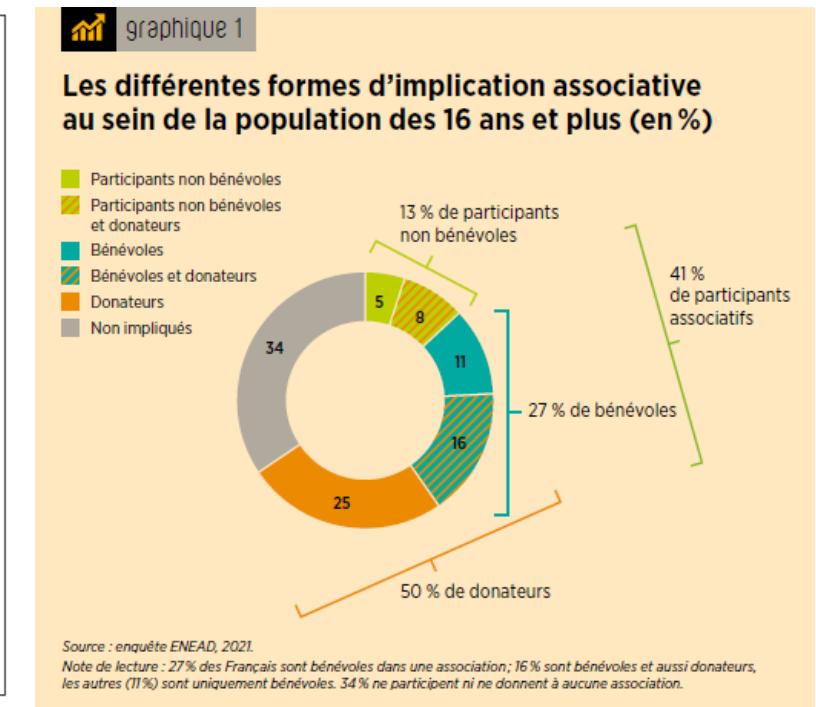
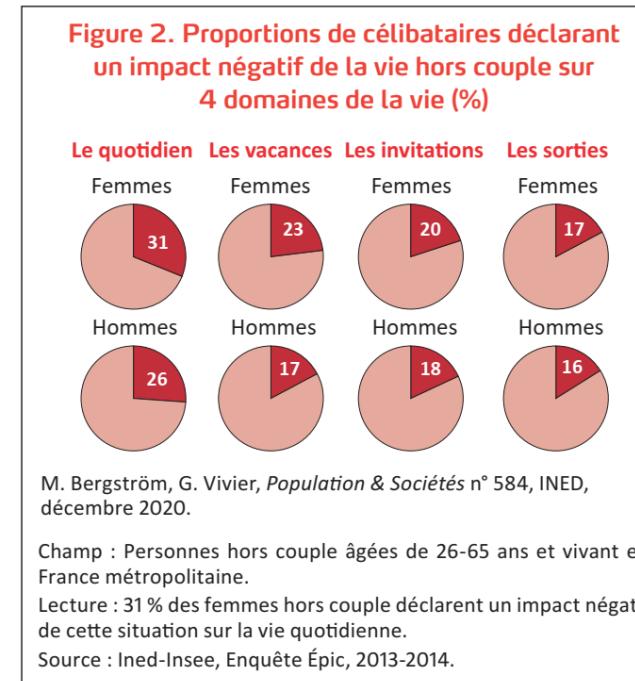


Paramétrage pour créer des diagrammes **simples** ou **empilés**, avec les **effectifs** ou les **pourcentages**, indiqués sur les axes ou les barres

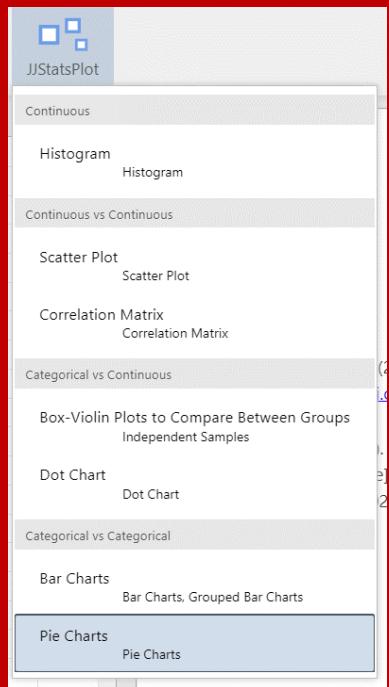
Diagrammes en secteurs

- Pour représenter une variable **dichotomique** (2 modalités)
- Pour représenter une variable **polytomique** (plusieurs modalités) **NON ORDONNÉE**



Sous Jamovi

Avec le module **JJStatsPlots**
(menu Analyses)



Pie Charts

Dependent Variable: instagra

Grouping Variable: (Optional)

Split By (Optional)

random_id4
survey
tps_rs
nbinsta
twitter
nbtwit
snapchat
nbsnap

Data

Exclude Missing (NA)

Analysis

Plot

Add GGStatsPlot Layer

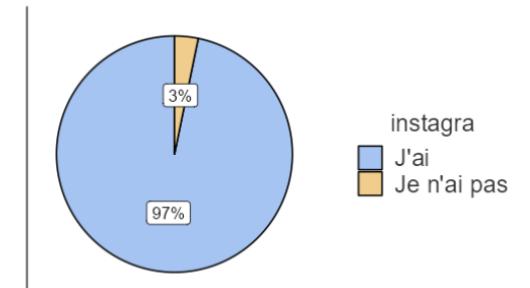
Résultats

Pie Charts

You have selected to use Pie Charts.

instagra

$\chi^2_{\text{gof}}(1) = 460.20, p = 4.35e-102, \hat{C}_{\text{Pearson}} = 0.68, \text{CI}_{\text{CI}} = [0.58, 0.78]$



$\log_e(\text{BF}_{01}) = -286.18, \alpha_{\text{Gungel-Dickey}} = 1.00$

Représenter les variables quantitatives

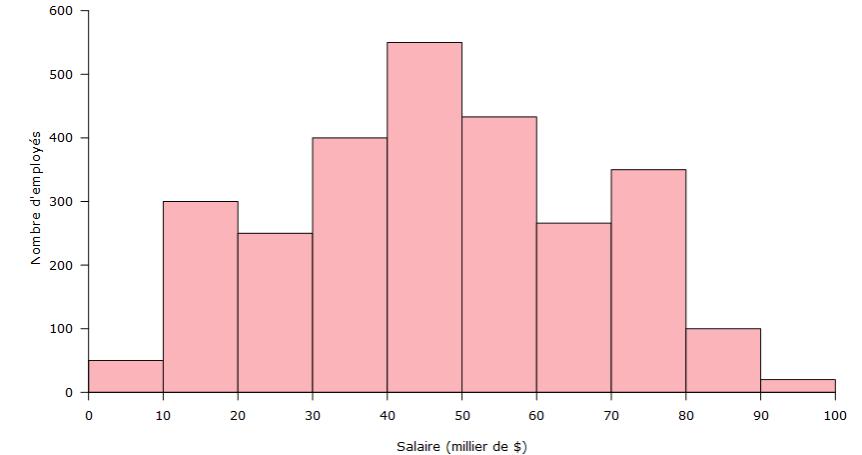
- Histogramme (*histogram*) avec ou sans courbe de densité
- Courbe (*line plot*)
- Nuage de points (*scatterplot*)
- Boîte à moustaches (*boxplot*) et alternatives (*violin plot*, etc.)

Histogrammes

- Histogramme « simple »

Source : Statistique Canada (2021)

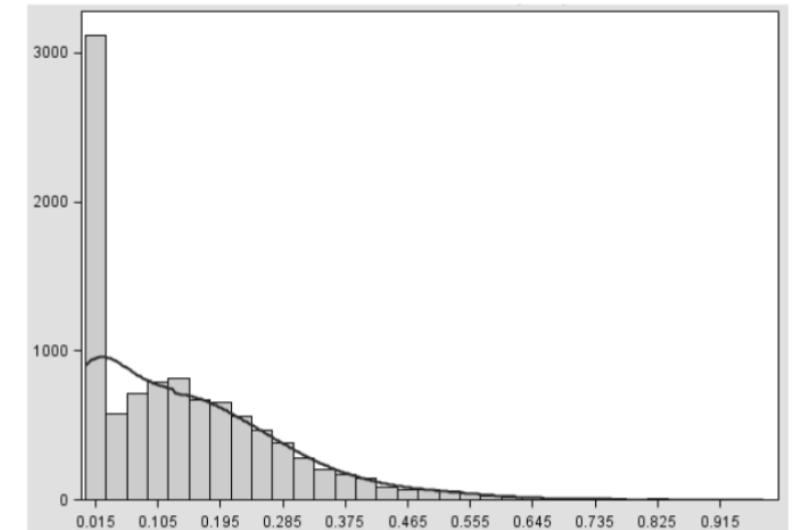
Graphique 5.7.1
Distribution des salaires des employés de la société ABC



- Histogramme avec courbe de densité

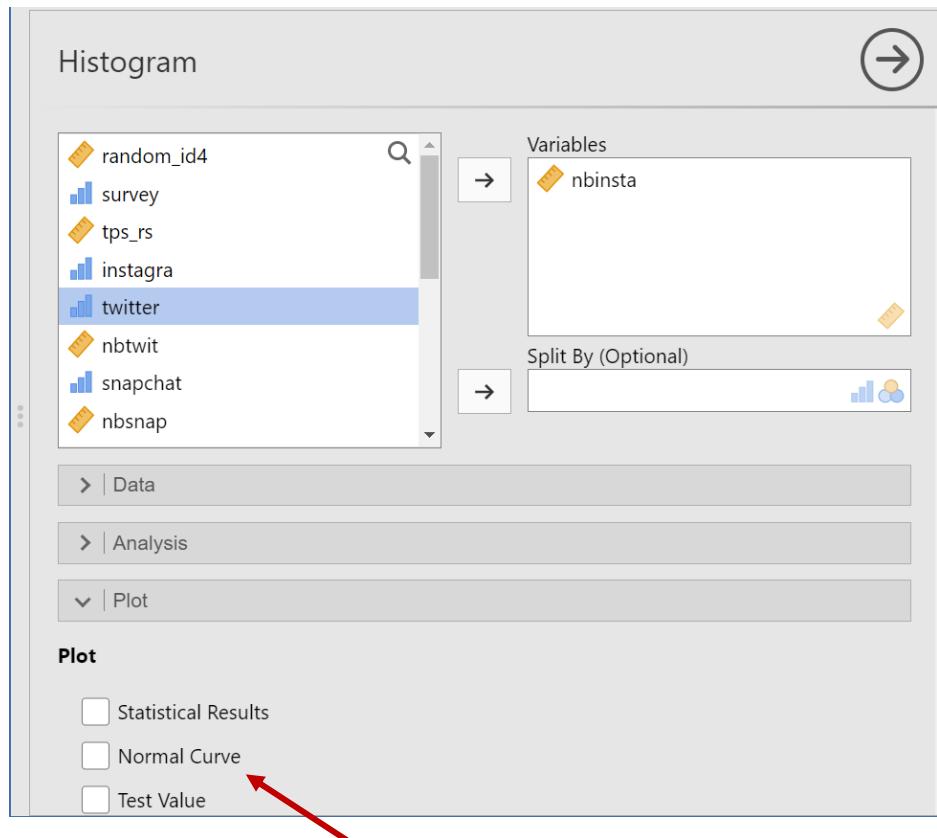
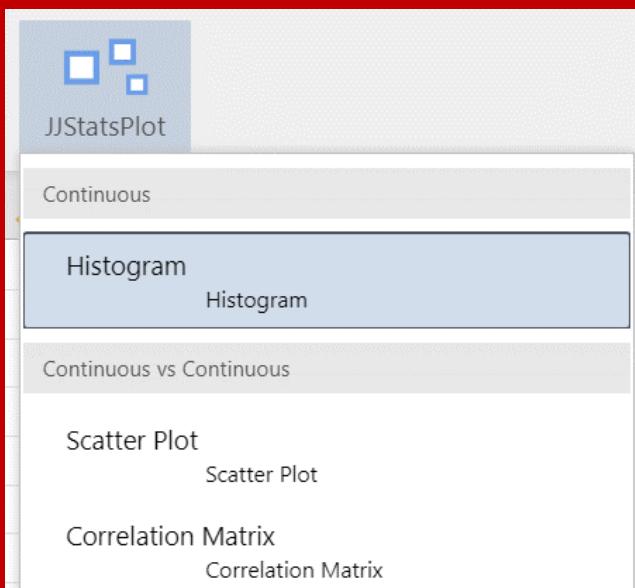
Source : Bouzidi, Btissam, et al. (2010).
Le pantouflage des énarques : une première analyse statistique »,
Revue française d'économie, vol. xxv, no. 3

Figure 9
Distribution du pourcentage de temps passé dans le secteur privé au cours de la carrière



Sous Jamovi

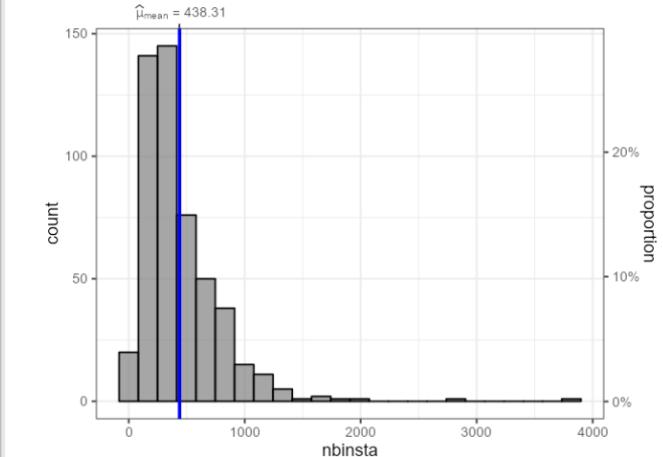
Avec le module **JJStatsPlots**
(menu Analyses)



Histogram

You have selected to make a histogram.

Histogram

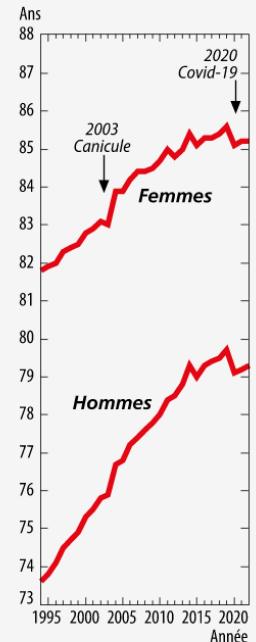


Possibilité d'ajouter la courbe de densité sur le graphique

Courbes

- Évolution au cours du temps

Figure 2. Évolution de l'espérance de vie à la naissance en France depuis 1994

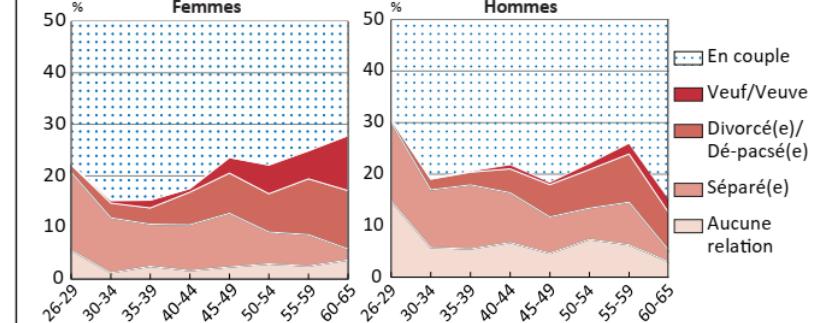


Gilles Pison, *Population & Sociétés*, n° 609, mars 2023, INED.

Source : Insee [1, 2].

- Aires sous la courbe

Figure 1. Situation vis-à-vis du célibat selon l'âge et le sexe (%)



M. Bergström, G. Vivier, *Population & Sociétés* n° 584, INED, décembre 2020.

Champ : Personnes âgées de 26-65 ans et vivant en France métropolitaine.

Lecture : Entre 26 et 29 ans, 22 % des femmes ne sont pas en couple ou engagées dans une relation amoureuse importante ; 7 % sont séparées d'une relation cohabitante.

Source : Ined-Insee, Enquête Épic, 2013-2014.

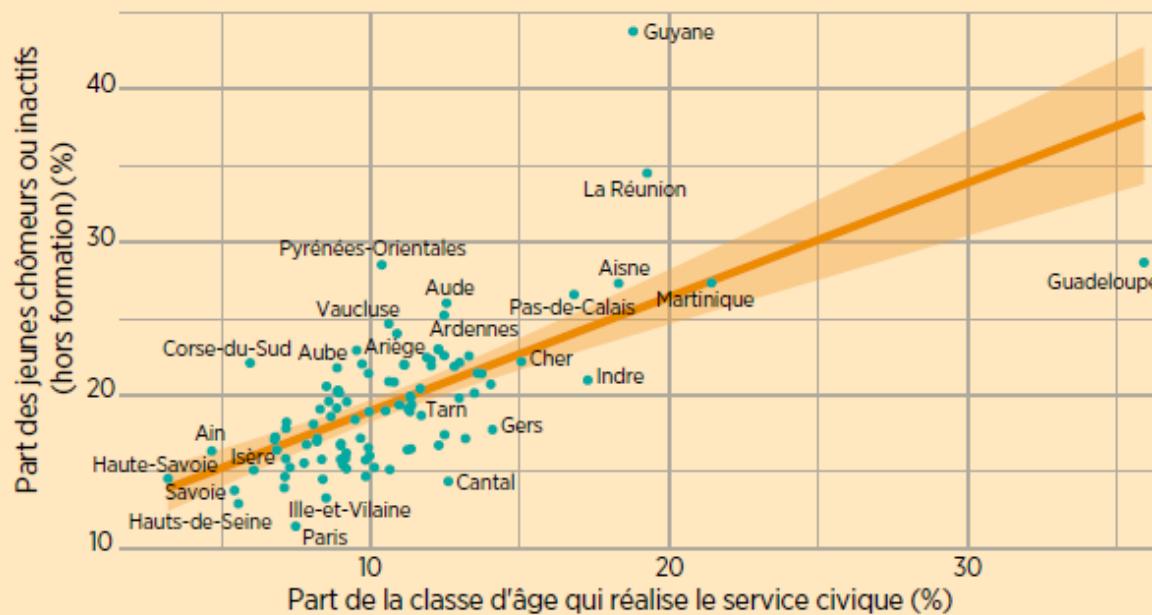
Nuage de points



figure 4

Probabilité de réalisation du service civique selon la part de jeunes chômeurs

16-25 ANS CHÔMEURS OU INACTIFS (HORS FORMATION), AU NIVEAU DÉPARTEMENTAL



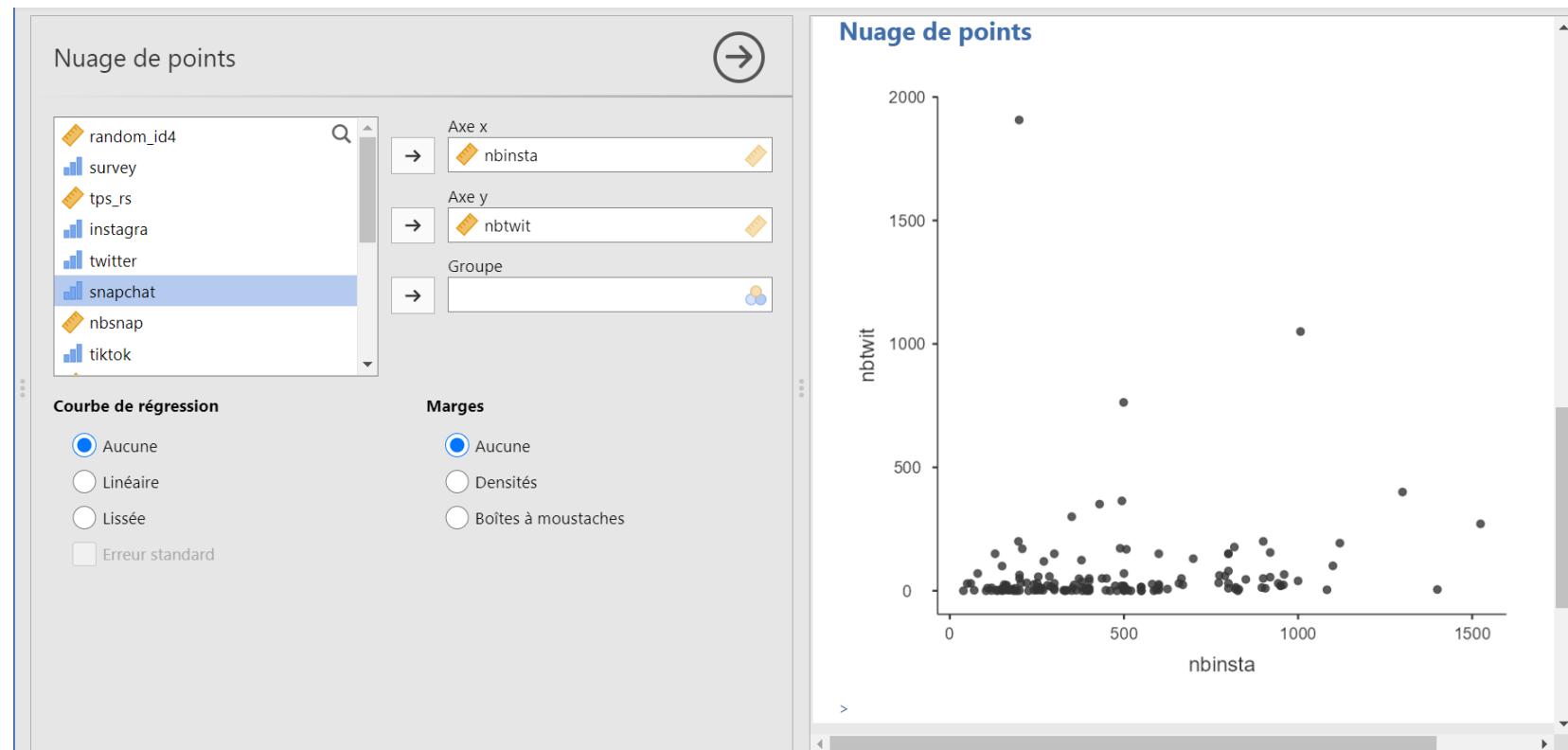
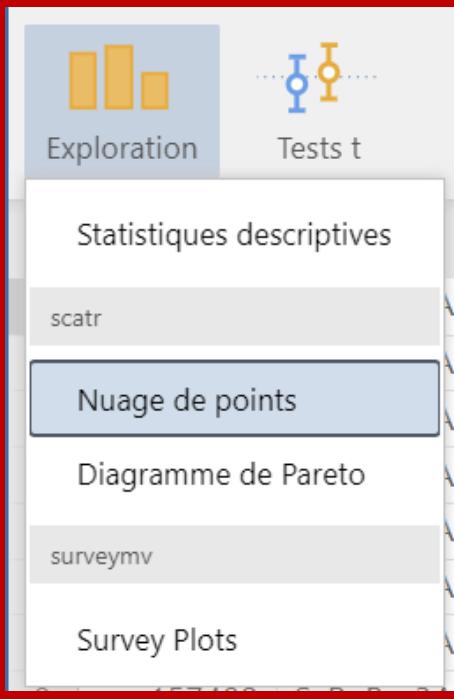
Source : ASP-ASC et recensement de population 2019, traitement INJEP, MEDES, RP INSEE 2019.

Champ : jeunes de 16 à 25 ans, France métropolitaine et DROM.

Note de lecture : la Haute-Savoie, l'Ain, la Savoie et les Hauts-de-Seine se caractérisent par des parts des 16-25 ans chômeurs ou inactifs (hors formation) et des indices de réalisation du service civique faibles. À l'inverse, l'Aisne, la Martinique, la Réunion, la Guyane et la Guadeloupe ont à la fois de forts taux de jeunes sans emploi et des ICSC élevés. La ligne orange représente la droite de régression linéaire et la zone orange clair, son intervalle de confiance à 95%.

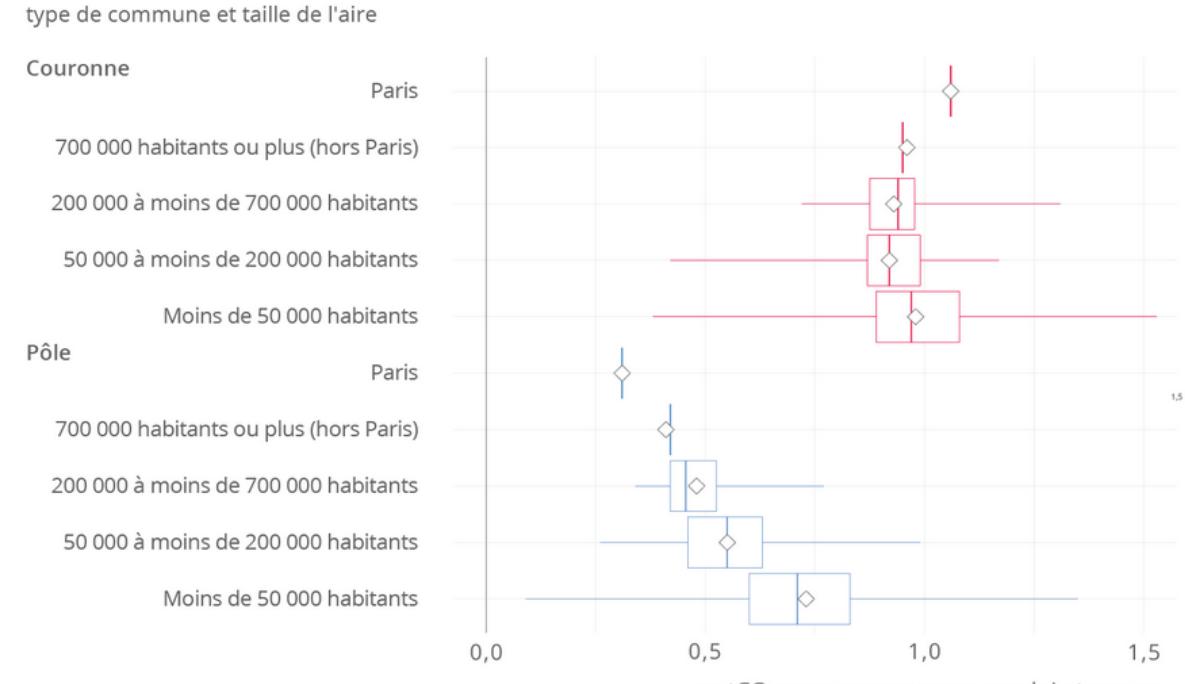
Sous Jamovi

Avec le module **scatr** (menu Analyses/Exploration)



Boxplot

Figure 3 – Dispersion des émissions de GES des trajets domicile-travail entre aires d'attraction des villes en 2019



Notes : La dispersion des émissions de gaz à effet de serre (GES) des trajets domicile-travail au sein de chaque zone (taille de l'aire croisée avec pôle/couronne) est représentée par une boîte à moustaches. Les différents segments des boîtes distinguent les quantiles à 25 %, 50 % (médiane) et 75 %. Les extrémités (moustaches) représentent le minimum et le maximum. La moyenne est représentée par un losange. Le nombre d'aires de 700 000 habitants ou plus hors Paris étant faible, seule la moyenne et la médiane sont indiquées dans ce cas ; /// : absence de résultat due à la nature des choses.

Lecture : En 2019, les émissions annuelles moyennes de GES par actif des trajets domicile-travail dépassent 0,71 tCO₂e dans la moitié des pôles des aires d'attraction des villes de moins de 50 000 habitants.

Champ : France métropolitaine, personnes en emploi de 15 ans ou plus qui se déplacent pour aller travailler et dont les distances domicile-travail sont inférieures à 10 km pour les piétons, 30 km pour les cyclistes et 100 km pour les autres modes de transport.

Sources : SDES-Insee, enquête Mobilité des personnes 2018-2019 ; Insee, recensement de la population 2019, exploitation complémentaire ; distancier Metric-OSRM, © les contributeurs d'OpenStreetMap et du projet OSRM.

Sous Jamovi

Via le menu **Analyses / Exploration / Statistiques descriptives**

Statistiques descriptives 

nbtwit
snapchat
nbsnap
tiktok
nbtik
instap

Séparer par 

Statistiques descriptives Variables sur plusieurs colonnes Tableau des fréquences 

▶ | Statistiques

▼ | Graphes

Histogrammes

Histogramme
 Densité

Graphes Q-Q

Q-Q

Boîtes à moustaches

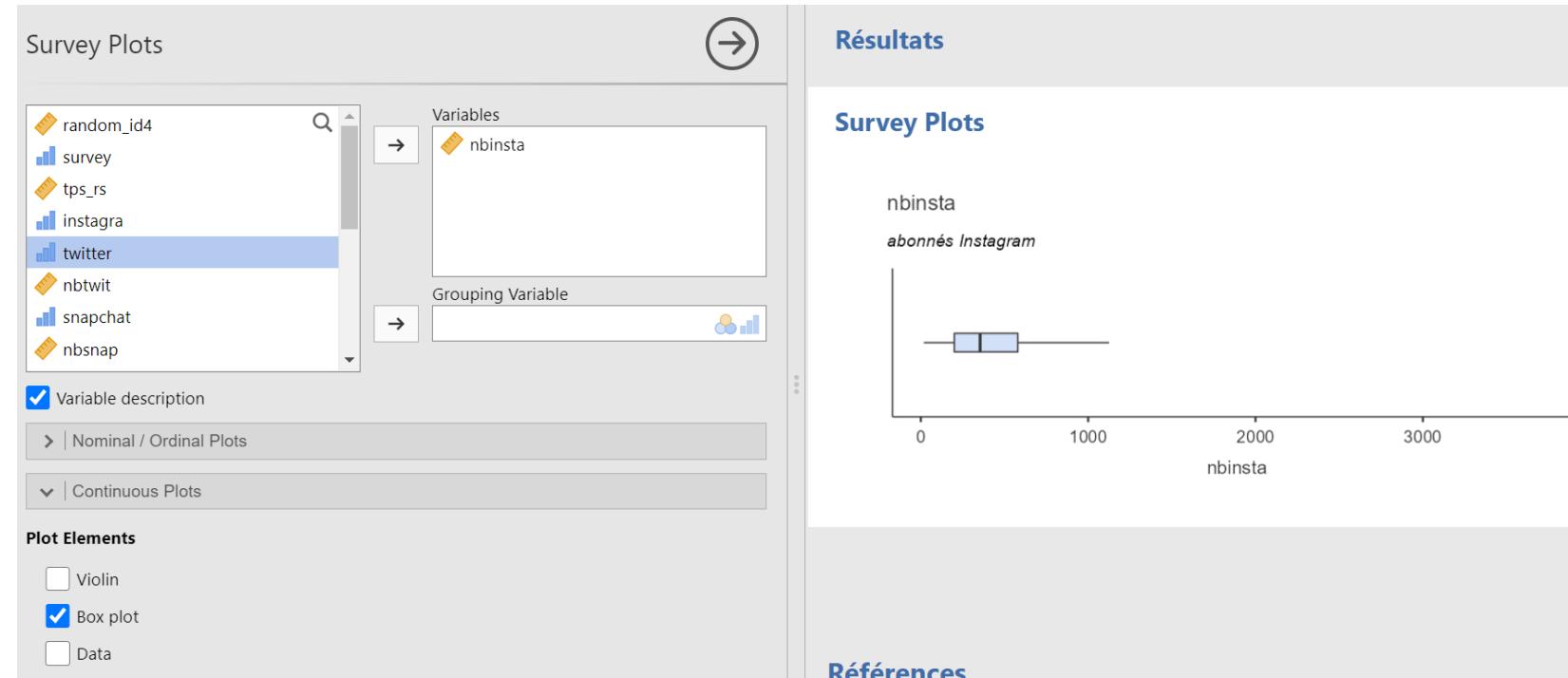
Boîte à moustaches
 Étiqueter les valeurs aberrantes
 Violon
 Données
Ecartées ▾
 Moyenne

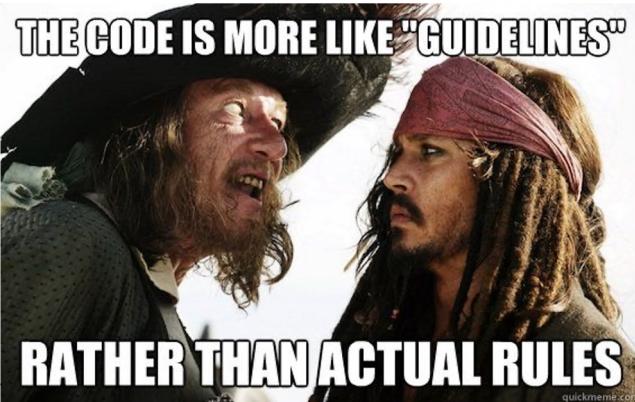
Diagrammes en barres

Diagramme en barres

Sous Jamovi

Avec le module **surveymv**
(menu
Analyses/Exploration)





“Règles” et pièges (1/2)

- Des **conventions** comme grandes lignes **directrices**... mais des **pièges / erreurs à éviter**
 - voir la page dédiée [Caveats](#) sur le site From data to viz
- **Quelques règles**
 - éviter de couper les axes des diagrammes/histogrammes
 - proportional ink
 - éviter les camemberts
 - réfléchir aux couleurs (cf. [Lisa Charlotte Muth](#))
 - conventions culturelles : blanc vs. noir / rouge vs. bleu (politique, température)
 - intelligible ?
 - enjeu d'accessibilité : [Colblindor](#)
 - choix de palettes couleurs : <https://flatuicolors.com/> ; <https://coolors.co/>

“Règles” et pièges (2/2)

Ne pas confondre diagramme en barres et histogramme !

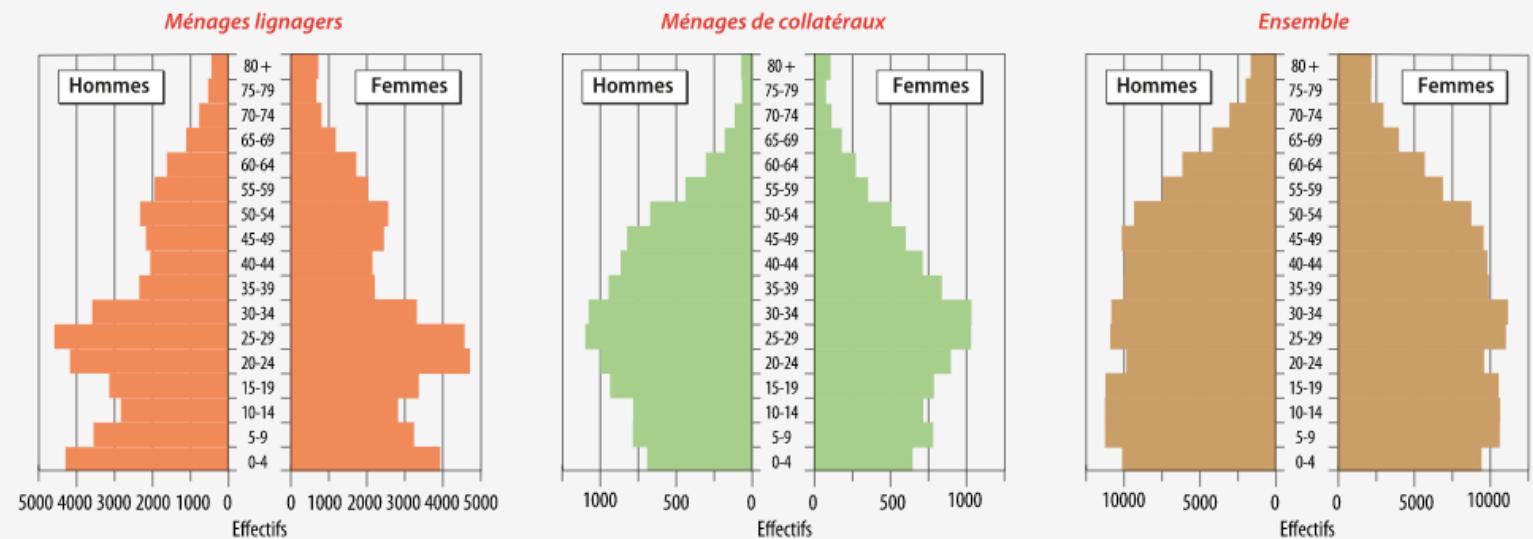
	Diagramme en barres	Histogramme
Utilisation	Pour comparer différentes catégories	Pour afficher la distribution d'une variable
Type de variable	Variables catégoriques	Variables numériques
Apparence	La fréquence de chaque catégorie est illustrée par une barre distincte	L'étendue des valeurs est divisée en une série d'intervalles qui ne se chevauchent pas. Les points de données sont regroupés et le nombre de points dans chaque intervalle correspond à une barre distincte
Espace entre les barres	Il peut y avoir de l'espace entre les barres	Il n'y a pas d'espace entre les barres
Réorganisation des barres	L'ordre peut être modifié pour les variables nominales	Impossible de modifier l'ordre des intervalles

Source : Statistique Canada (2021) <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/histo/5214822-fra.htm#shr-pg0>

Exemple

Une représentation particulière : la pyramide des âges

Figure 1. Pyramides des âges des personnes en ménage lignager, de collatéraux et de l'ensemble



Leïla Fardeau, Éva Lelièvre pour l'équipe Atolls, *Population & Sociétés*, n° 616, novembre 2023, Ined.

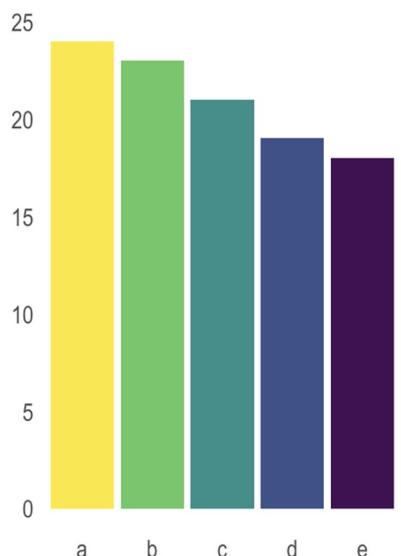
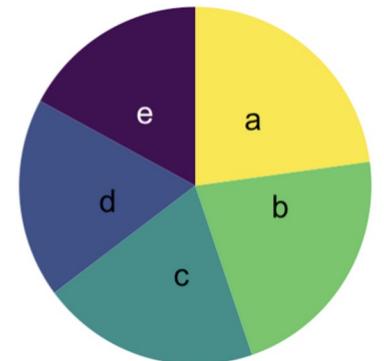
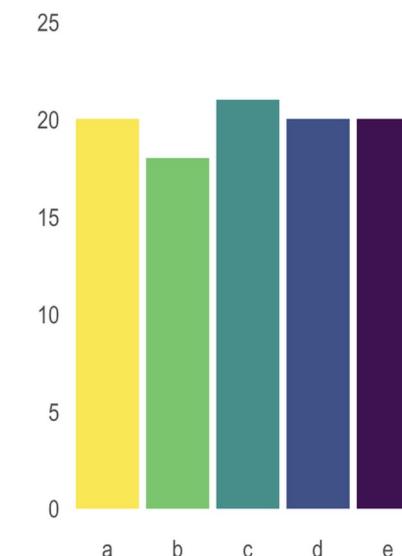
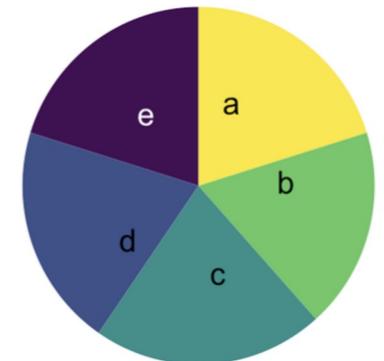
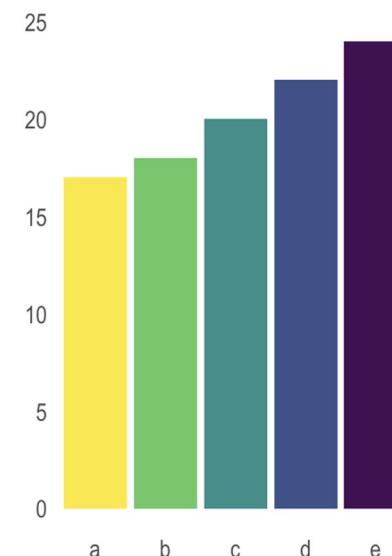
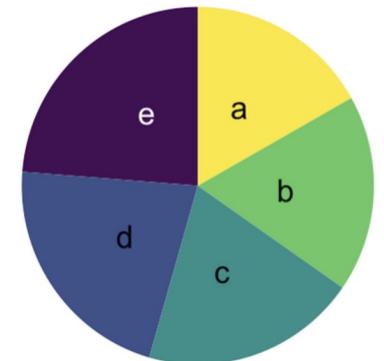
Champ : Personnes résidant dans des logements ordinaires.

Source : ISPF – Recensement 2017 de la Polynésie française.

Exemple

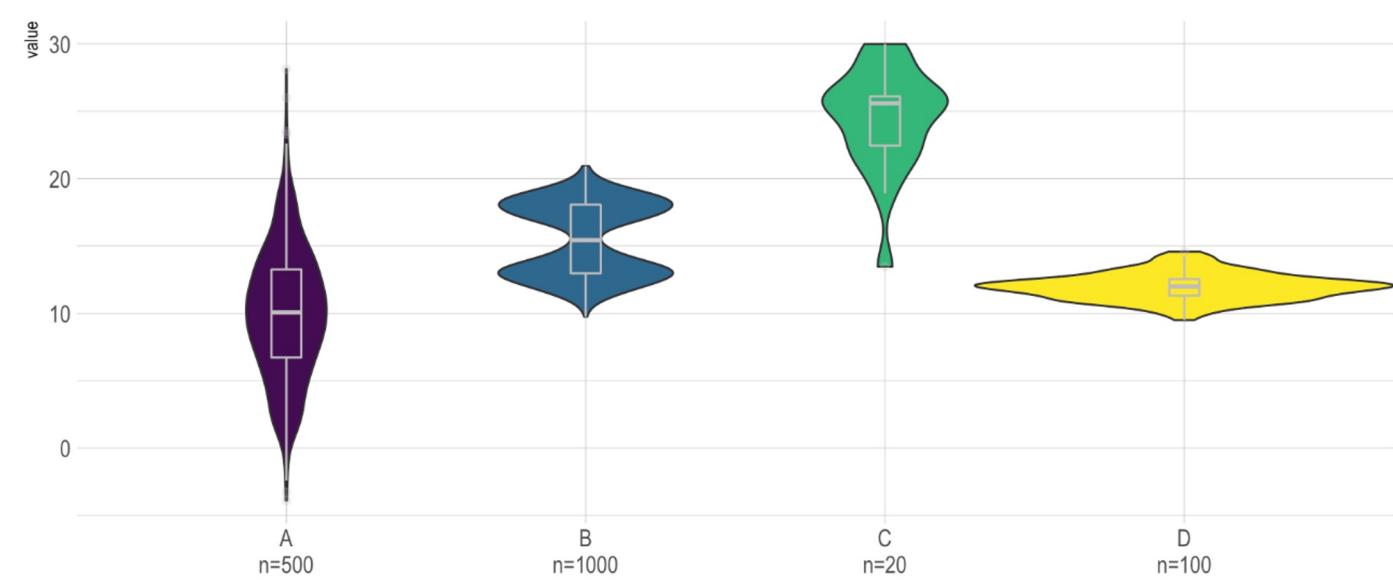
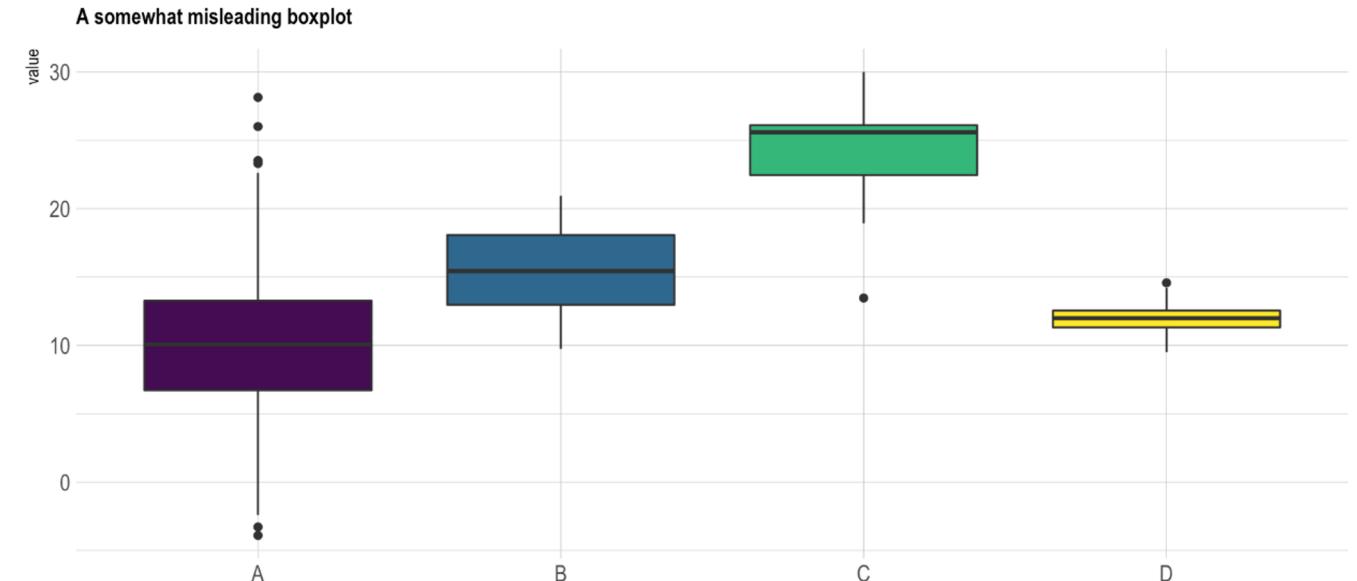
Problème avec les pie charts

Source : [data to viz](#)



Exemple

Le risque de masquer/rater des infos (box plot)
Source : [data to viz](#)



Ressources

- **Informations, guides, etc.**
 - [From Data to viz](#)
 - [Datavizcatalogue](#)
 - Le [blog](#) de datawrapper
- **Outils (ici Jamovi suffit, mais pour aperçu)**
 - <https://www.rawgraphs.io/>
 - <https://flourish.studio/>
 - <https://www.datawrapper.de/>
 - Python ([Plotly](#), [Seaborn](#), [Matplotlib](#)) / R ([ggplot2](#))

Bibliographie sélective

- Chanvril-Ligneel F and Le Hay V (2014) Méthodes statistiques pour les sciences sociales. Paris: Ellipses. [Chapitre 8]
- Py B (2013) La Statistique sans Formule Mathématique. 3e édition. Statistiques et probabilité. Pearson. [Chapitre 3]
- Benveniste C (2009) Épisode 3. Des graphiques et des tableaux. *Hors-Série. Numéro Spécial : Savoir Compter, Savoir Conter*

Application des notions abordées à son jeu de données

Travail en groupe