

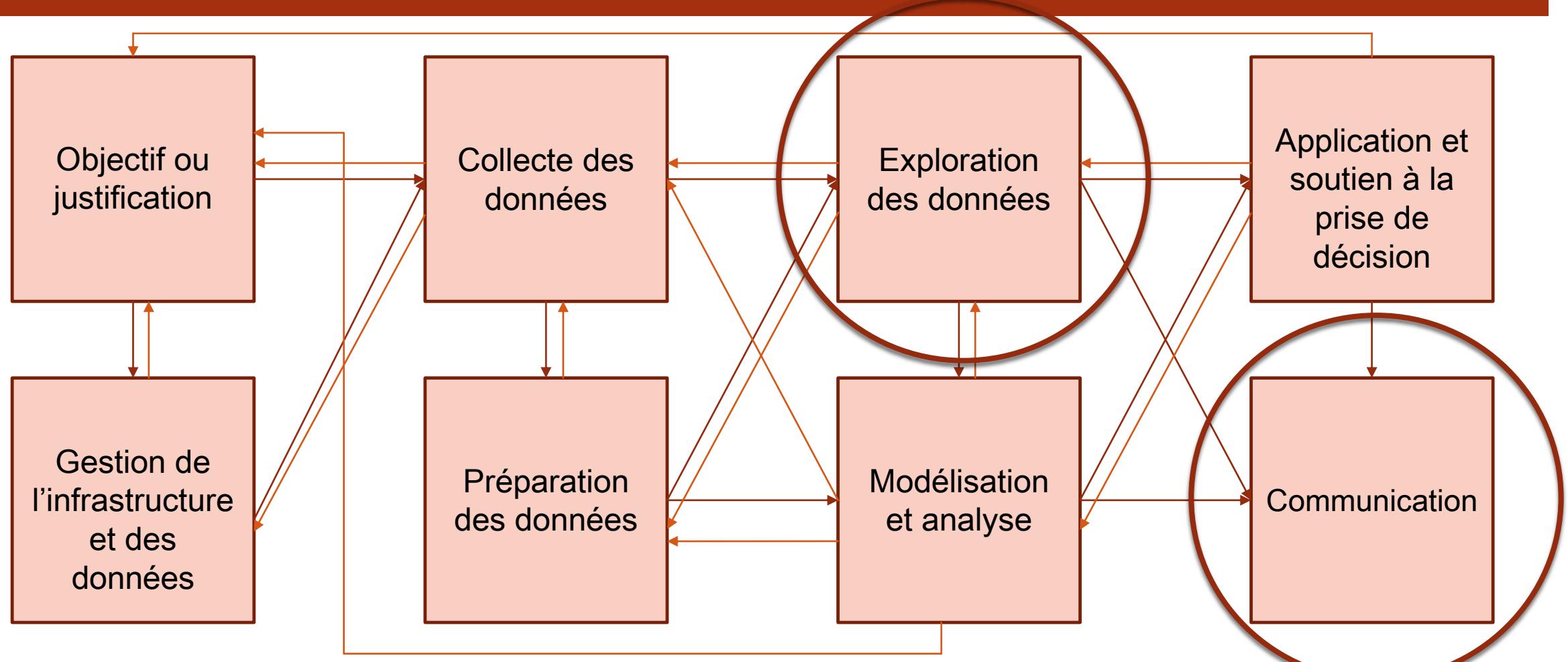
# EXPLORATION ET VISUALISATION DES DONNÉES

FORMATION AVANCÉE EN SCIENCE DES DONNÉES I

« L'horizon des découvertes n'est plus limité par la collecte et le traitement des données, mais plutôt par leur gestion, leur analyse et leur visualisation. »

@DamianMingle

# LE PROCESSUS D'ANALYSE (DANS TOUT SON DÉSORDRE)



# APERÇU

1. Exploration des données
2. Visualisation des données avant l'analyse
3. Visualisation des données après l'analyse
4. Catalogue de visualisations
5. Tableau d'honneur et tableau d'horreur

# OBJECTIFS D'APPRENTISSAGE

Comprendre les différents rôles de la visualisation de données dans le processus d'analyse.

Améliorer la compréhension des méthodes de représentation simultanée de dimensions multiples.

Améliorer la capacité à juger du nombre de dimensions représentées dans un graphique.

Comprendre certaines des stratégies et des considérations permettant de concevoir de bonnes visualisations après l'analyse.

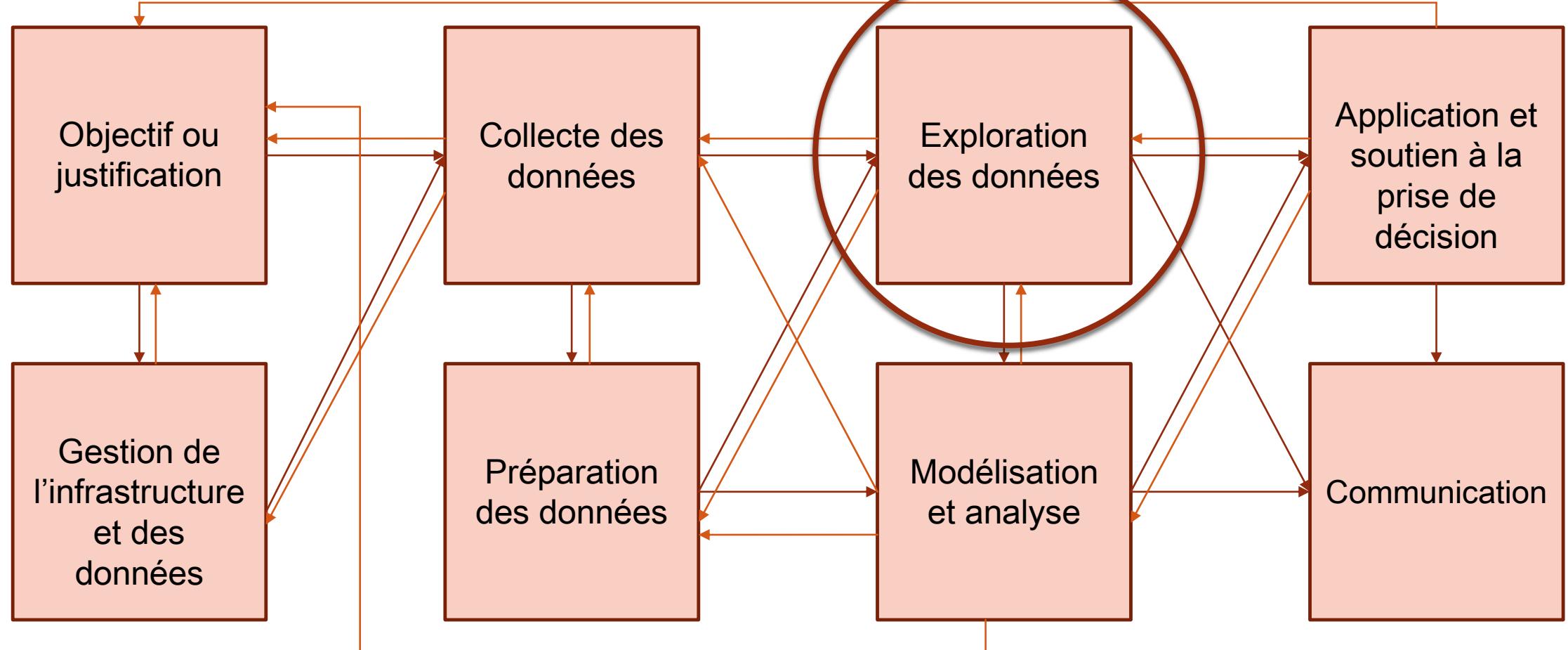
Comprendre la différence entre une visualisation et une infographie.

Améliorer la capacité à juger de la qualité des visualisations de données.

# EXPLORATION DES DONNÉES

EXPLORATION ET VISUALISATION DES DONNÉES

# LE PROCESSUS D'ANALYSE (DANS TOUT SON DÉSORDRE)



# APPRENDRE À CONNAÎTRE VOS DONNÉES



Bonjour, jeu de données, moi c'est Mathieu!



Enchanté Mathieu, je suis le jeu de données!

## QUESTIONS DE BASE

Quel système est représenté par vos données – objets, caractéristiques, relations?

Comment vos données représentent-elles ce système – quel est le modèle de données?

Qui a créé le jeu de données? Quand? Dans quel but?

À supposer qu'il s'agit d'un fichier bidimensionnel (fichier plat), que représentent les rangées? Que représentent les colonnes?

Avez-vous toute les informations nécessaires (**métadonnées**) pour répondre à ces questions? Où pouvez-vous obtenir davantage d'information?

# SOMMAIRE DES DONNÉES SANS VISUALISATION

	C1	N03	NH4
Min.	: 0.222	: 0.000	: 5.00
1st Qu.	: 10.994	: 1.147	: 37.86
Median	: 32.470	: 2.356	: 107.36
Mean	: 42.517	: 3.121	: 471.73
3rd Qu.	: 57.750	: 4.147	: 244.90
Max.	: 391.500	: 45.650	: 24064.00
NA's	: 16	: 2	: 2

season

Length: 340

Class :character

Mode :character

autumn spring summer winter

80 84 86 90

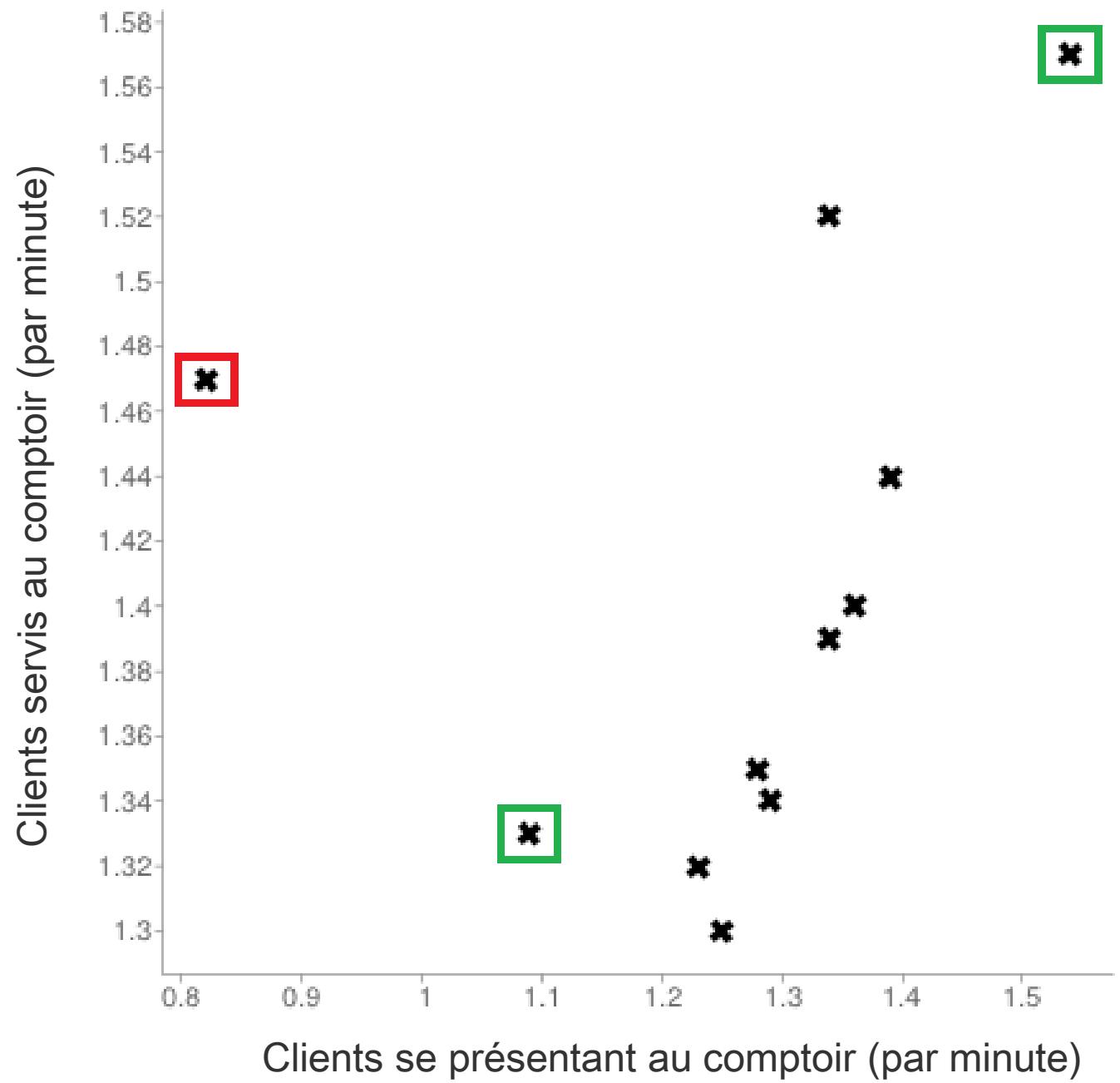
# VISUALISATION DES DONNÉES AVANT L'ANALYSE

EXPLORATION ET VISUALISATION DES DONNÉES

## UTILISATION AVANT L'ANALYSE

La visualisation des données peut être utile pour préparer l'analyse :

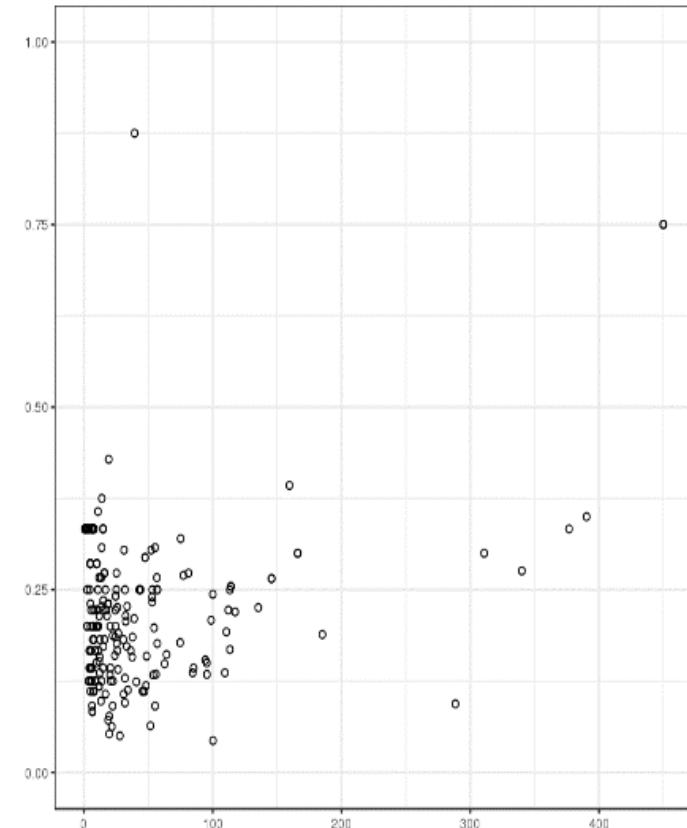
- **Détection des anomalies**  
Entrées invalides, valeurs manquantes, données aberrantes
- **Mise en forme des transformations de données**  
Compartimentage, uniformisation, transformations de Box-Cox, transformations de style analyse en composantes principales (ACP)
- **Familiarisation avec les données**  
L'analyse de données est un art, analyse exploratoire
- **Détection de structures de données cachées**  
Agrégation, associations, motifs renseignant la prochaine étape de l'analyse



# REPRÉSENTATION D'OBSERVATIONS À VARIABLES MULTIPLES

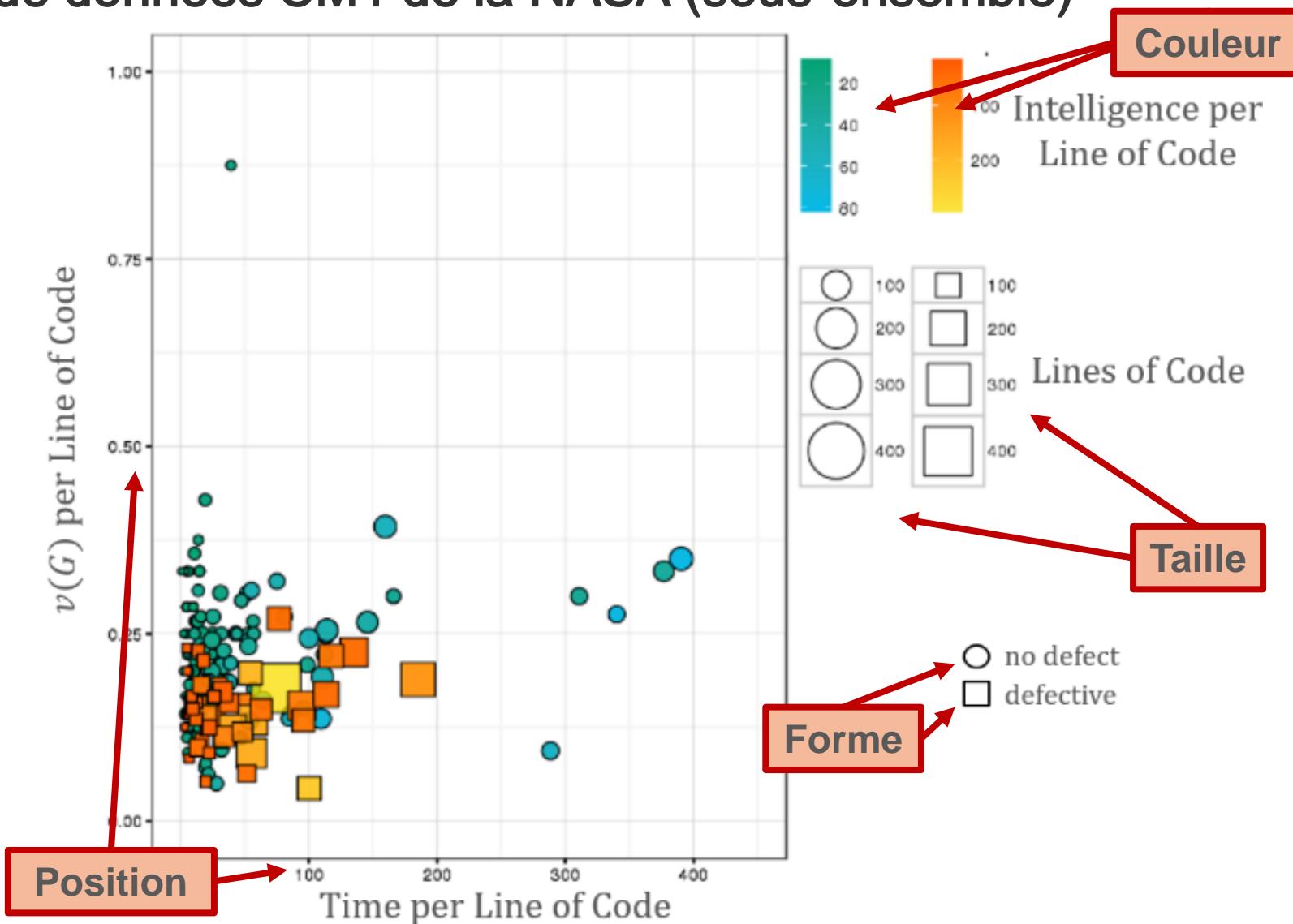
Deux variables peuvent être représentées selon la position sur un plan. Des facteurs additionnels peuvent être représentés par différents moyens :

- taille
- couleur
- valeur
- texture
- orientation d'une droite
- forme
- (mouvement?)



Jeu de données CM1 de la NASA  
(sous-ensemble) [data-action-lab.com](http://data-action-lab.com) 

## Jeu de données CM1 de la NASA (sous-ensemble)



# VISUALISATIONS COURANTES POUR L'EXPLORATION DES DONNÉES

Graphique linéaire/graphique à traits/droite numérique

Histogramme

(Diagramme à moustaches)

Graphique linéaire

Diagramme en bâtons

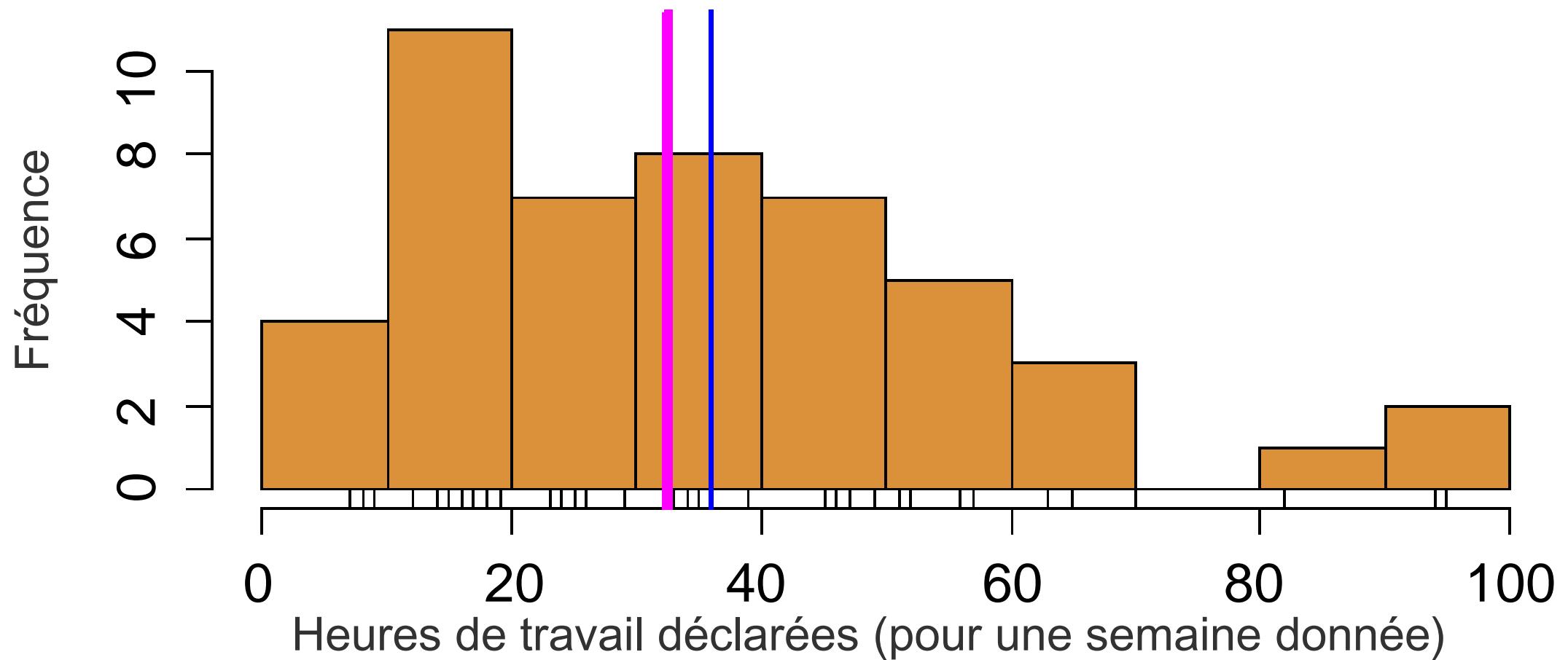
Nuage de points

# GRAPHIQUE À TRAITS

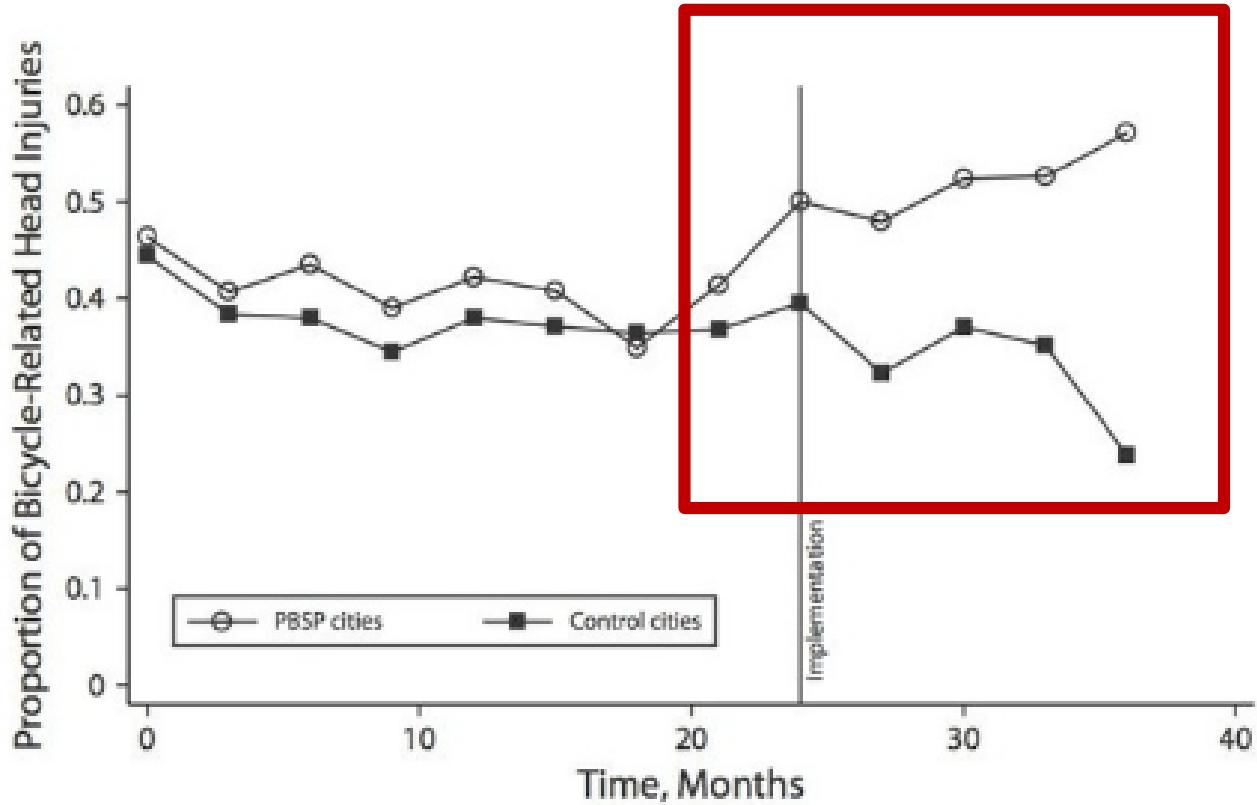


- Les trous le long de la droite numérique indiquent l'absence de ces valeurs dans le jeu de données.
- Rappel : Ce graphique ne représente pas nécessairement les données du jeu de données dans l'ordre – il s'agit d'une droite numérique, elle n'indique donc que les valeurs présentent dans le jeu.
- Les valeurs identiques se chevauchent.

# HISTOGRAMME



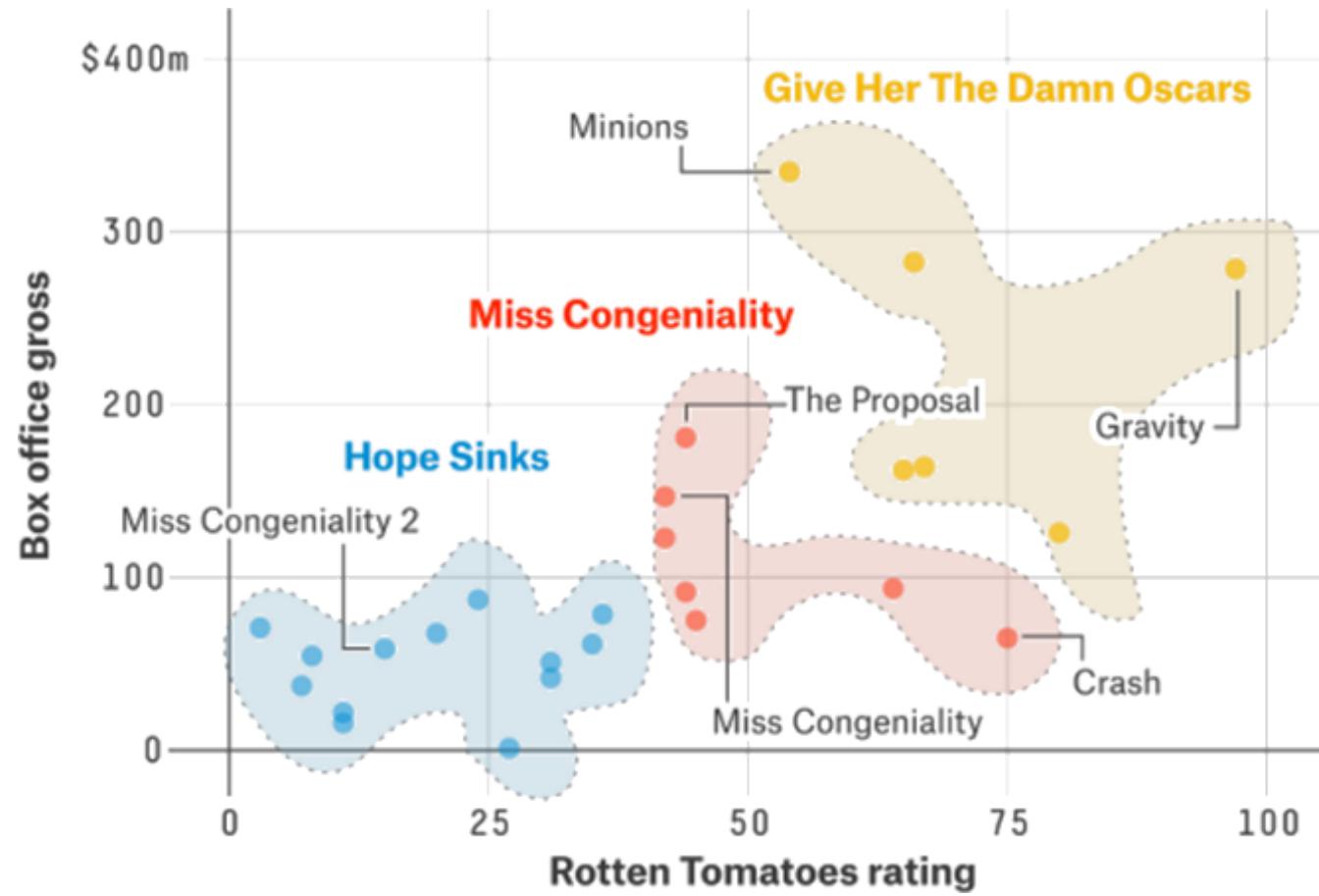
## GRAPHIQUE LINÉAIRE

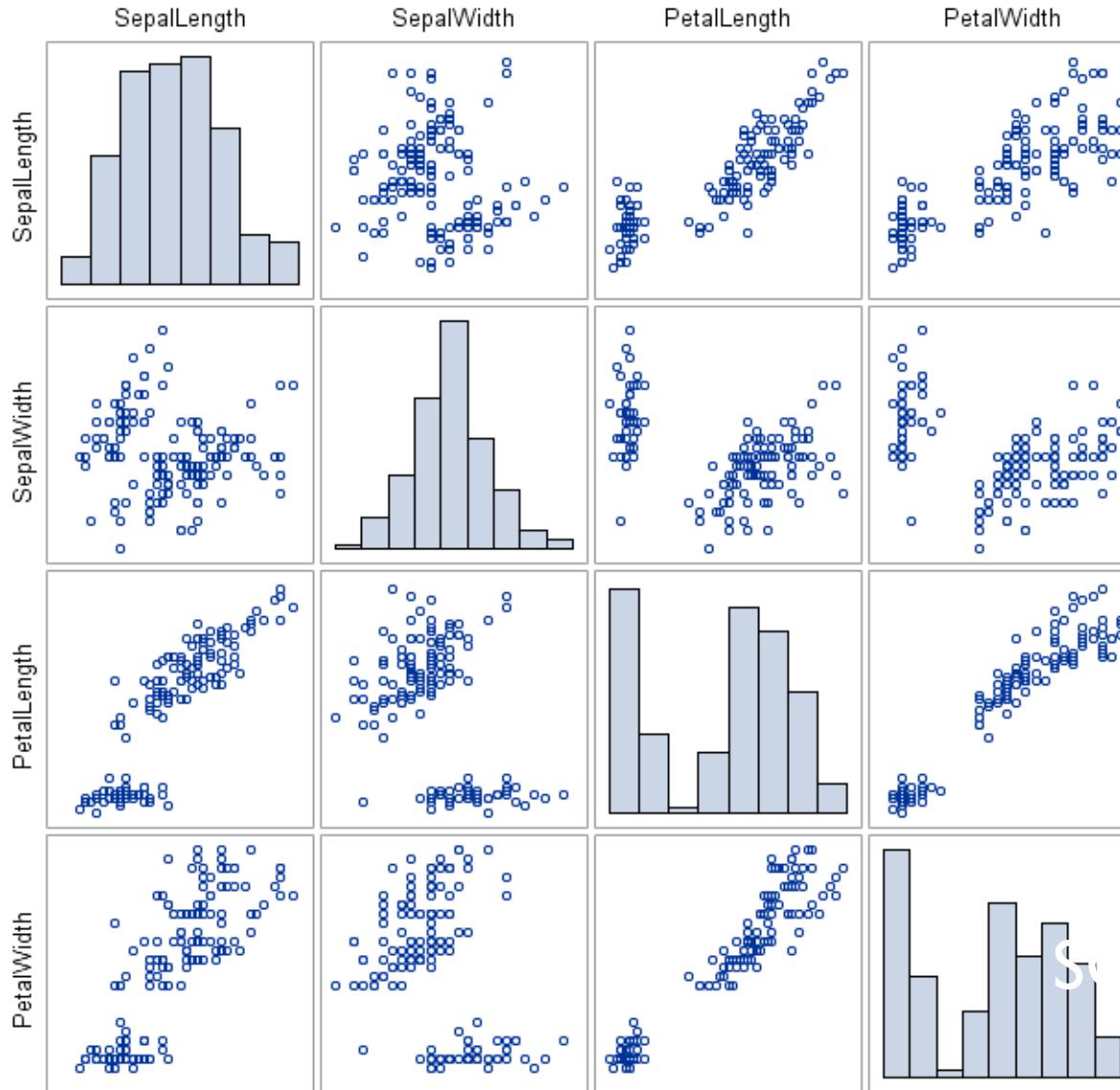


Proportion des blessures à la tête parmi l'ensemble des blessures liées au cyclisme dans les villes possédant un service de vélopartage et les villes de référence, en fonction de la date de l'intervention (ligne verticale); Amérique du Nord.

[Graves et coll., *Am.J.Phys.Health*, 2014]

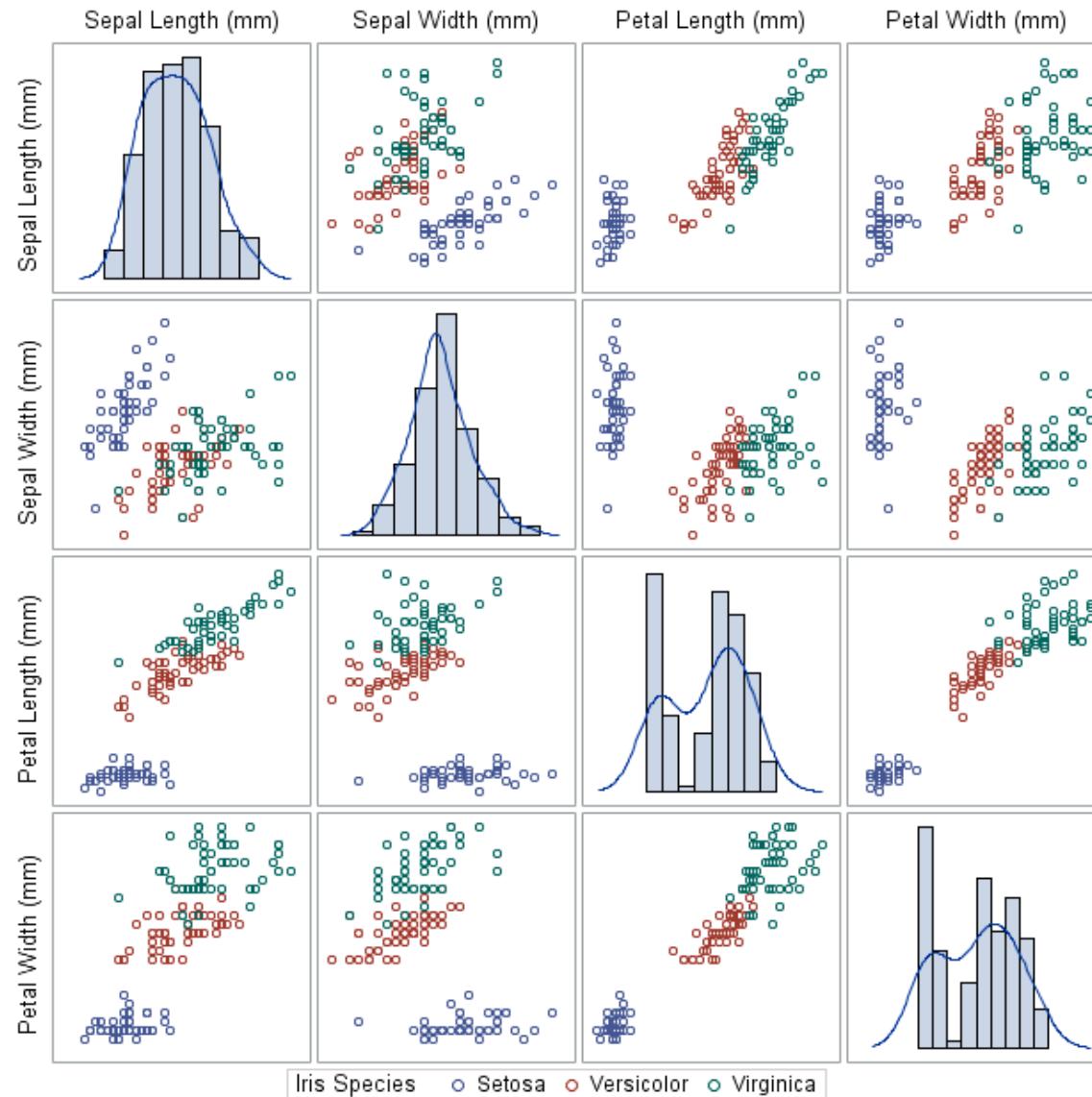
# NUAGE DE POINTS



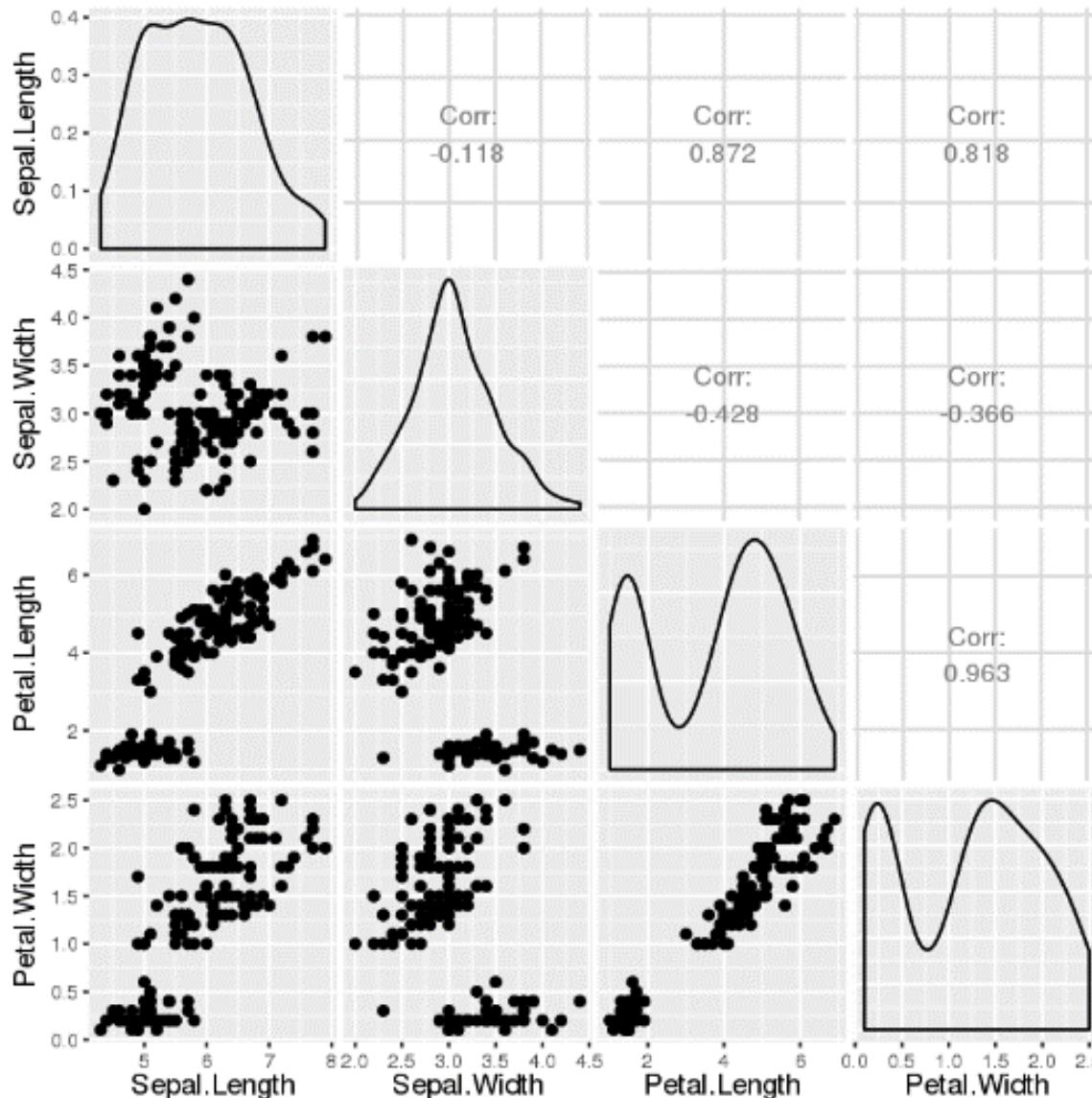


[Fait avec la procédure corr de SAS]  
data-action-lab.com



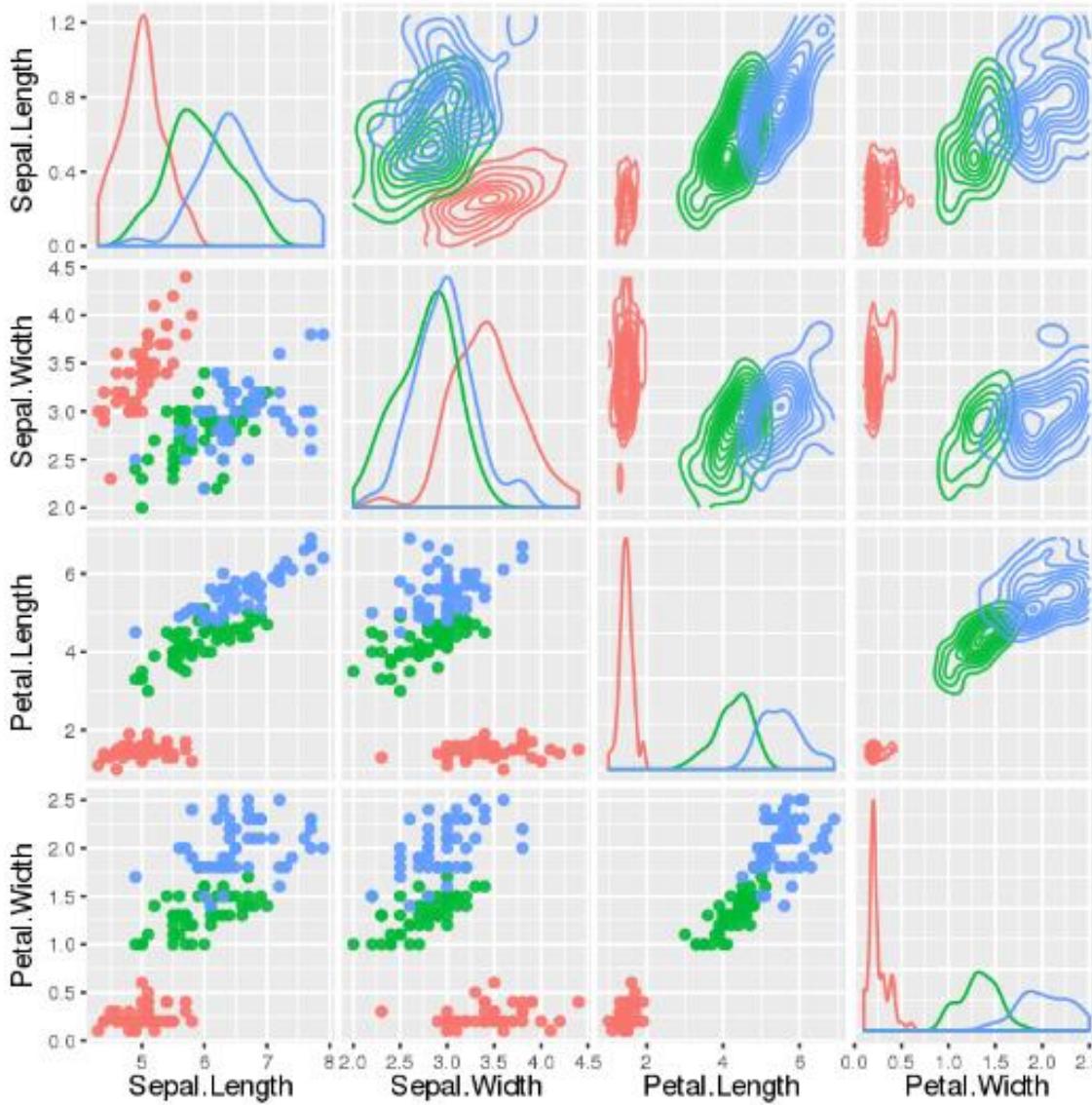


[Fait avec la procédure sgscatter de SAS]



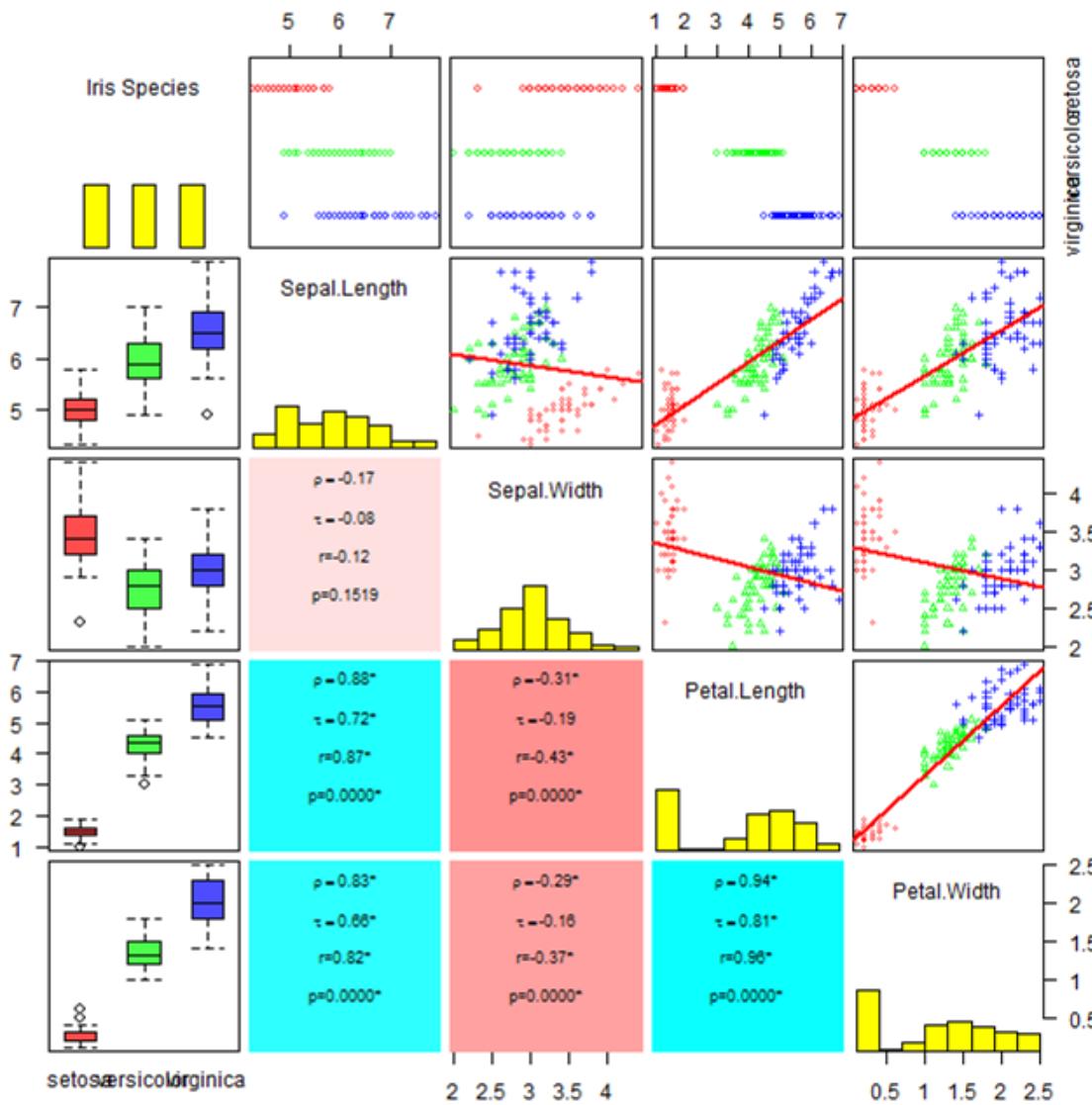
[Fait avec la commande ggpairs de R]  
data-action-lab.com





[Fait avec la commande ggpairs de R]  
data-action-lab.com



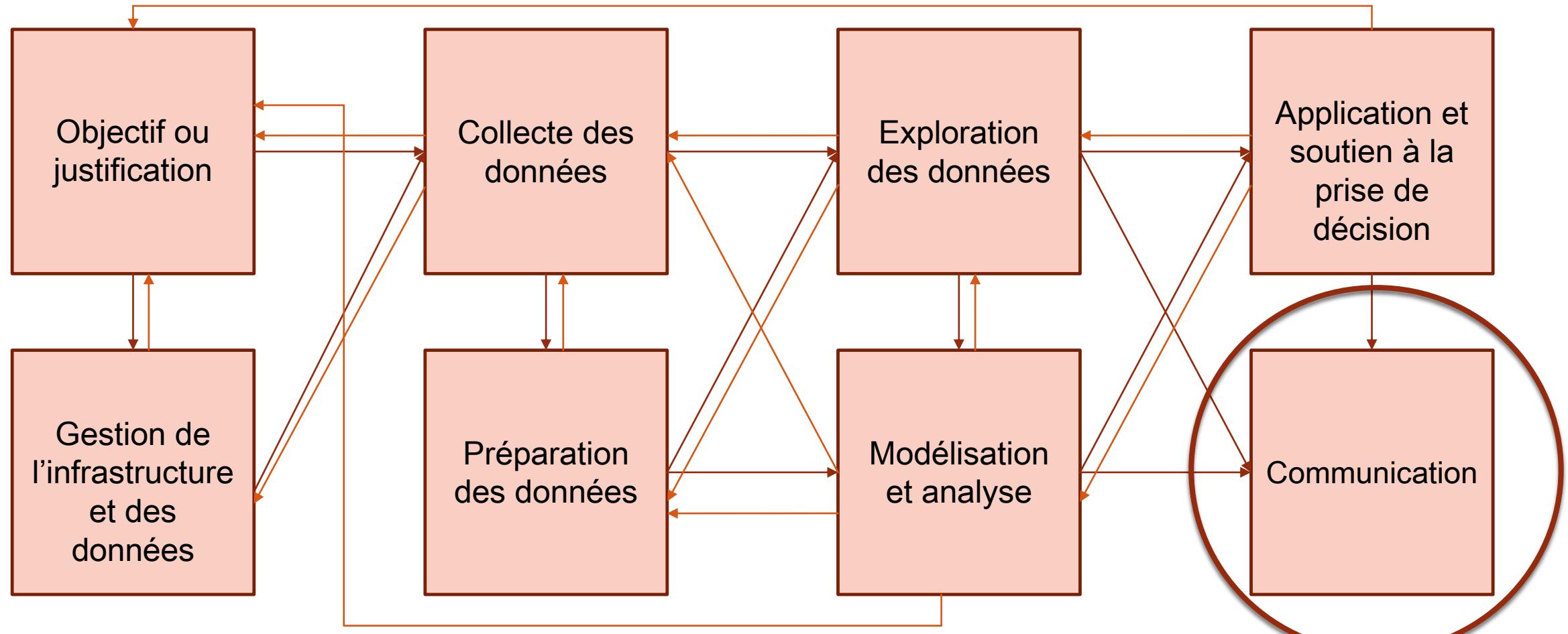


Ce graphique commence-t-il à être trop chargé?

# VISUALISATION DES DONNÉES APRÈS L'ANALYSE

EXPLORATION ET VISUALISATION DES DONNÉES

# LE PROCESSUS D'ANALYSE (DANS TOUT SON DÉSORDRE)



# PRINCIPES FONDAMENTAUX DU DESIGN ANALYTIQUE

Le **raisonnement et la communication** sont interreliés dans nos vies et notre univers causal, dynamique et multivarié.

La **symétrie** dans les visualisations : les consommateurs devraient rechercher exactement ce que les producteurs offrent, soit :

- des comparaisons pertinentes
- des réseaux causaux et leur structure sous-jacente
- des relations multivariées
- des données intégrées et pertinentes
- une documentation transparente
- un accent sur le contenu

## ACCESSIBILITÉ

On peut traduire un tableau en braille assez facilement, mais ce n'est pas toujours possible pour un graphique.

L'une des solutions peut être de décrire les caractéristiques et les structures de la visualisation... **à condition de pouvoir les repérer.**

Les analyses doivent produire des visualisations claires et pertinentes, mais ils doivent également les décrire d'une façon qui permet d'en « saisir » la portée.

## ACCESSIBILITÉ

Les analystes doivent avoir compris tous les éléments d'information transmis, ce qui n'est pas nécessairement réaliste.

### **Perception des données :**

- représentations texturées
- conversion texte-parole
- utilisation de sons ou de musique
- représentations odorantes ou axées sur le goût (?!?)

# INFOGRAPHIE

## Crée pour raconter une histoire (subjectivité)

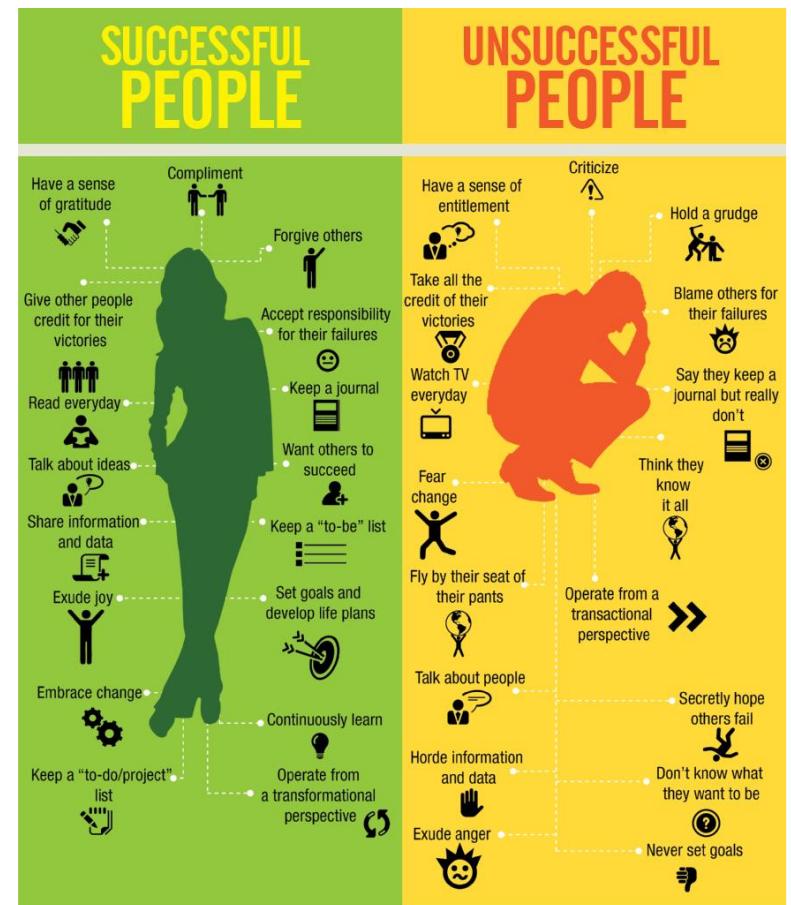
# Cible un public précis

## Autonome et indépendante

La conception graphique est un aspect clé

Ne peut généralement pas être réutilisée avec d'autres données

Peut comprendre de l'information impossible à quantifier



# VISUALISATION DES DONNÉES

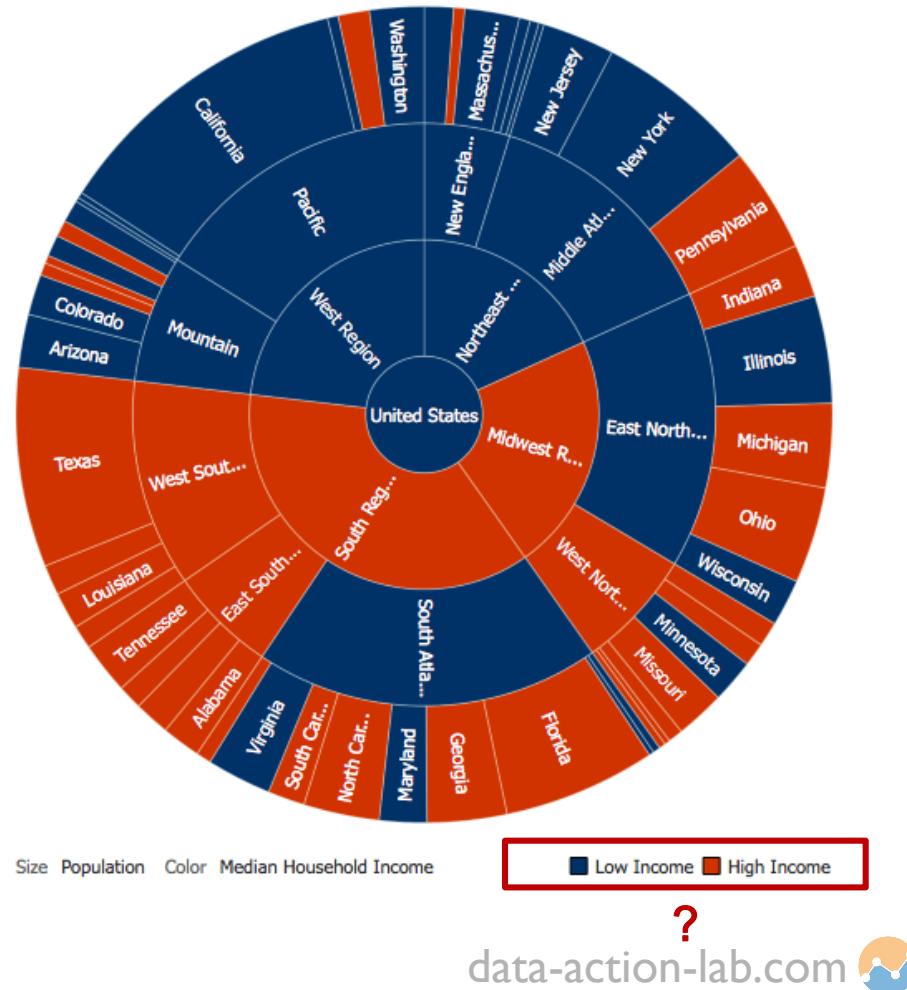
Une méthode et un objet à la fois (**objectivité**)

Met généralement l'accent sur des données **quantifiables**

Sert à extraire le sens des données ou à les rendre **accessibles** (les jeux de données peuvent être imposants et difficiles à manipuler)

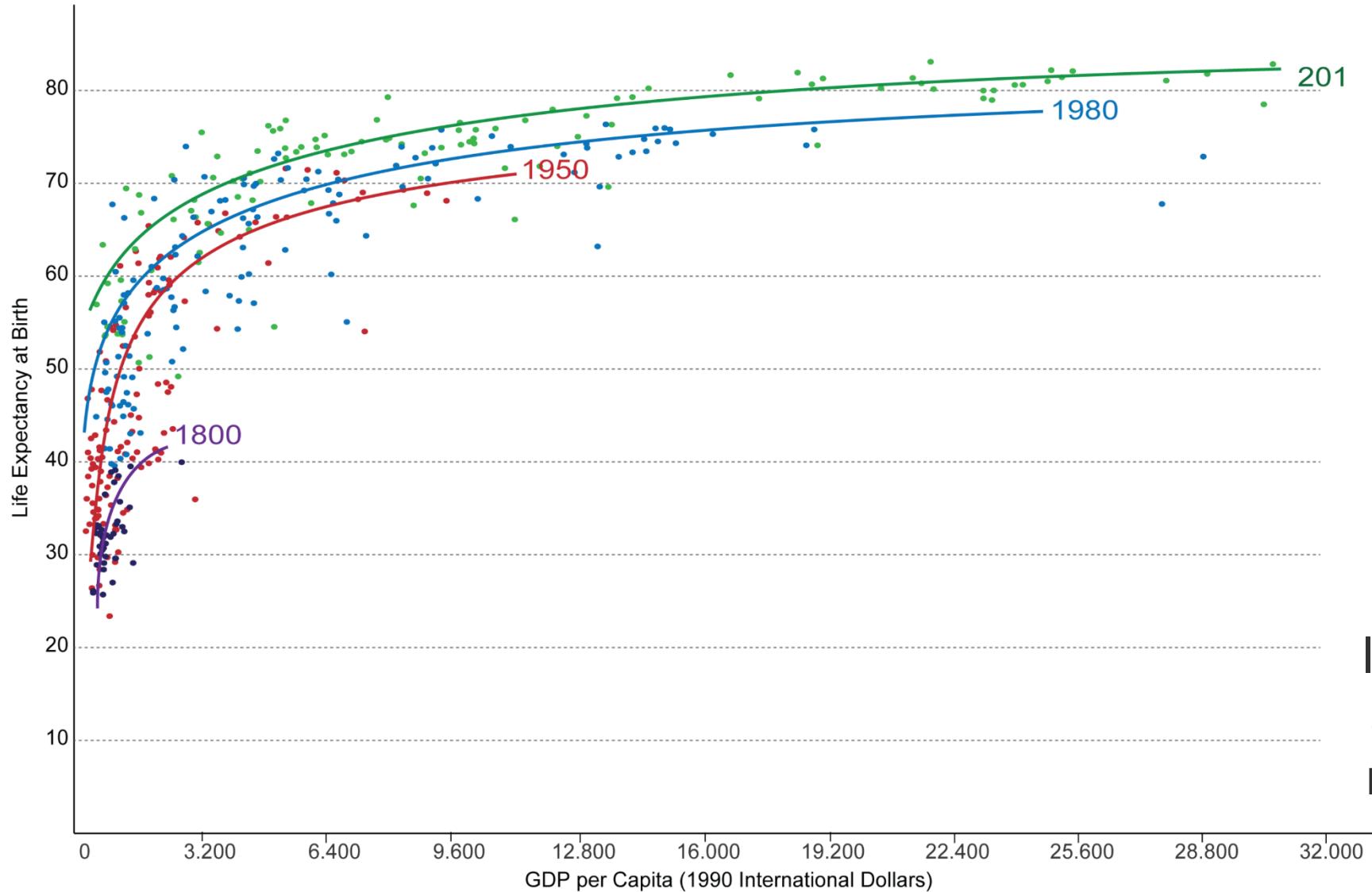
Peut être générée automatiquement

L'apparence est moins importante que l'**information** transmise par les données



## Life Expectancy vs. GDP per Capita from 1800 to 2012 – by Max Roser

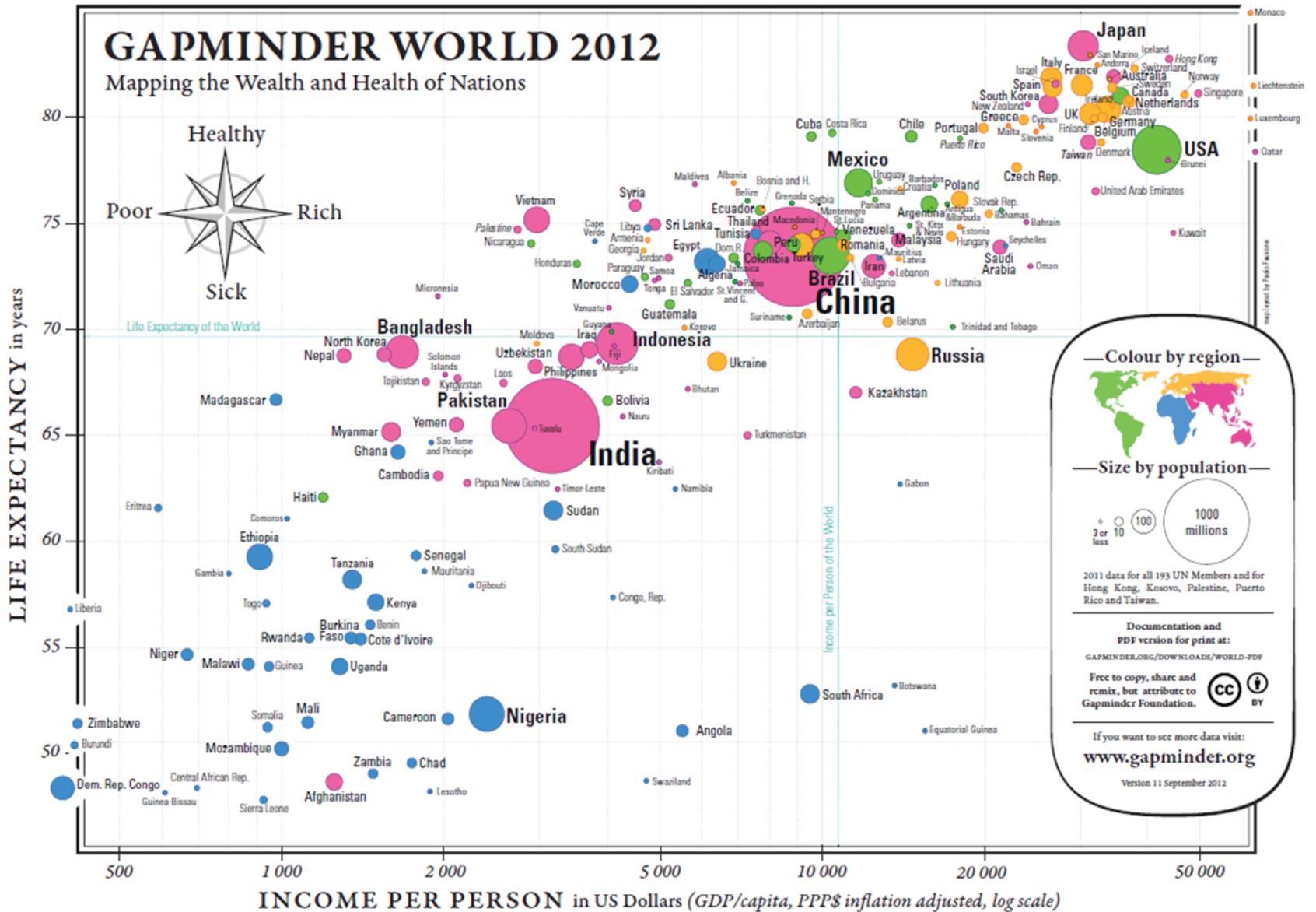
GDP per capita is measured in International Dollars. This is a currency that would buy a comparable amount of goods and services U.S. dollar would buy in the United States in 1990. Therefore incomes are comparable across countries and across time.



Ce graphique représente la relation entre l'espérance de vie et le PIB par habitant.

En général, plus le PIB d'un pays élevé, meilleure est son espérance de vie.

La corrélation semble suivre une courbe logarithmique : l'augmentation de l'espérance de vie par unité additionnelle de PIB est de moins en moins importante à mesure que le PIB augmente.



## PRÉSENTATION DES RÉSULTATS DE L'ANALYSE

Les graphiques devraient être **clairs** et **attrayants**.

Ce ne sont pas toutes les jolies images qui ont une histoire à raconter, mais s'il est impossible de raconter une histoire à l'aide d'une jolie image, peut-être qu'il est temps de revoir l'histoire...

De nouvelles techniques de représentation graphique apparaissent régulièrement – il est trop tôt pour déterminer lesquelles résisteront à l'épreuve du temps.

Il ne faut pas avoir peur d'essayer quelque chose de nouveau tant que cela permet de **transmettre l'information souhaitée**.

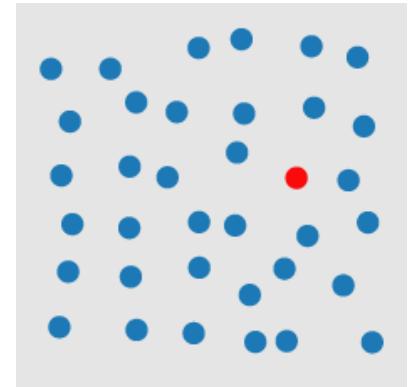
# TRAITEMENT VISUEL

La perception est **fragmentée** – les yeux sont constamment en mode balayage.

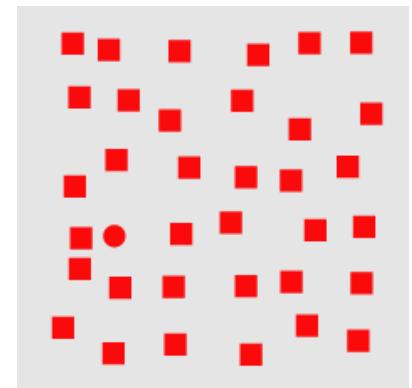
Les centres de traitement visuel sont constamment à la recherche de motifs.

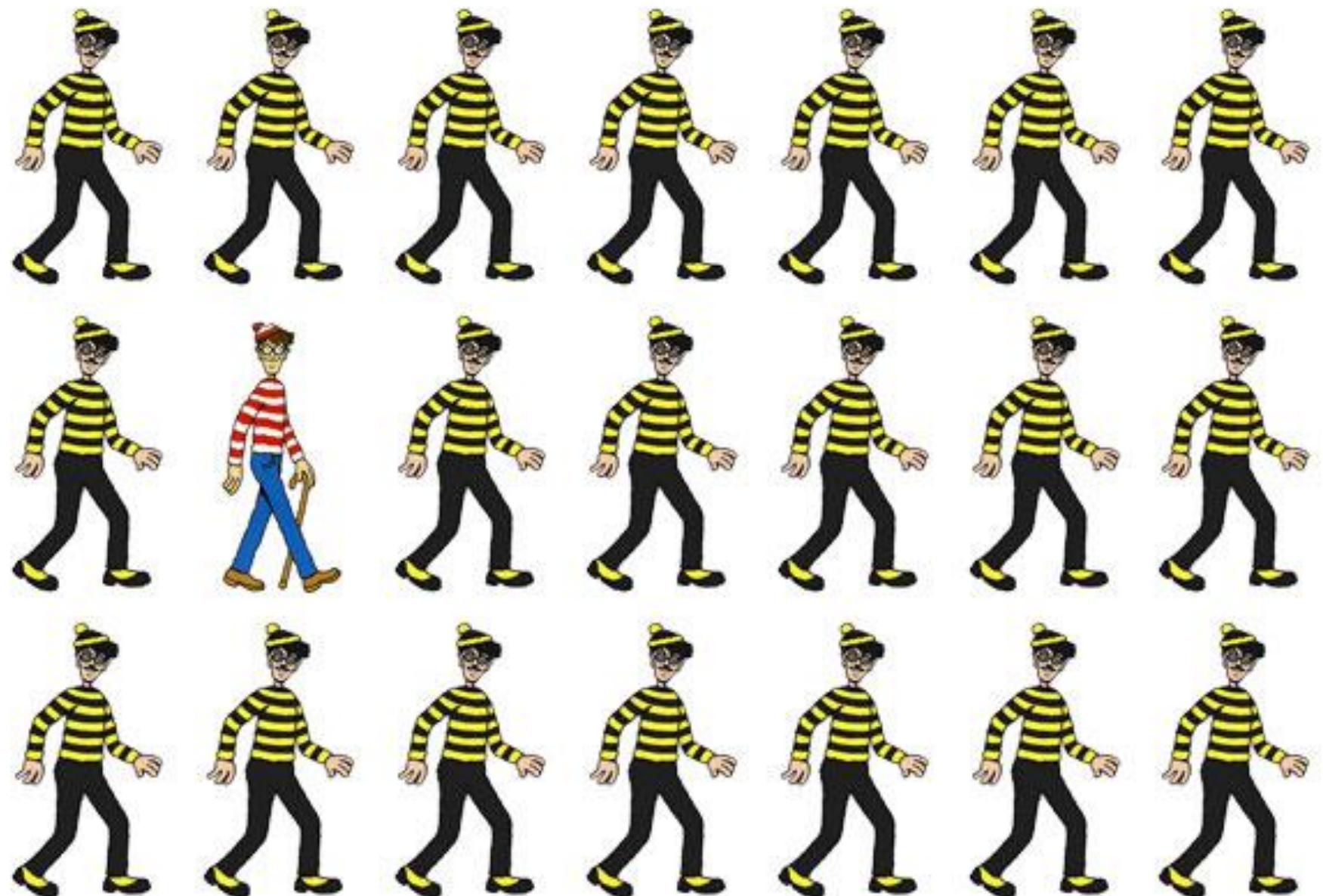
- **Traitement préattentif** : rapide , instinctif, efficace, superficiel, collecte d'information et détection de motifs.  
caractéristiques → motifs → objets
- **Traitement attentif** : lent, délibéré, focalisé, découverte de caractéristiques à l'intérieur des motifs.  
objets → motifs → caractéristiques

Vision préattentive



Vision attentive







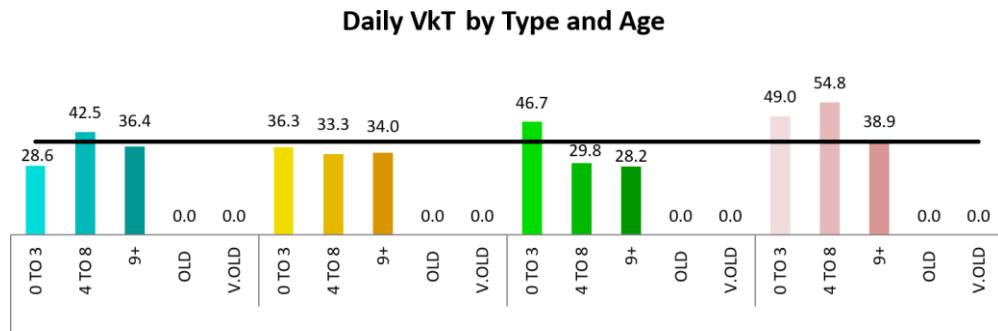
## RÈGLES DE BASE

### 1. Examiner les données

Aberrations, pics, anomalies.

### 2. Expliquer l'encodage

Ne pas présumer que le lecteur comprend la signification de tous les éléments.



### 3. Étiqueter les axes

Il est important d'afficher l'échelle.

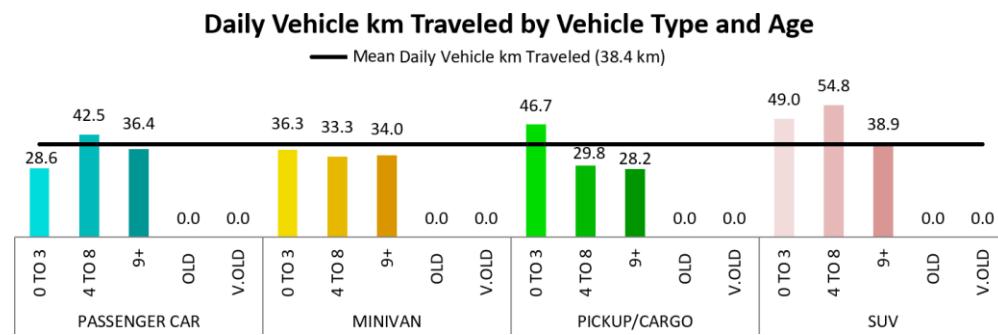
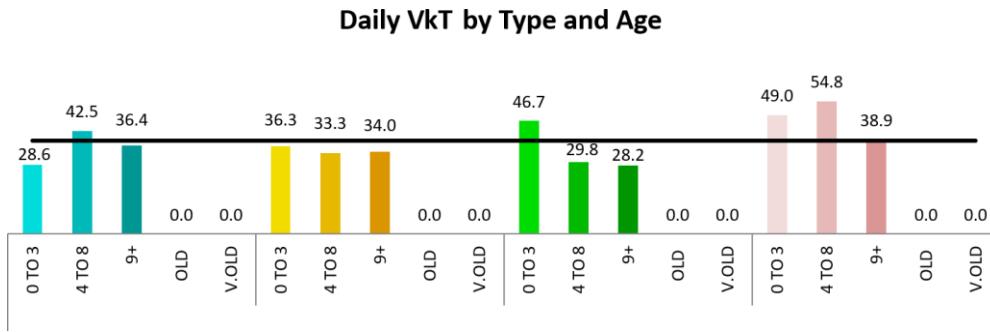
# RÈGLES DE BASE

## 1. Examiner les données

Aberrations, pics, anomalies.

## 2. Expliquer l'encodage

Ne pas présumer que le lecteur comprend la signification de tous les éléments.



## 3. Étiqueter les axes

Il est important d'afficher l'échelle.

## RÈGLES DE BASE

### 4. Afficher les unités

Ne pas forcer le lecteur à faire des suppositions.



### 5. Respecter les principes géométriques

L'échelle des cercles et des formes en deux dimensions est définie par leur superficie, celle des diagrammes à bâtons, par leur longueur.

### 6. Indiquer les sources

Éviter tout risque d'accusation de plagiat et permettre aux lecteurs d'en apprendre plus.

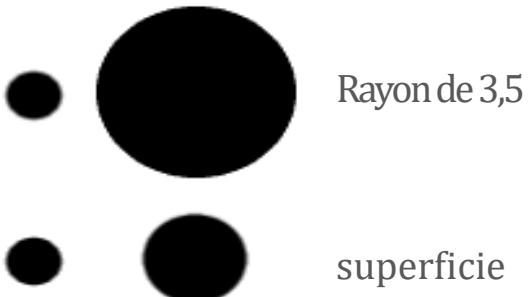
### 7. Penser au public

Une affiche peut contenir plus de texte, mais un diaporama devrait être concis.

## RÈGLES DE BASE

### 4. Afficher les unités

Ne pas forcer le lecteur à faire des suppositions.



### 5. Respecter les principes géométriques

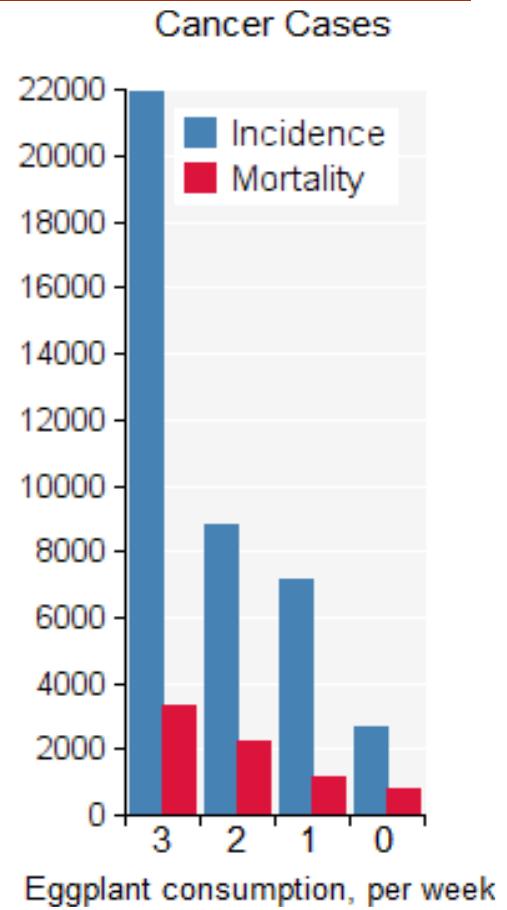
L'échelle des cercles et des formes en deux dimensions est définie par leur superficie, celle des diagrammes à bâtons, par leur longueur.

### 6. Indiquer les sources

Éviter tout risque d'accusation de plagiat et permettre aux lecteurs d'en apprendre plus.

### 7. Penser au public

Une affiche peut contenir plus de texte, mais un diaporama devrait être concis.



## RÈGLES DE BASE

### 4. Afficher les unités

Ne pas forcer le lecteur à faire des suppositions.



Rayon de  
3,5

Superficie  
de 3,5

### 5. Respecter les principes géométriques

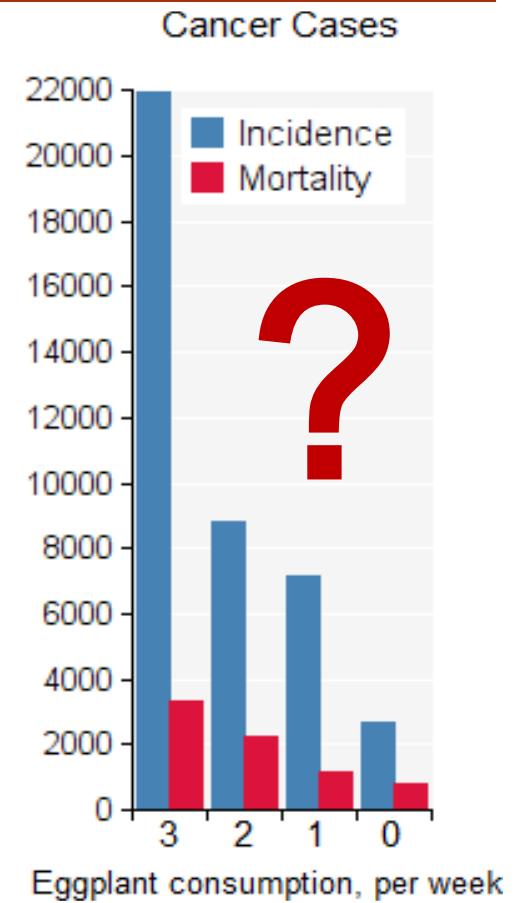
L'échelle des cercles et des formes en deux dimensions est définie par leur superficie, celle des diagrammes à bâtons, par leur longueur.

### 6. Indiquer les sources

Éviter tout risque d'accusation de plagiat et permettre aux lecteurs d'en apprendre plus.

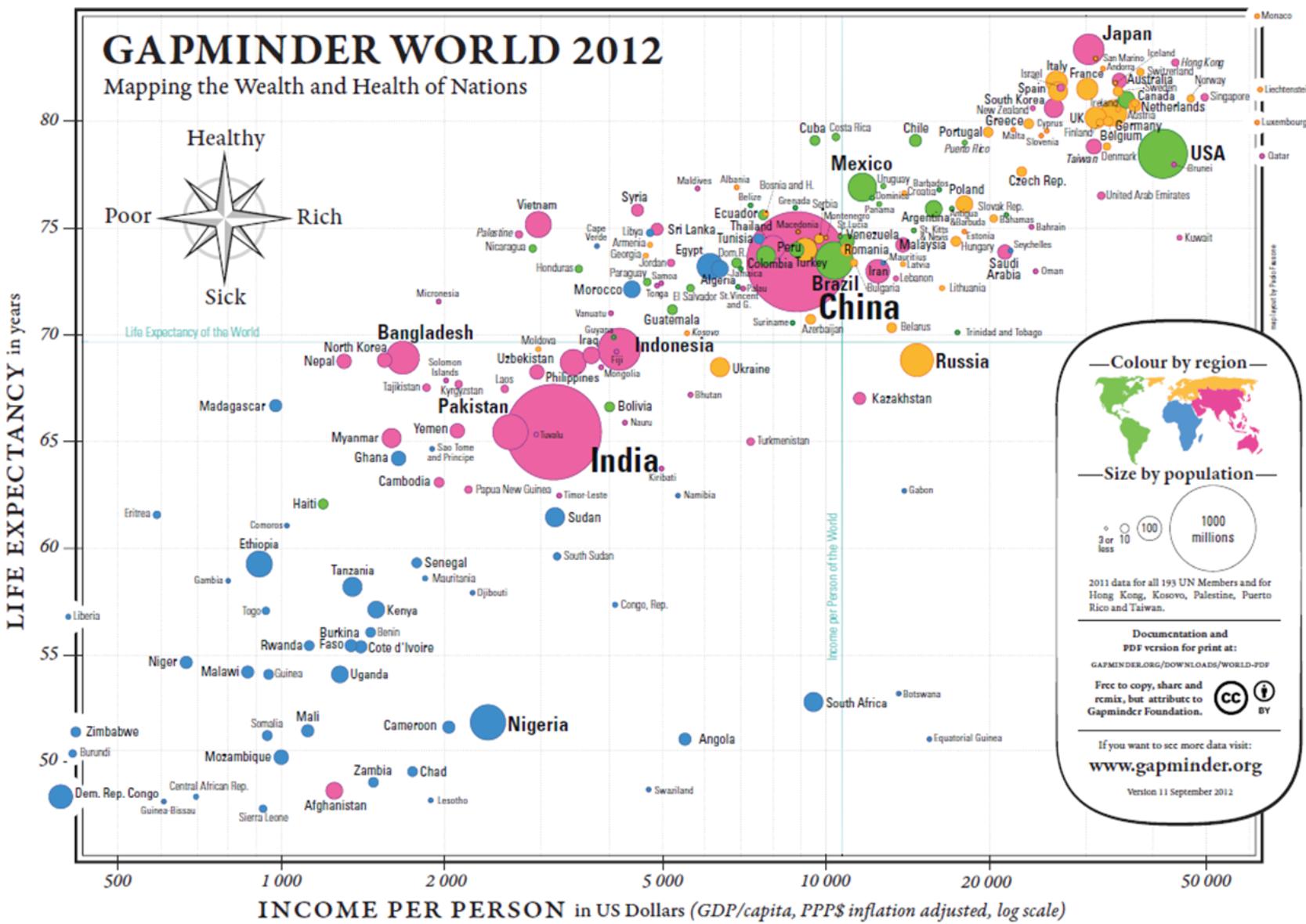
### 7. Penser au public

Une affiche peut contenir plus de texte, mais un diaporama devrait être concis.



## Exercice de groupe

- Comment ce visuel aide-t-il le public à comprendre les données?
- Repérez-vous des motifs intéressants dans ce visuel?



# DISCUSSION

**Le message passe-t-il?** L'intégration des données contribue à transmettre l'information importante.

Dans *La sémiologie graphique*, Bertin affirme que **les variables rétiniennes n'ont pas toutes le niveau d'efficacité** pour relayer ou représenter de l'information. Il peut être nécessaire de faire des essais pour trouver le meilleur choix dans un contexte donné.

L'addition de certain éléments conceptuels peut améliorer la compréhension des données.

La façon dont nous percevons les motifs influence notre interprétation de la représentation des données.

Les représentations de données ne devraient pas reposer sur une méthode de visualisation choisie au hasard. Le résultat variera selon la structure des données et la combinaison des questions étudiées.

# CATALOGUE DE VISUALISATION

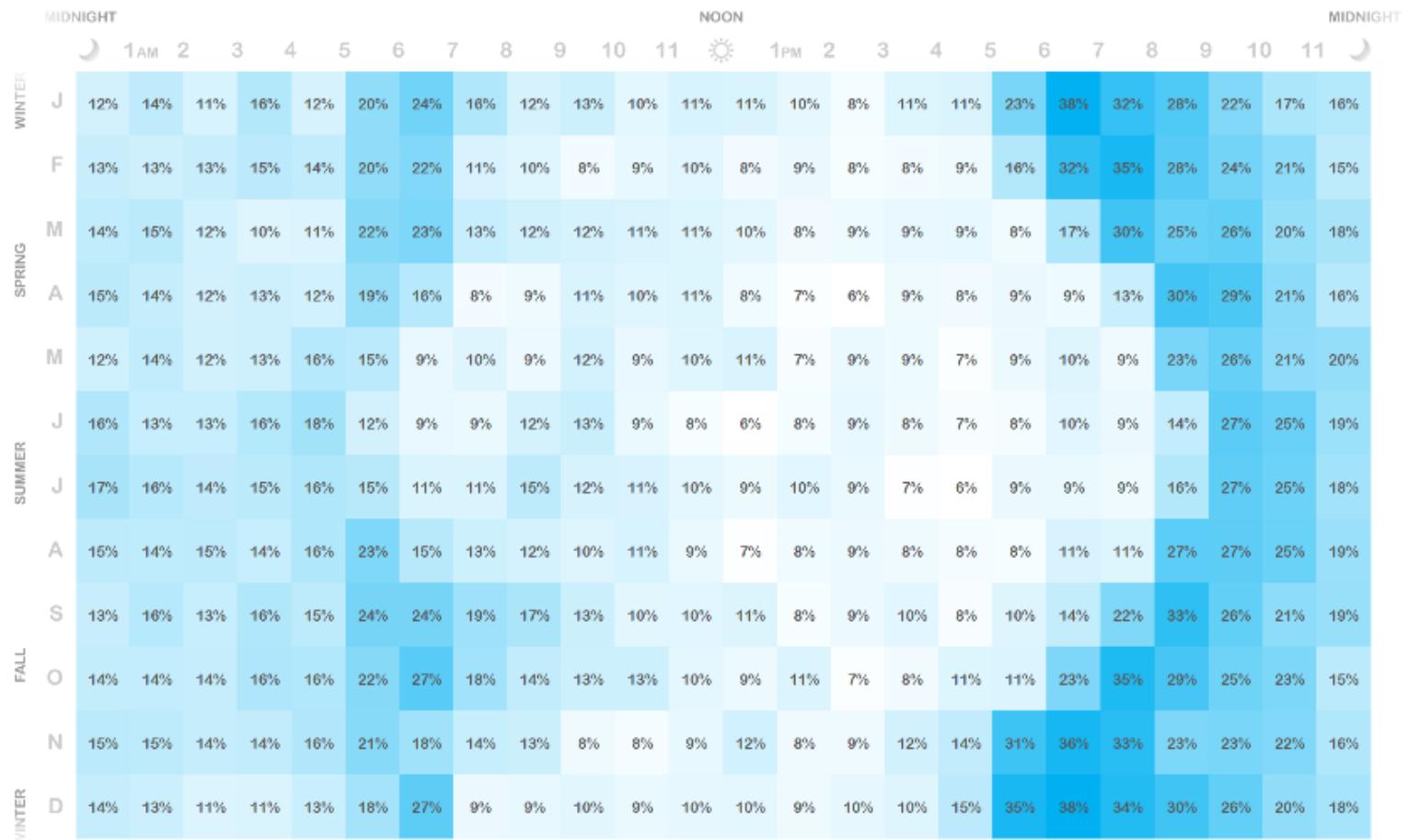
# CARTES DE DENSITÉ

## L'horizon du risque pour les piétons

Il s'agit du taux d'accidents de la route mortels impliquant des piétons, selon l'heure de la journée, tout au long de l'année.

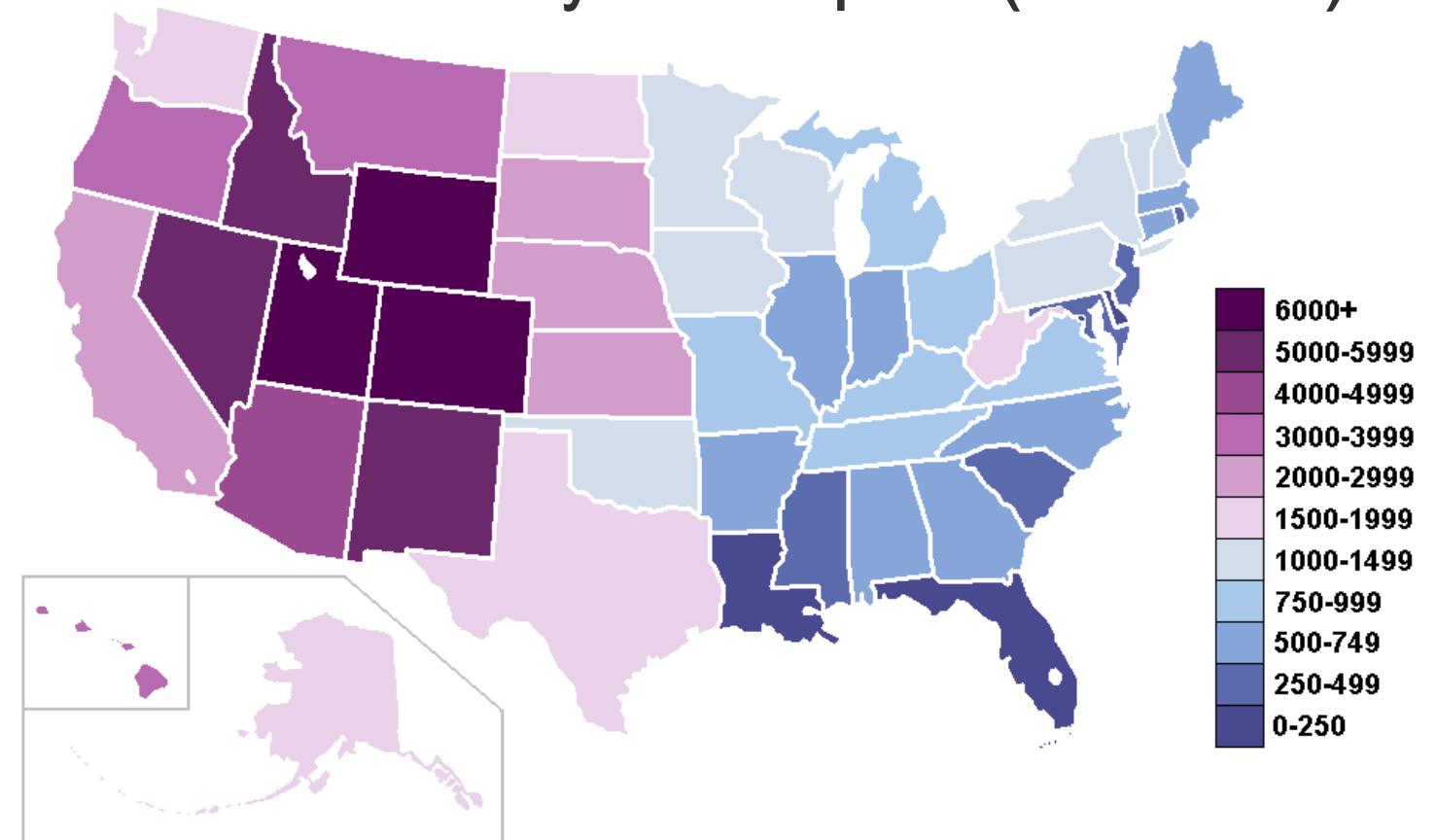
Le déplacement du soleil au fil des saisons se reflète par une courbe de risque élevé qui fait écho à la courbature de la terre même (**Remarque : ???**).

Source : Fatality Analysis  
Reporting System (NHTSA 2006-  
2010).



# CARTE DE DENSITÉ (CHOROPLÈTE)

Altitude moyenne en pieds (selon l'État)



## CARTE DE DENSITÉ

Idéale pour étudier les relations entre 3 ou 4 variables :

- l'une d'elle doit correspondre à un pourcentage ou à une valeur située dans une fourchette définie (pour permettre d'établir l'échelle de couleur aux fins de comparaison);
- l'autre peut être une variable de taille ou une variable nominale.

Il vaut mieux **compartimenter les données**, même si les variables des axes sont continues (réduire le nombre d'observations nécessaires pour donner un résultat utile).

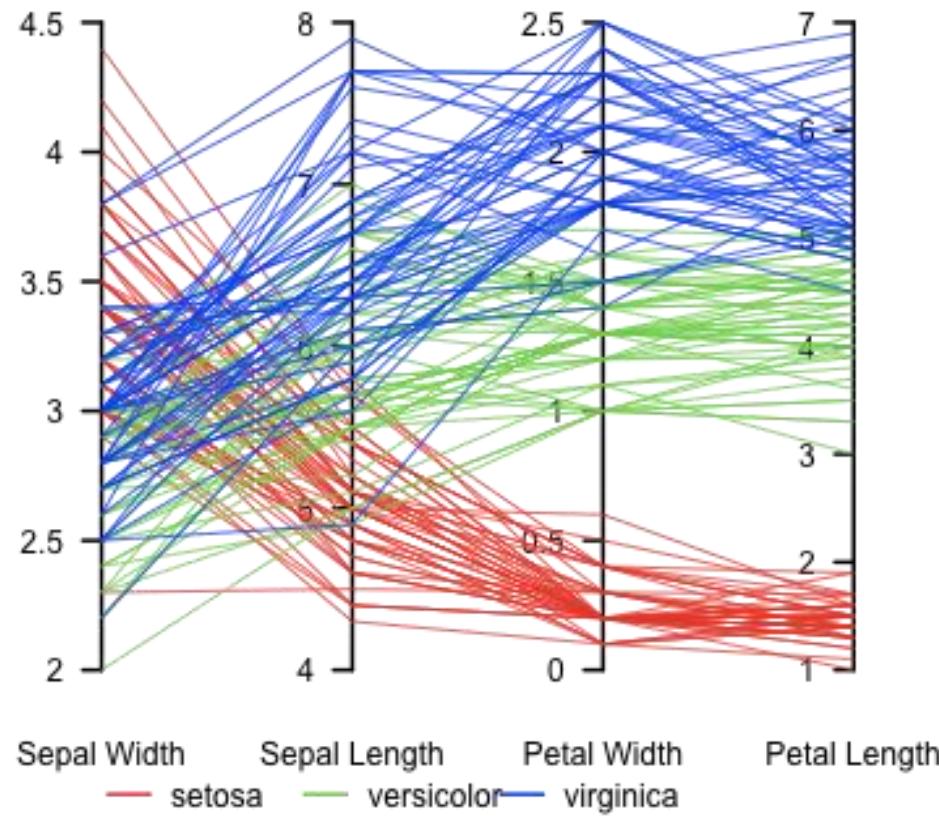
Plus facile à lire lorsque les couleurs sont choisies en fonction de la progression naturelle des tons.

**Rouge → Vert** ou **Rouge → Jaune → Vert**

Il s'agit d'un exemple (qui n'est pas optimal pour les daltoniens).

# COORDONNÉES PARALLÈLES

## Jeu de données sur les iris



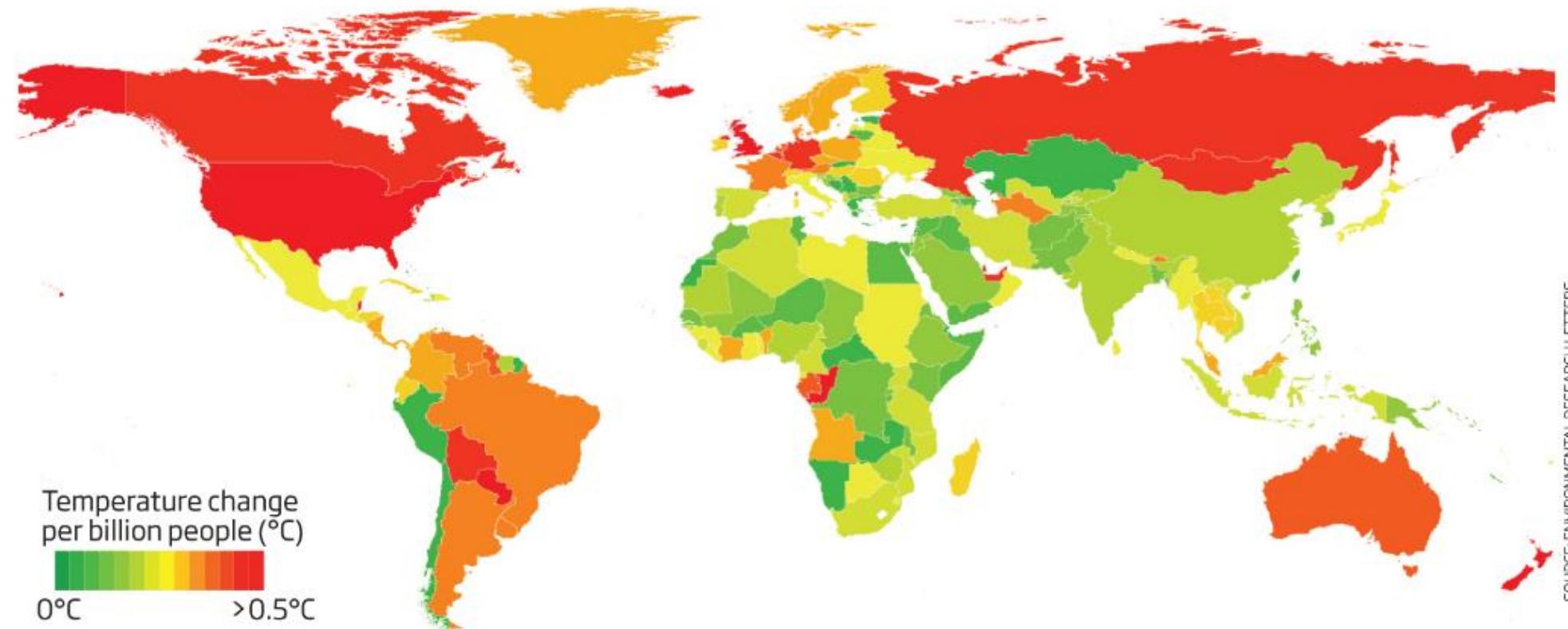
Fonctionne bien pour un nombre limité de sujet, ou lorsqu'il existe des distinctions claires entre les groupes.

Le chevauchement des tracés peut rendre ce type de graphique difficile à lire.

# CARTES GÉOGRAPHIQUES

## Global warming culprits, judged by population

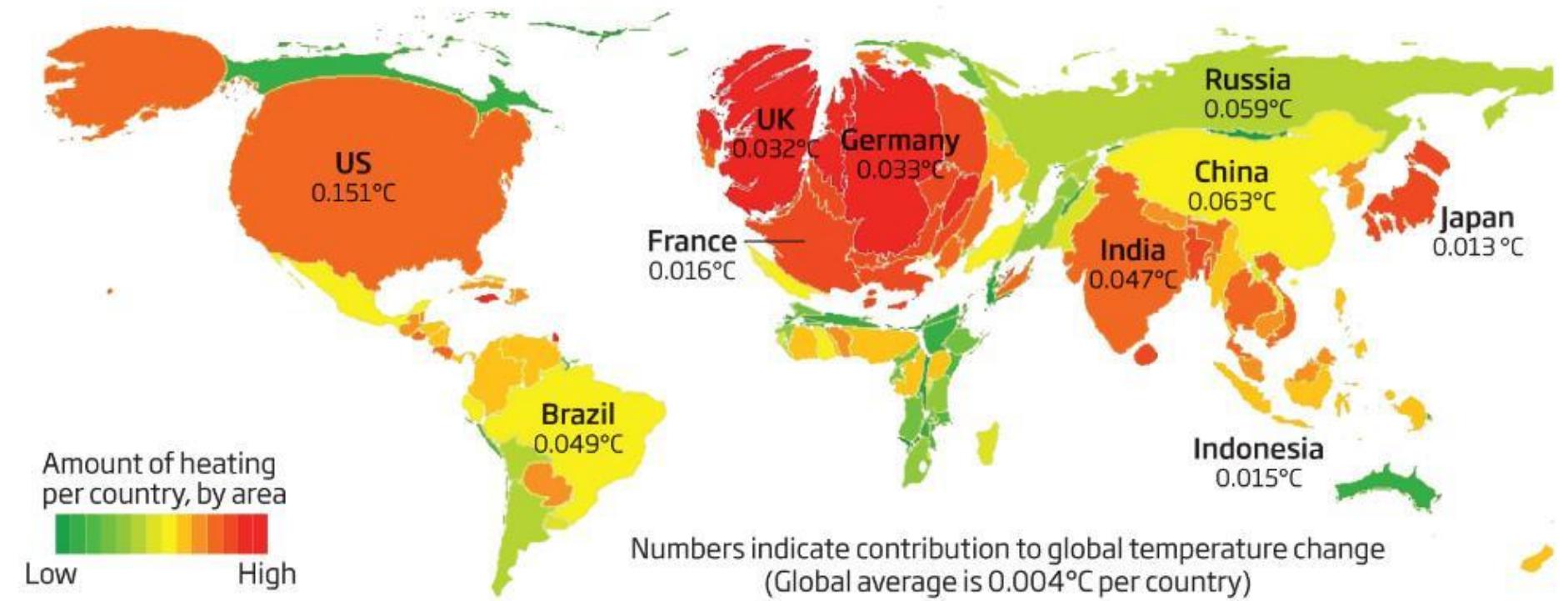
Countries that have caused more global warming per billion people are coloured red and low-emitters are dark green

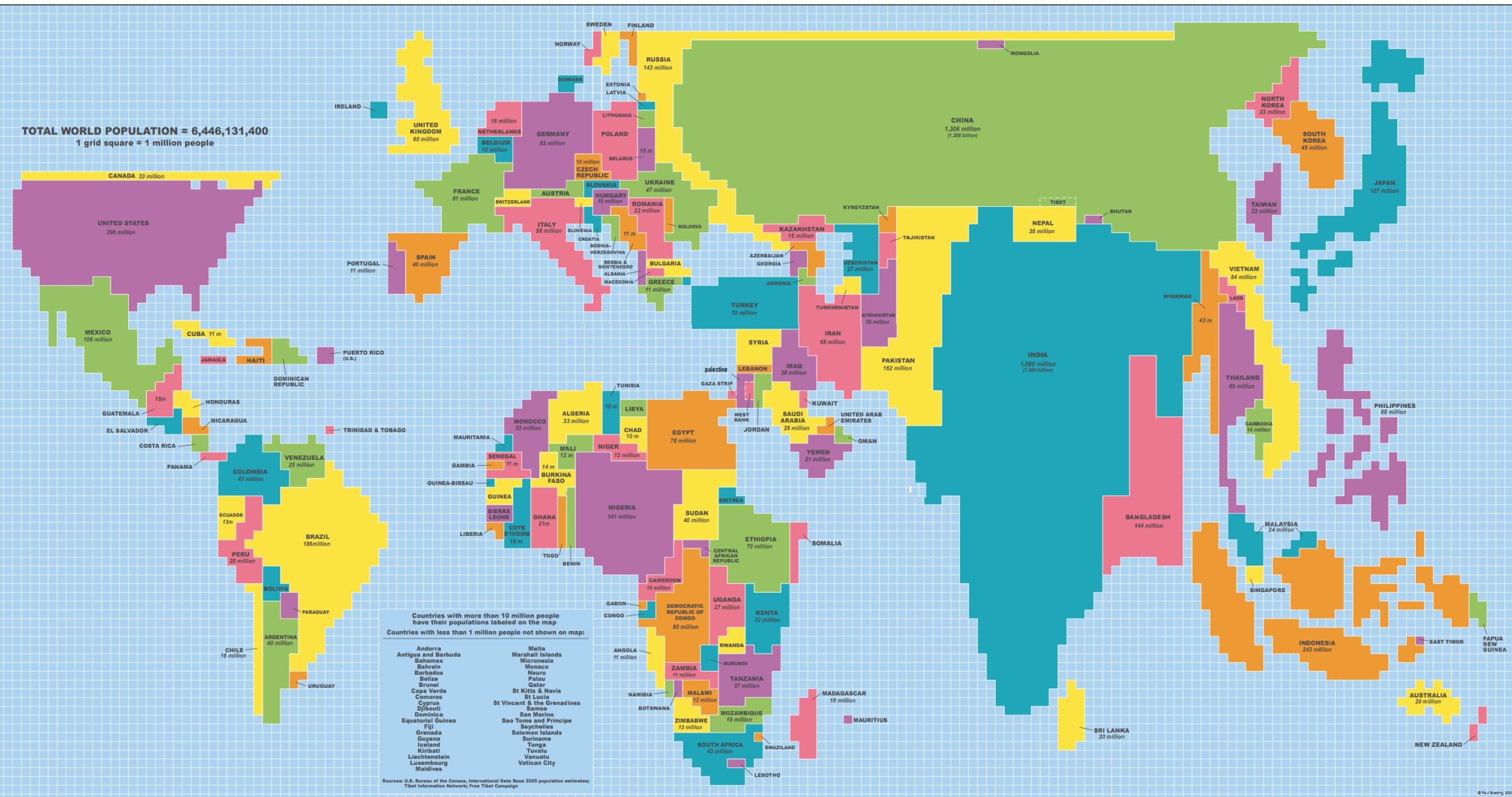


# CARTES GÉOGRAPHIQUES

## Global warming culprits, judged by size

Countries that have caused disproportionately more global warming than their area would suggest are shown swollen, while low-emitters in relation to their size are shrunken





# CARTES

La plupart des gens ont l'habitude d'utiliser des cartes géographiques, elles sont donc généralement faciles à interpréter.

Ces cartes peuvent avoir un effet marquant quand la visualisation des données produit un **résultat inattendu**

- qui masque des renseignements importants
- ou qui correspond à un manque de renseignements importants
- ou qui change la perspective du lecteur.

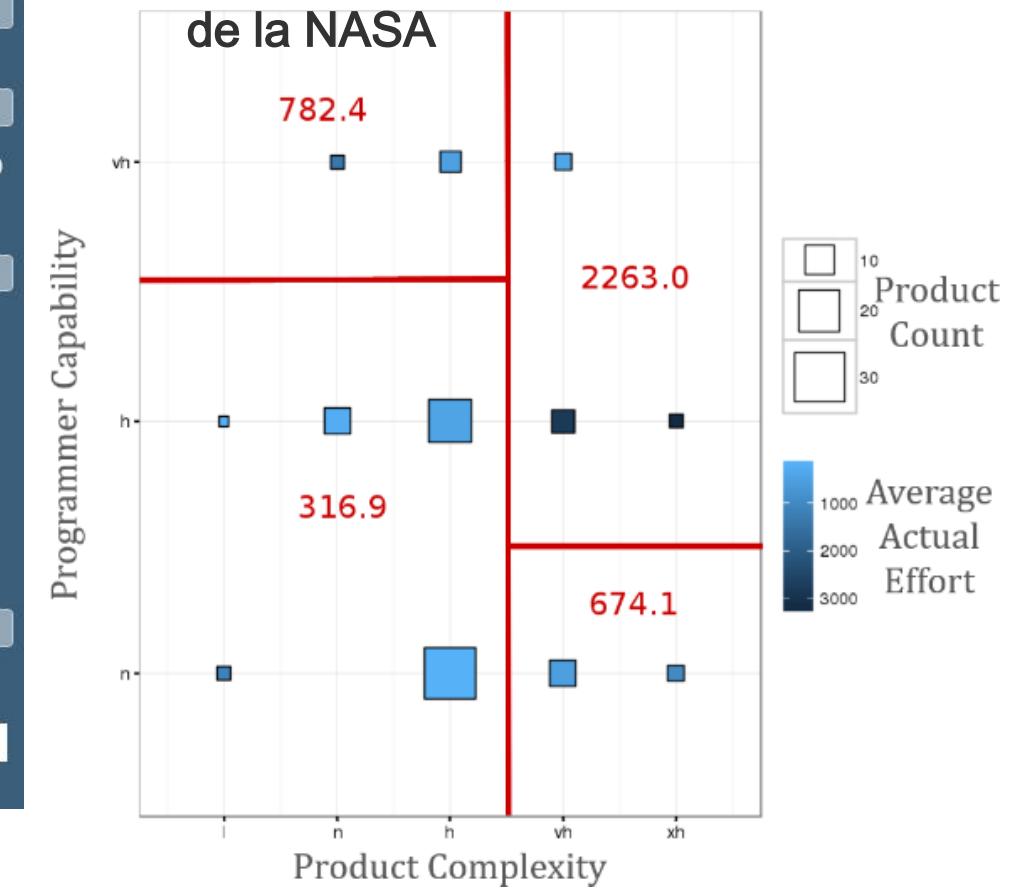


# GRAPHIQUE À BULLES

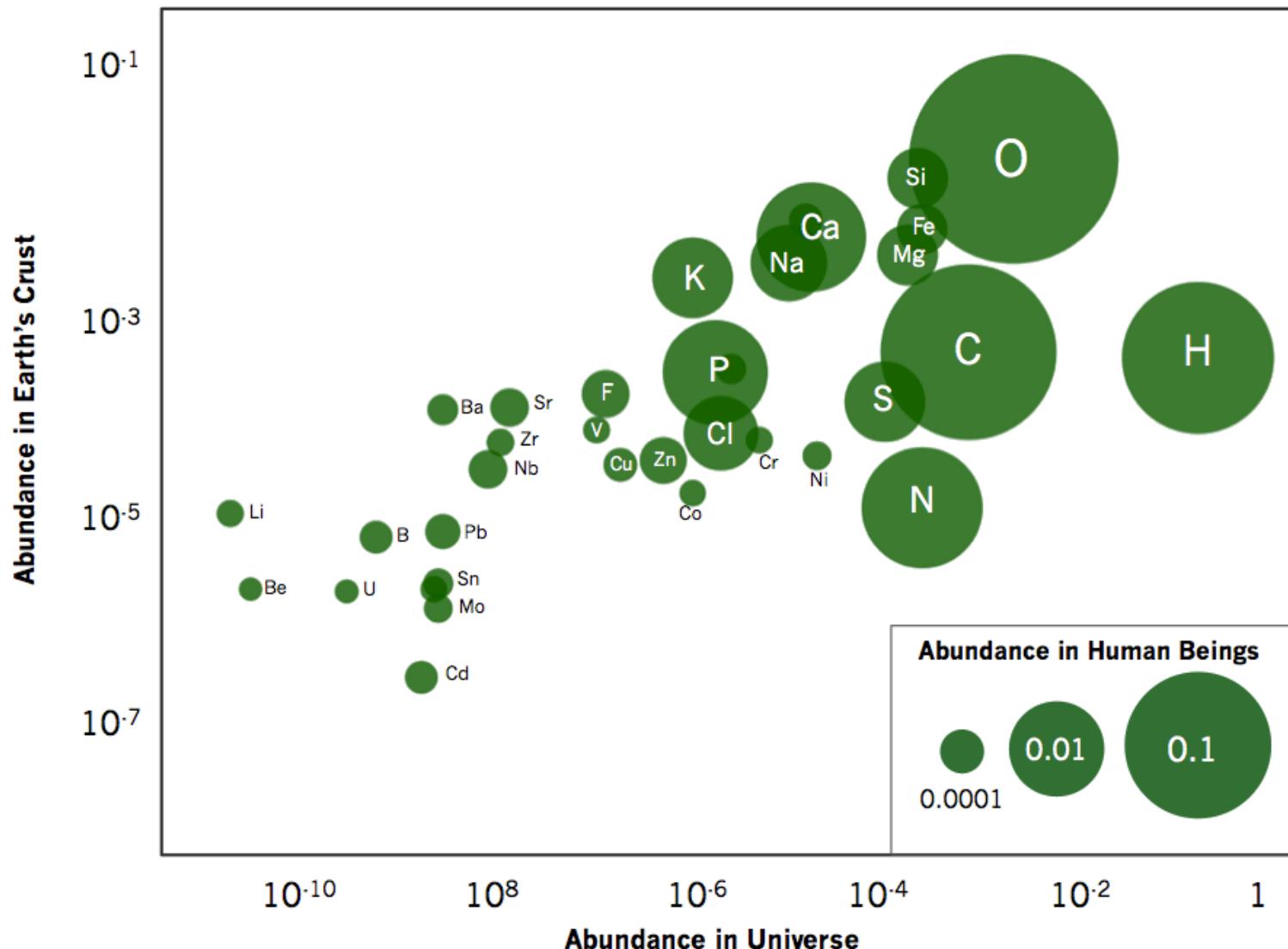


Utilisation des joueurs (Sénateurs d'Ottawa)

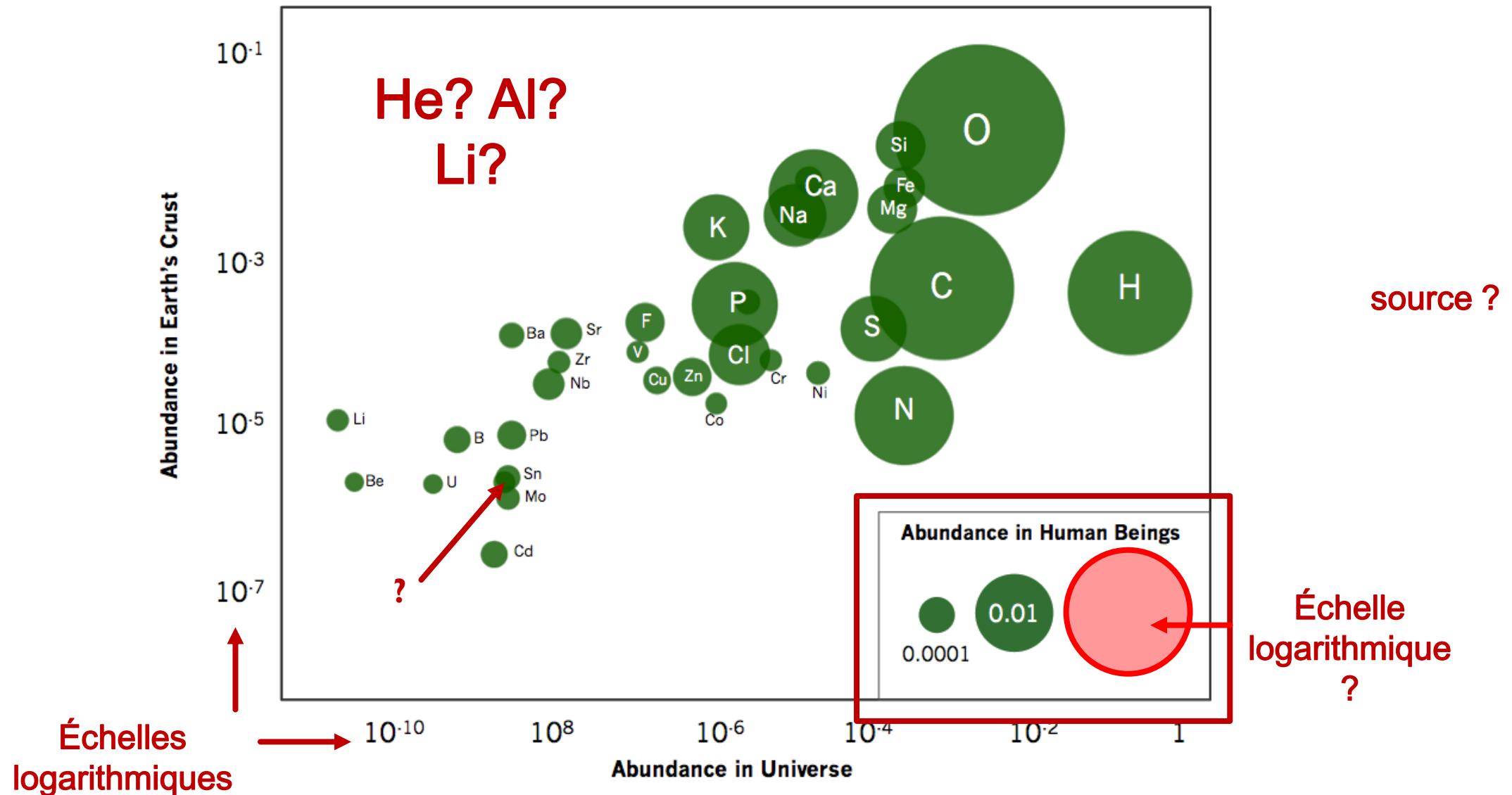
## Jeu de données COCOMO de la NASA



# Abondance des éléments chimiques



# Abondance des éléments chimiques



## GRAPHIQUE À BULLES

**La couleur et la géométrie** nous permettent de représenter (au moins) deux variables additionnelles sur un nuage à point en deux dimensions.

Il peut être nécessaire de revoir l'échelle ou de compartimenter les données.

L'utilisation d'une vidéo permet d'ajouter une variable ordinaire.

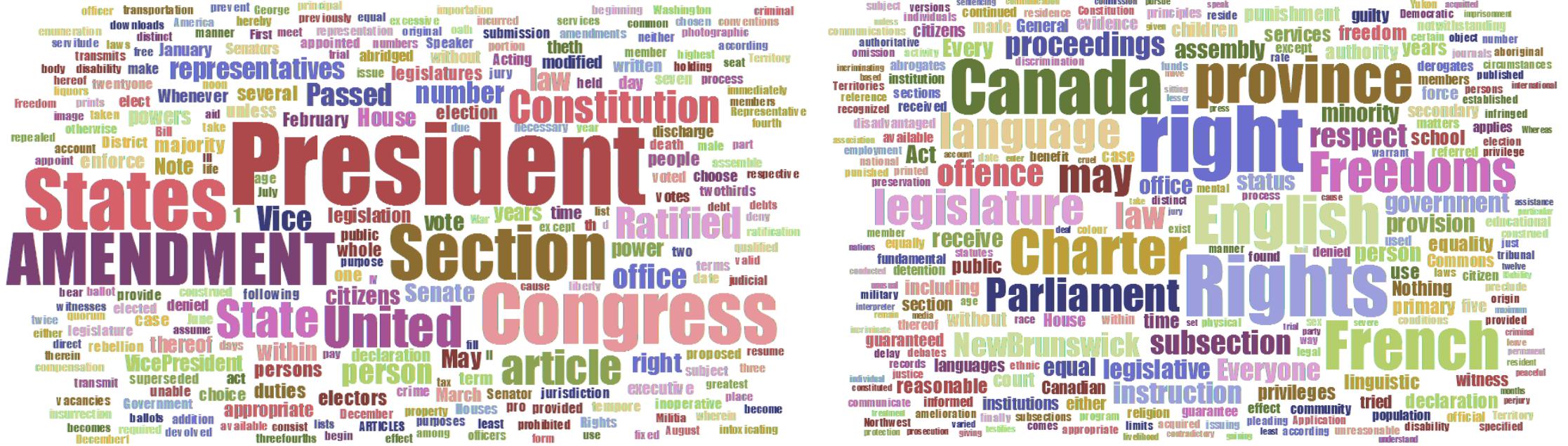
**Il est aussi possible d'ajouter du texte** pour représenter une variable nominale additionnelle.

À son efficacité maximale quand le graphique **n'est pas trop chargé**.

Un de mes **graphiques préférés** – bon équilibre entre caractéristiques modernes et traditionnelles.

# NUAGE DE MOTS

## Constitution des É.-U. et Charte canadienne des droits et libertés



## NUAGE DE MOTS

Pour une efficacité maximale, la taille de police devrait dépendre de la fréquence.

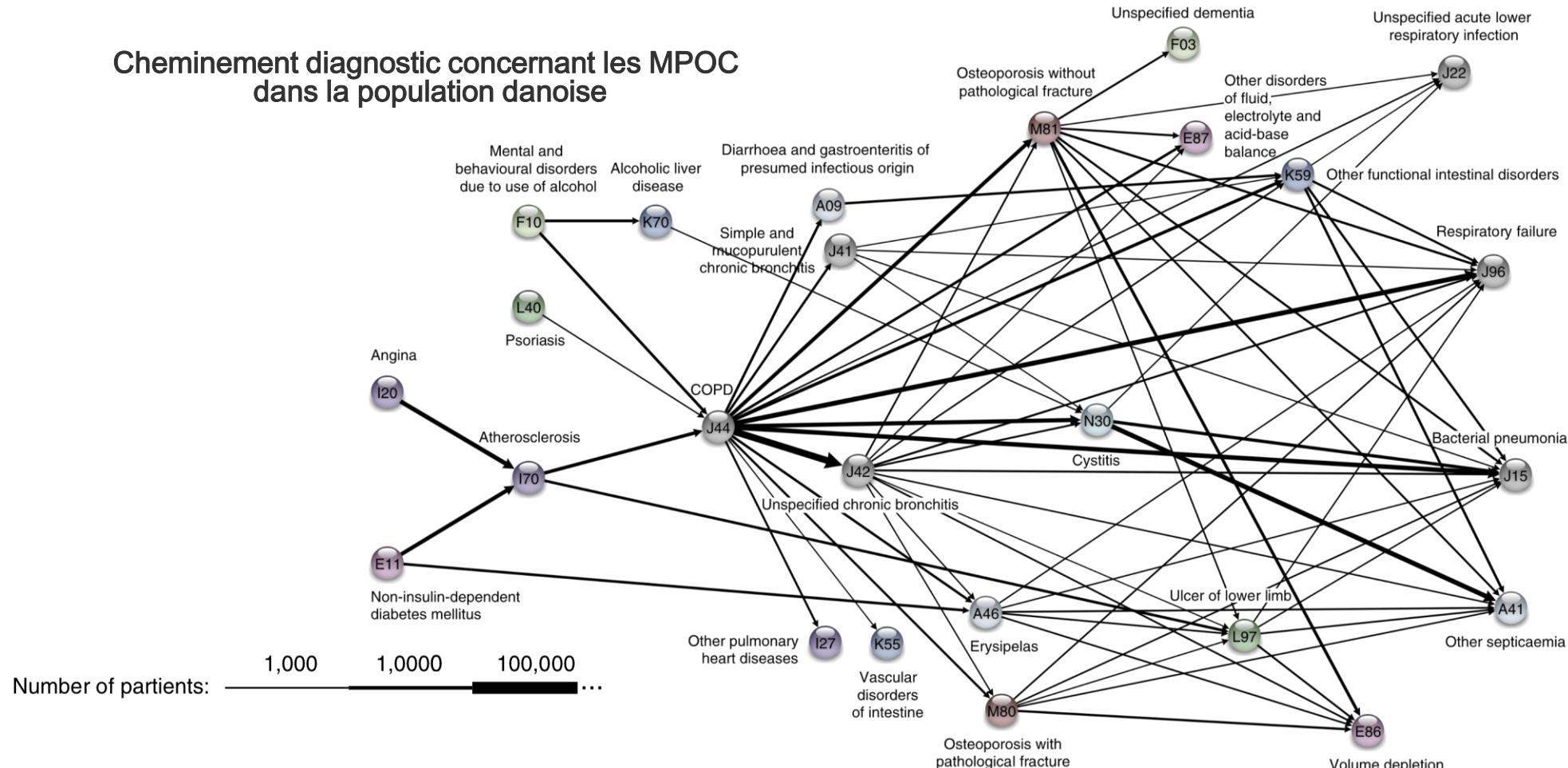
Généralement utilisé pour représenter des données nominales à une variable, mais il est possible d'intégrer d'autres variables en utilisant des **miniatures**, la **taille** des nuages, le **positionnement** des mots, la **couleur** et la **tonalité chromatique**.

L'algorithme de positionnement et de choix de couleurs est « secret ».

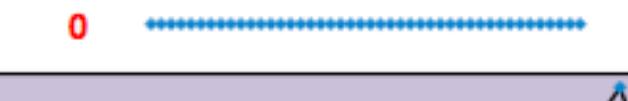
Cette caractéristique peut servir de preuve d'identité du créateur.

# DIAGRAMME DE RÉSEAU

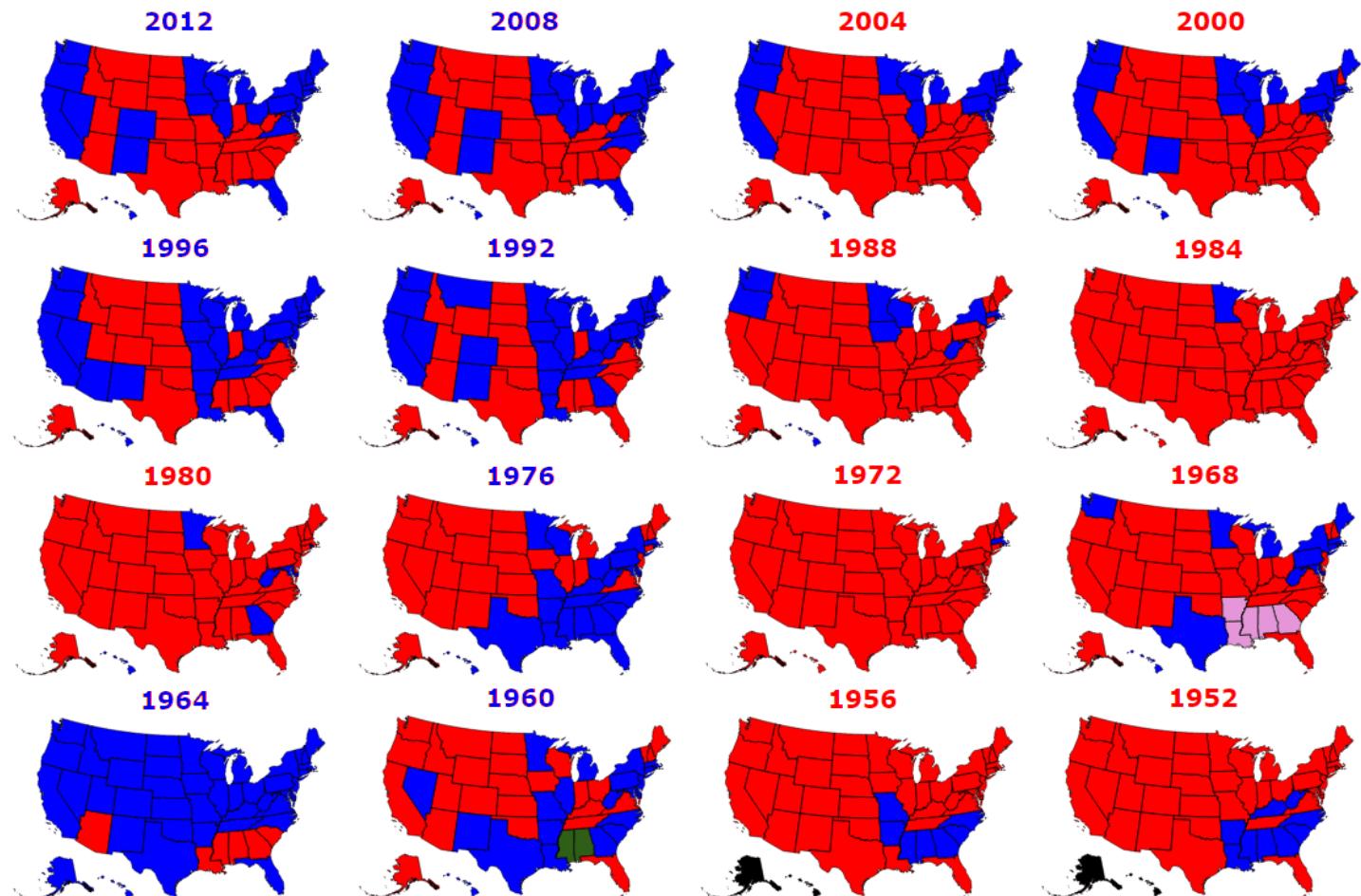
**Cheminement diagnostic concernant les MPOC dans la population danoise**



# GRAPHIQUE DE TENDANCE ET MINIATURES

	Start	Monthly Number of Cases	End	Low	High	Mean	Std Dev	Blanks	Zeros	Trend
Total	19502		17265	15150	25072	19903	2612	0.0	0.0	379.2
Hôpital n° 1	46		19	3	46	19	9	0.0	0.0	-1.6
Hôpital n° 2	156		240	101	326	194	60	0.0	0.0	9.7
Hôpital n° 3	16		11	2	76	15	15	0.0	0.0	-2.9
Hôpital n° 4	3		13	0	105	9	15	0.0	0.4	-1.8
Hôpital n° 5	42		50	25	91	61	16	0.0	0.0	1.2
Hôpital n° 6	48		53	34	169	67	25	0.0	0.0	0.6
Hôpital n° 7	0		N.A.	0	0	0	0	2.2	9.8	0.0
Hôpital n° 8	56		104	34	150	73	25	0.0	0.0	4.6

# MINIATURES



# VISUALISATIONS INTERACTIVES ET ANIMÉES

Une visualisation animée **n'est pas toujours** la meilleure solution.

Quelles renseignements peuvent être transmis grâce à l'interactivité? Tout dépend des données et de la visualisation.

## Exemples :

- The Clubs That Connect the World Cup, NY Times, 2014
- Who Marries Whom, Bloomberg, 2016
- HipparcosStar Mapper, European Space Agency, 2016
- The Internet of Things – a Primer, Information is Beautiful, 2016
- The Genealogy and History of Popular Music Genres, Musicmap, 2016

# VISUALISATIONS INTERACTIVES ET ANIMÉES

## Exemples (suite) :

- [Sequences Sunburst](#), Kerry Rodden, 2015
- [Health and Wealth of Nations](#), Gapminder Foundation
- [Mobius Transformations Revealed](#), Arnold D.N, Rogness, J, 2007
- [Visualizing the Riemann Function and Analytic Continuation](#), 3Blue1Brown, 2016
- [Small Arms and Ammunition – Imports and Exports](#), Google, 2012
- [The Evolution of the Web](#), Google, Hyperakt, Vizzuality, 2012
- [peoplemovin](#), Carlo Zapponi, 2012

## DISCUSSION ET POINT À RETENIR

« Si certains types de visualisation deviennent dominant, il y a toujours un risque que les questions qui se prêtent particulièrement bien à ces visualisations deviennent dominantes, ce qui aurait une incidence sur les méthodes de collecte de données, l'accessibilité des données, le choix des pistes de recherche et ainsi de suite. » [TRADUCTION]

(P. Boily)

### À retenir :

- Explorer les données
- Essayer plusieurs méthodes

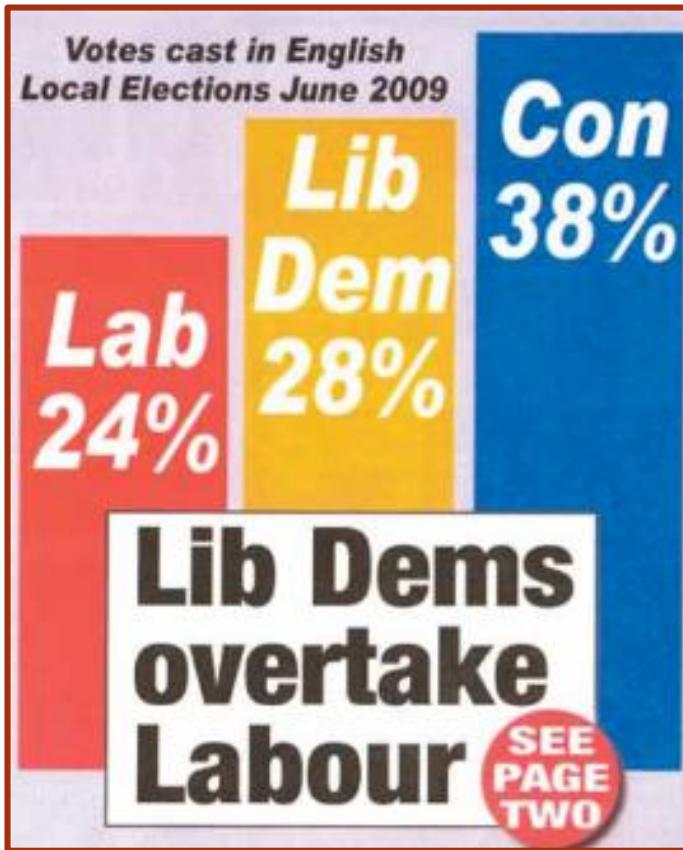
# TABLEAU D'HONNEUR ET TABLEAU D'HORREUR

EXPLORATION ET VISUALISATION DES DONNÉES

## GRAPHIQUES TROMPEURS



## GRAPHIQUES TROMPEURS



# GRAPHIQUES TROMPEURS

**Problèmes :** information fallacieuse, sélective ou traitée de façon incompétente.

**Solutions :**

- Échelles et unités de mesure uniformes
- Séries chronologiques complètes
- Ne pas choisir arbitrairement la fourchette de données
- La troncation d'un axe peut exagérer certains effets
- Les nombres doivent être sensés

## À SURVEILLER

Certaines méthodes produisent des graphiques impressionnantes, mais trompeurs.

Se méfier :

- **de la manipulation des axes et des échelles linéaires;**
- **des effets d'échelle**, lorsque des données sont représentées par des formes ou des volumes;
- **des choix arbitraires** permettant d'omettre certaines observations.

Pour les jeux de données dont le nombre de dimensions est réduit, un **tableau** peut être aussi informatif et comporter moins de risque de mésinterprétation.

## À SURVEILLER

Différentes manières d'évaluer le caractère trompeur d'un graphique :

- **Facteur de mensonge** : rapport entre la taille de l'effet affichée dans le graphique et la taille de l'effet dans les données.
- **Densité des données** : rapport entre le nombre d'observations et la superficie du graphique.
- **Rapport de bric-à-brac graphique** : rapport entre la superficie nécessaire pour transmettre l'information et la superficie du graphique.

On souhaitera habituellement que le facteur de mensonge et le rapport de bric—à—brac graphique se rapprochent autant que possible de 1, tandis que la densité des données devrait être « élevée » (dans la limite du raisonnable).

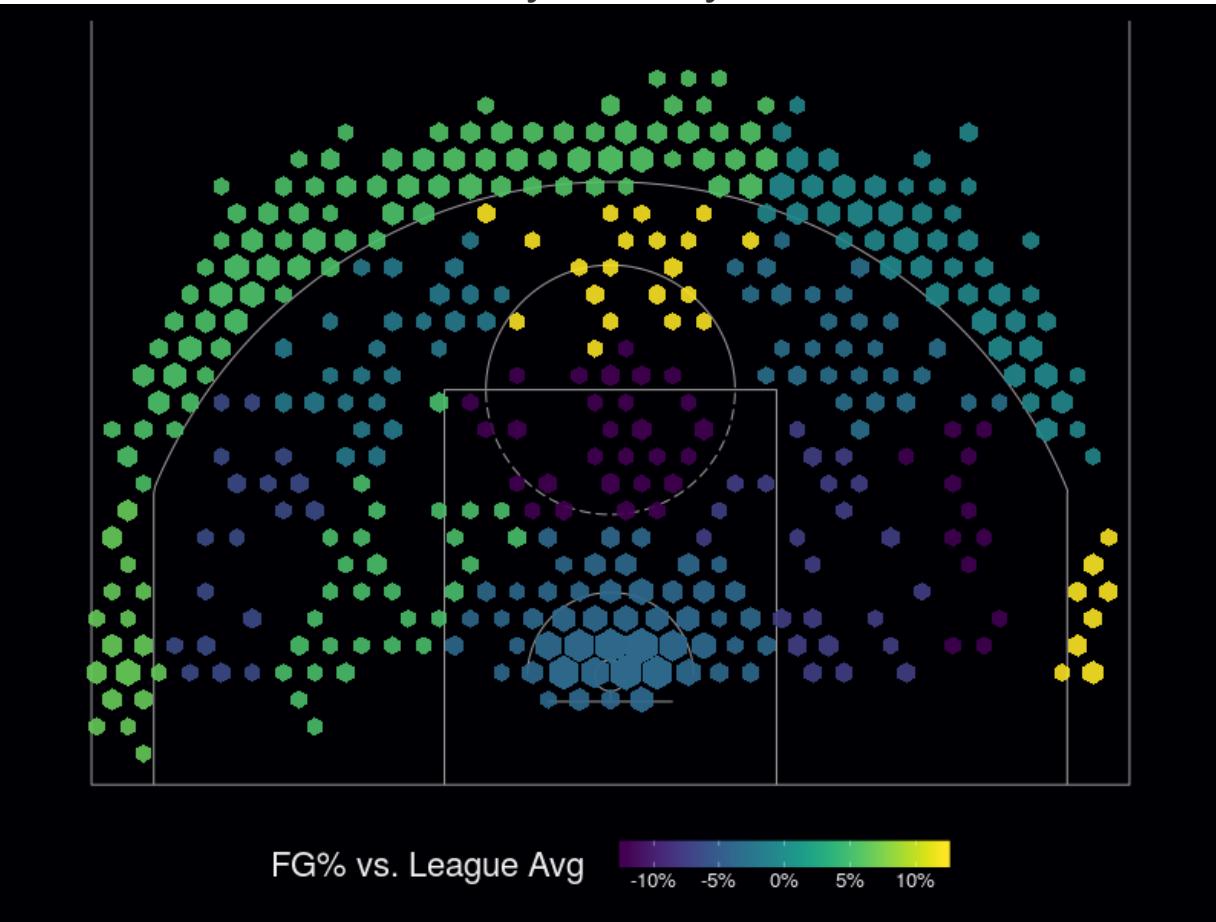
## À VOUS DE JUGER

Certaines des visualisations suivantes (peuvent être considérées comme) bonnes. Mais pas toutes!

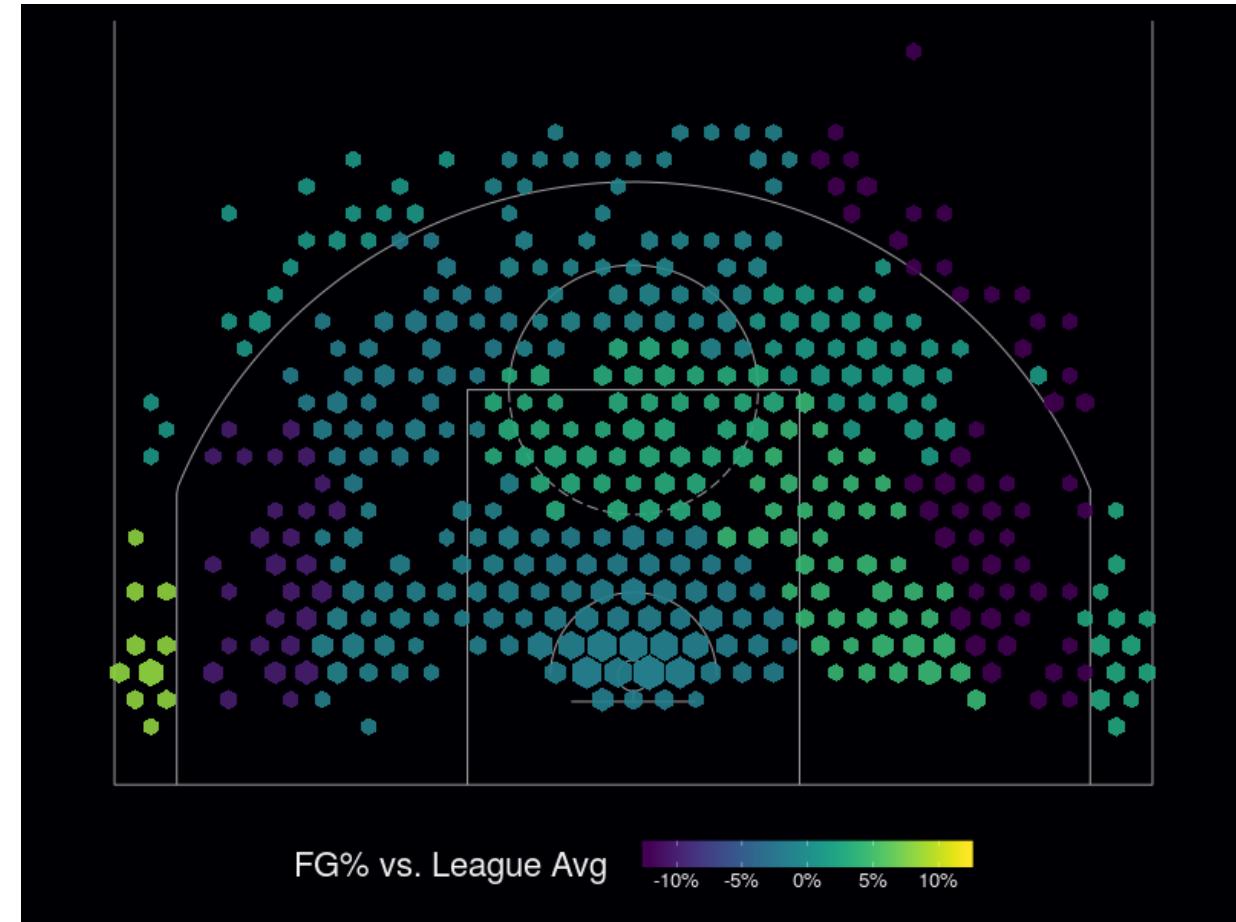
Dans quelle catégorie tombe chacune? À vous de juger...

# Paniers marqués (%) dans la NBA par rapport à la moyenne de la ligue (2015–2016)

Kyle Lowry

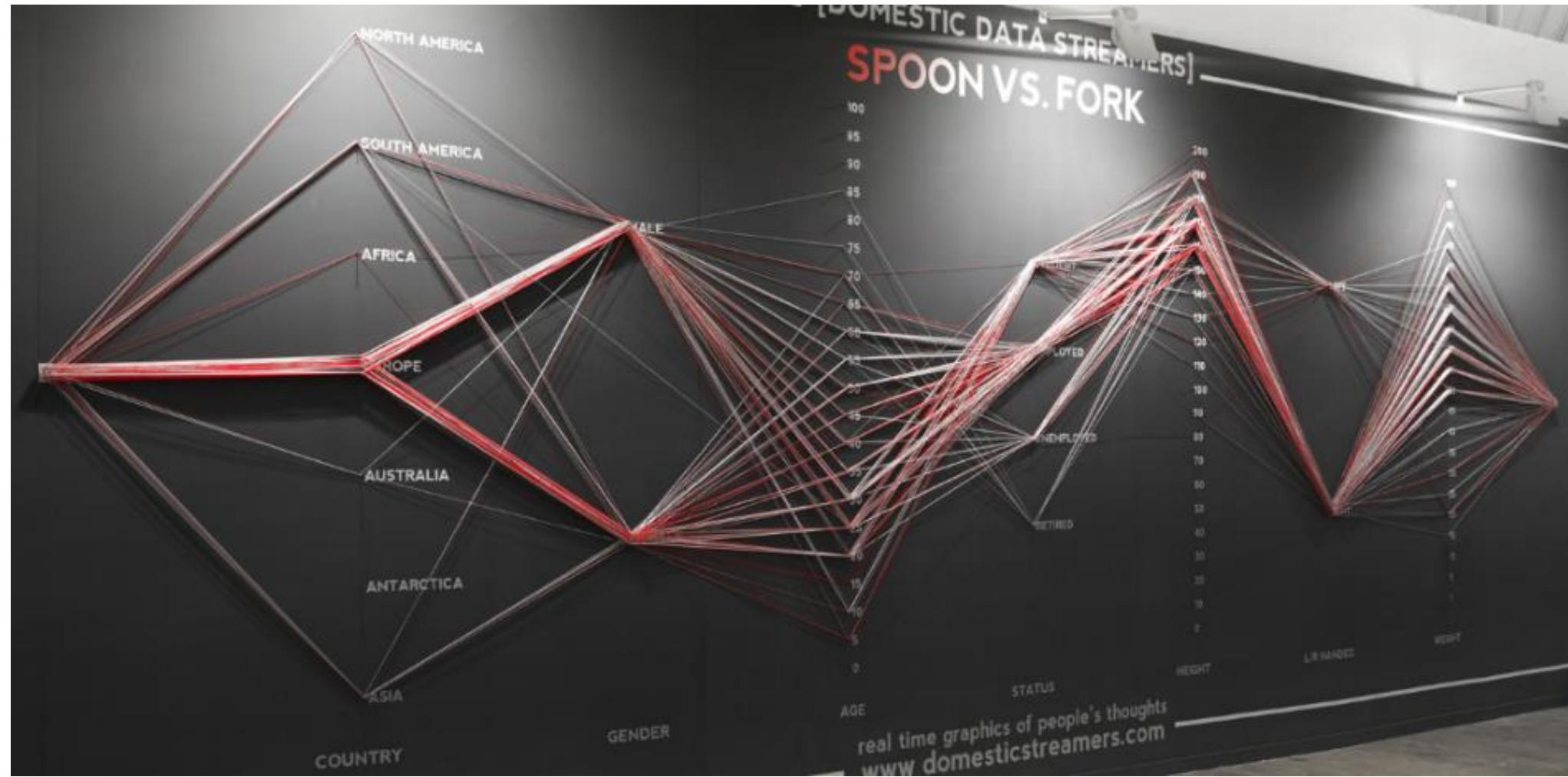


DeMar DeRozan



Quelles comparaisons pouvez-vous faire? Comprenez-vous l'encodage? Le contexte?

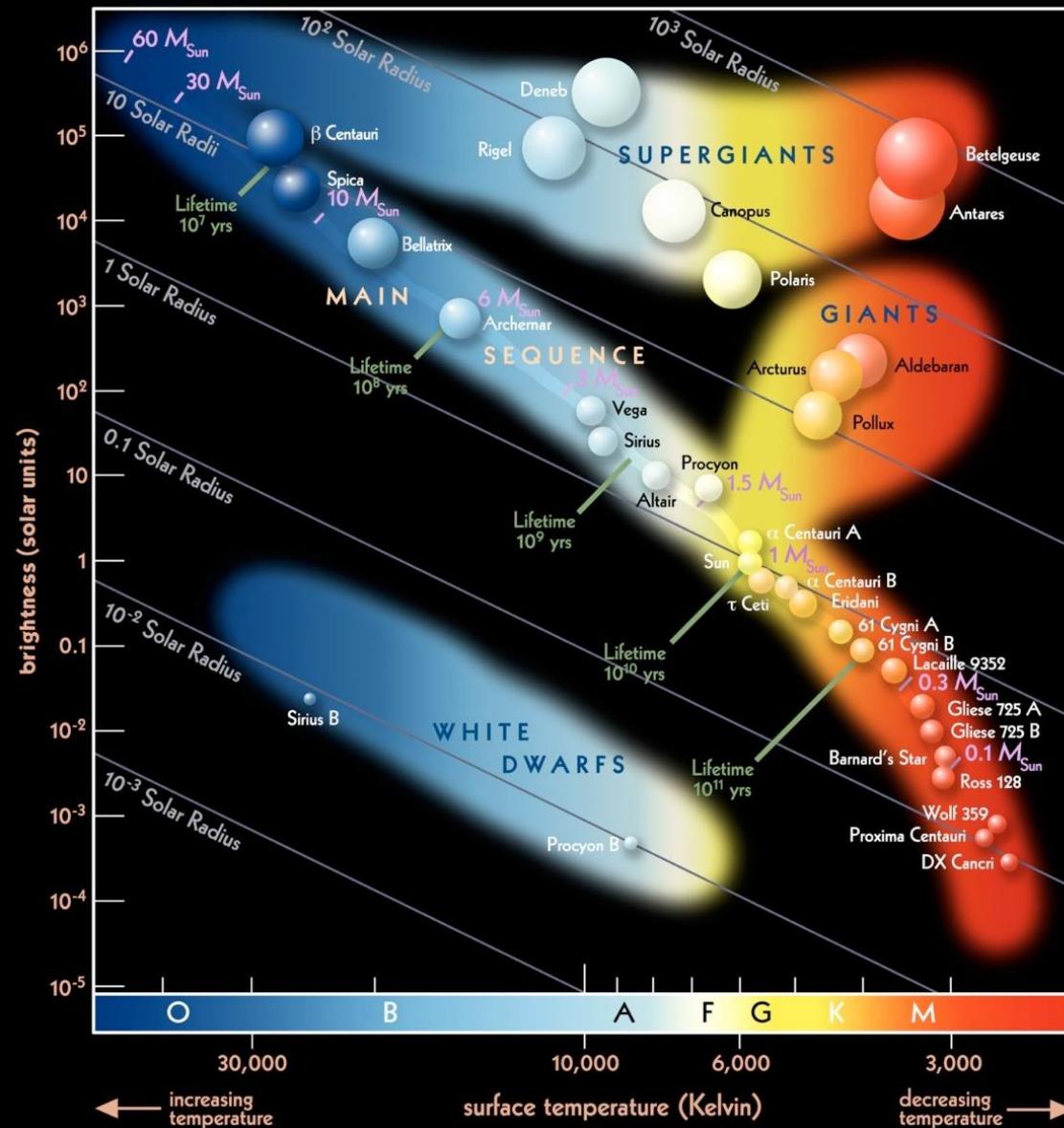
# Cuillère ou fourchette?



Existent-ils des problèmes de collecte de données? Où pensez-vous que cette activité s'est déroulée? La compétition « cuillère ou fourchette » est-elle une distraction?



# Diagramme Hertzsprung-Russell



## Éléments de données

- Rayon des étoiles (x 2)
- Température à la surface (x 2)
- Classe spectrale
- Luminosité
- Masse
- Durée de vie
- Nom

## Structure sous-jacente

- 4 regroupements
- La durée de vie, la masse et le rayon sont liés à la luminosité et à la température à la surface sur la séquence principale.

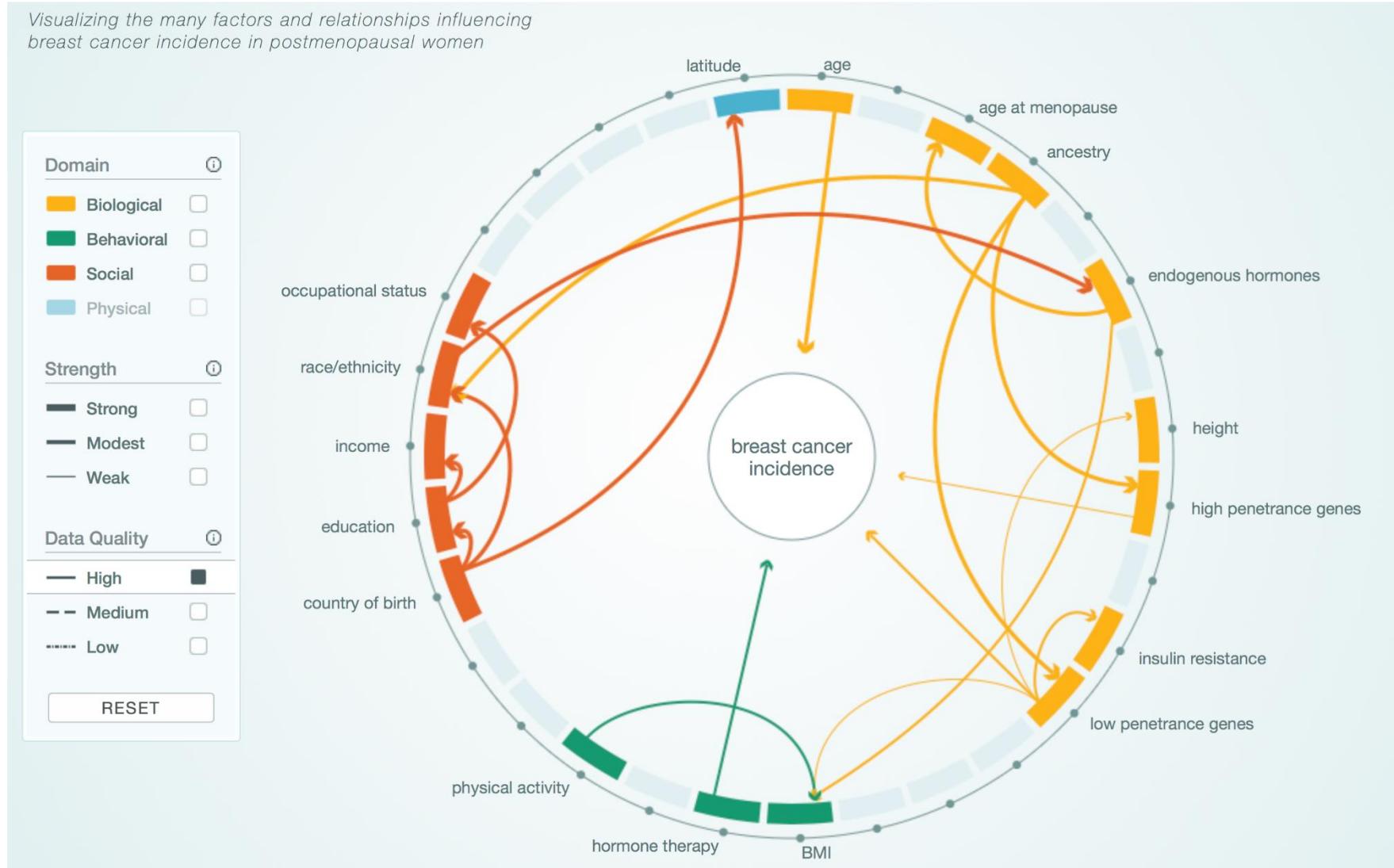
Le diagramme ne montre qu'un sous-ensemble d'étoiles.

# Modèle des causes du cancer du sein



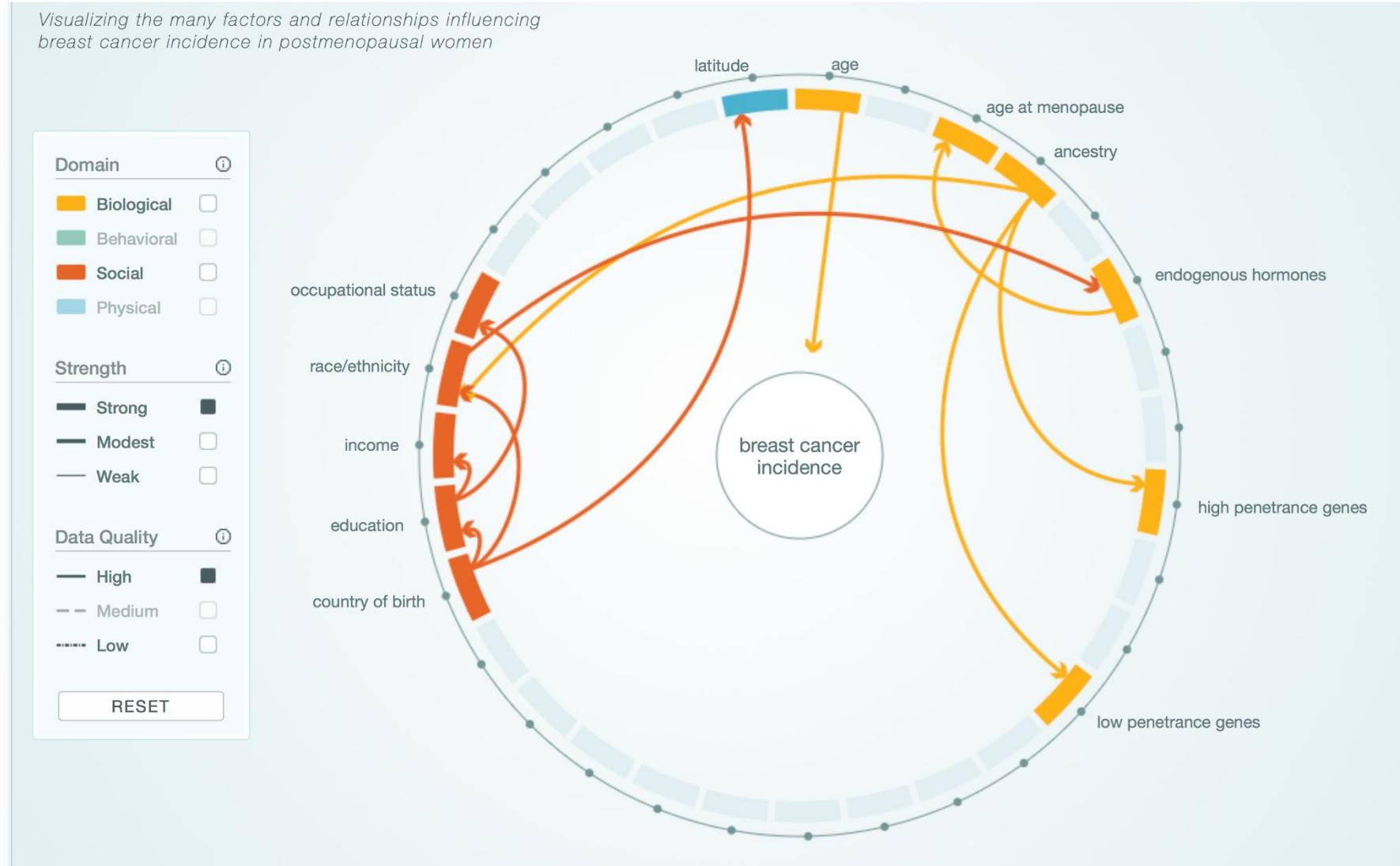
Pouvez-vous inférer une relation de cause à effet de ce diagramme?

# Modèle des causes du cancer du sein



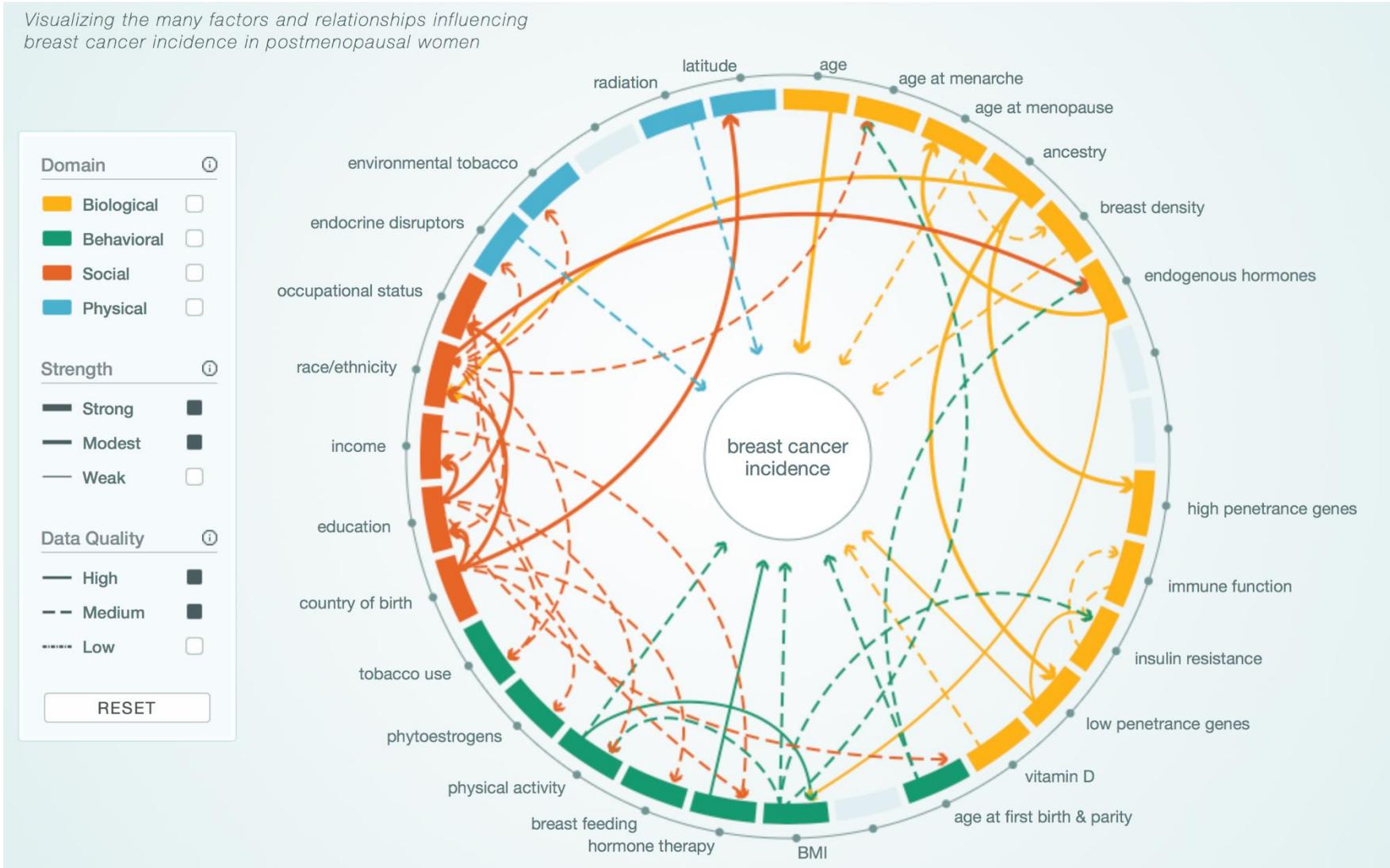
Pouvez-vous inférer une relation de cause à effet de ce diagramme?

# Modèle des causes du cancer du sein



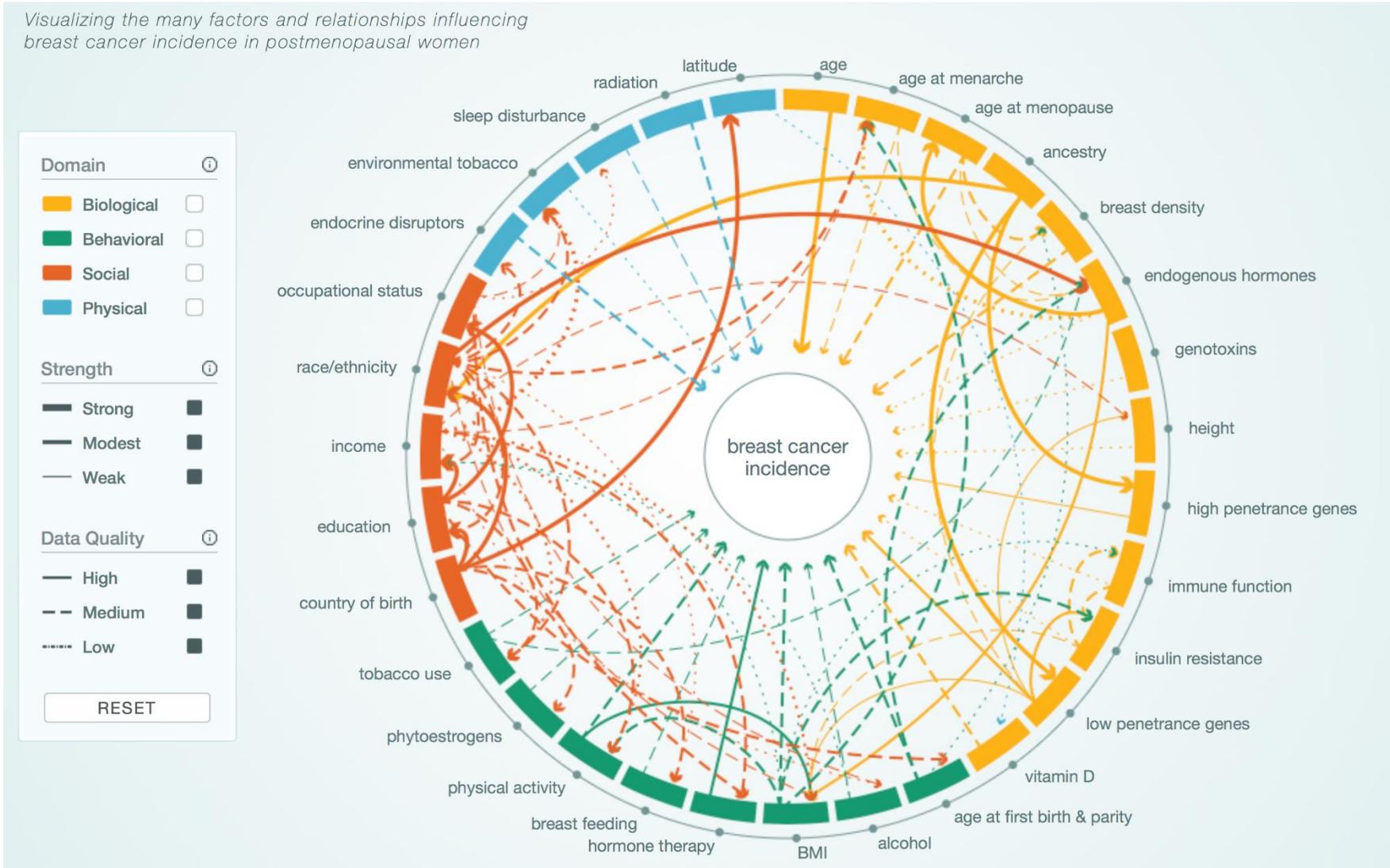
Pouvez-vous inférer une relation de cause à effet de ce diagramme?

# Modèle des causes du cancer du sein



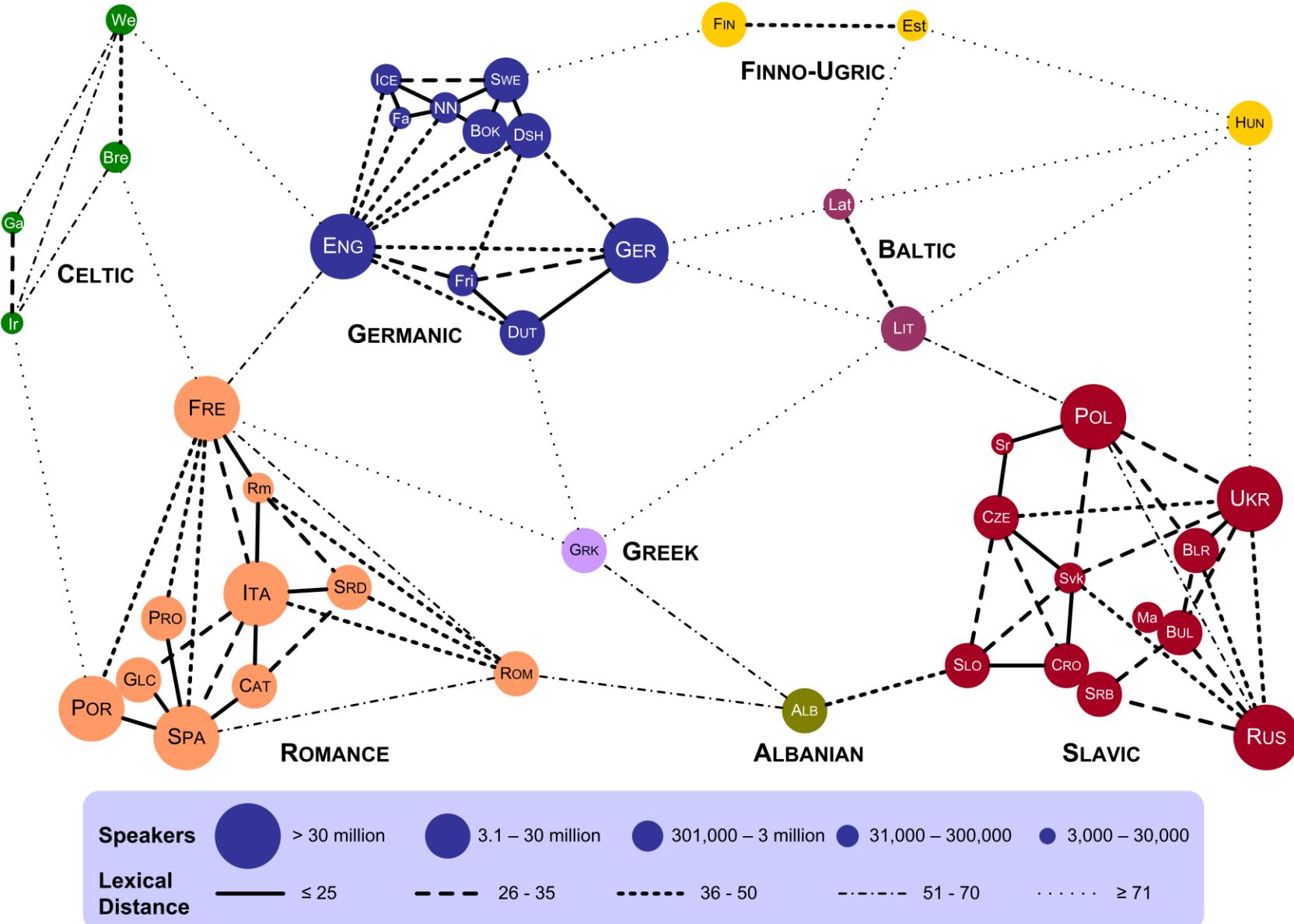
Pouvez-vous inférer une relation de cause à effet de ce diagramme?

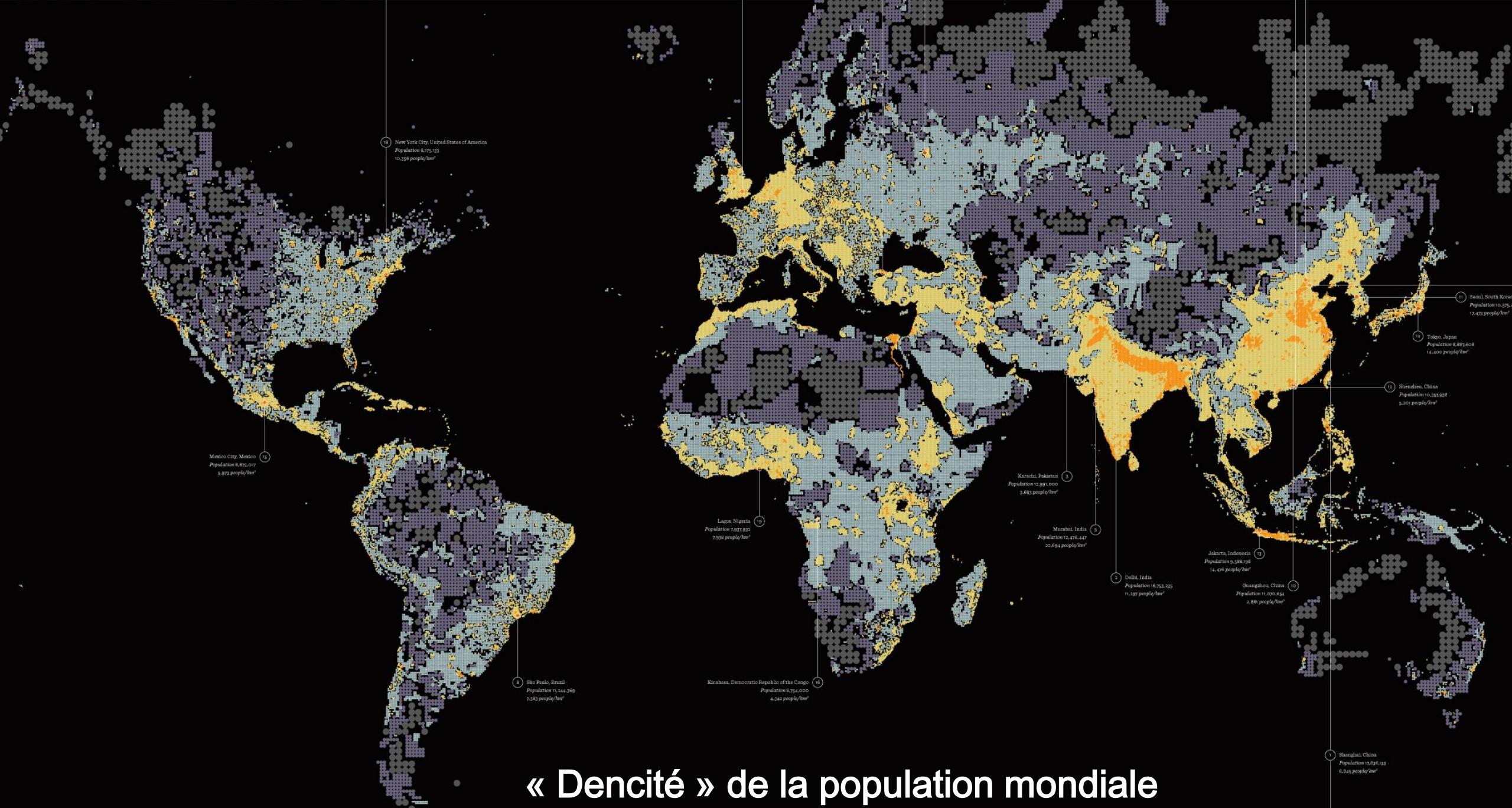
# Modèle des causes du cancer du sein



Pouvez-vous inférer une relation de cause à effet de ce diagramme?

# Distance lexicale entre les langues européennes





« Dencité » de la population mondiale

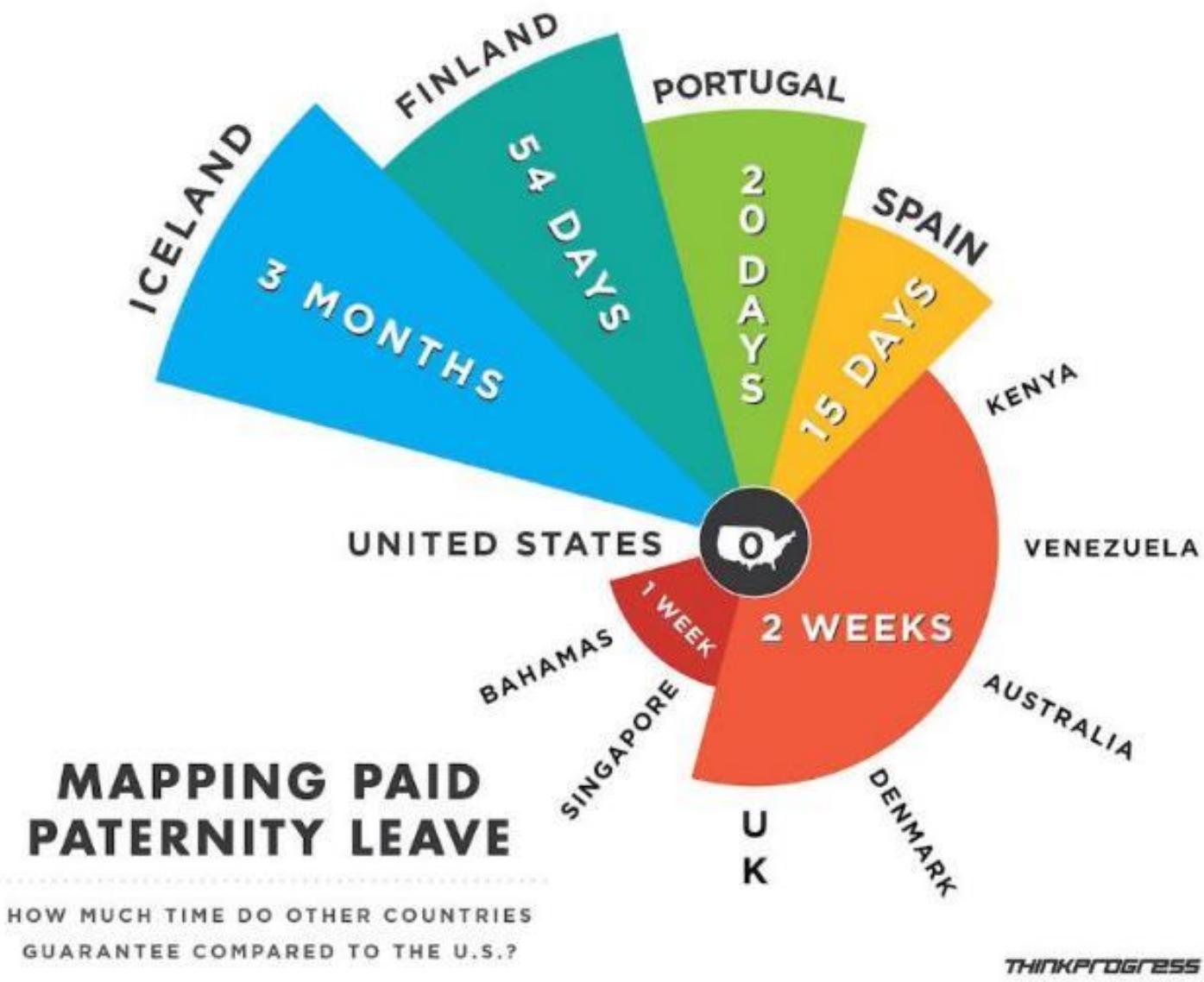
Faible densité de données

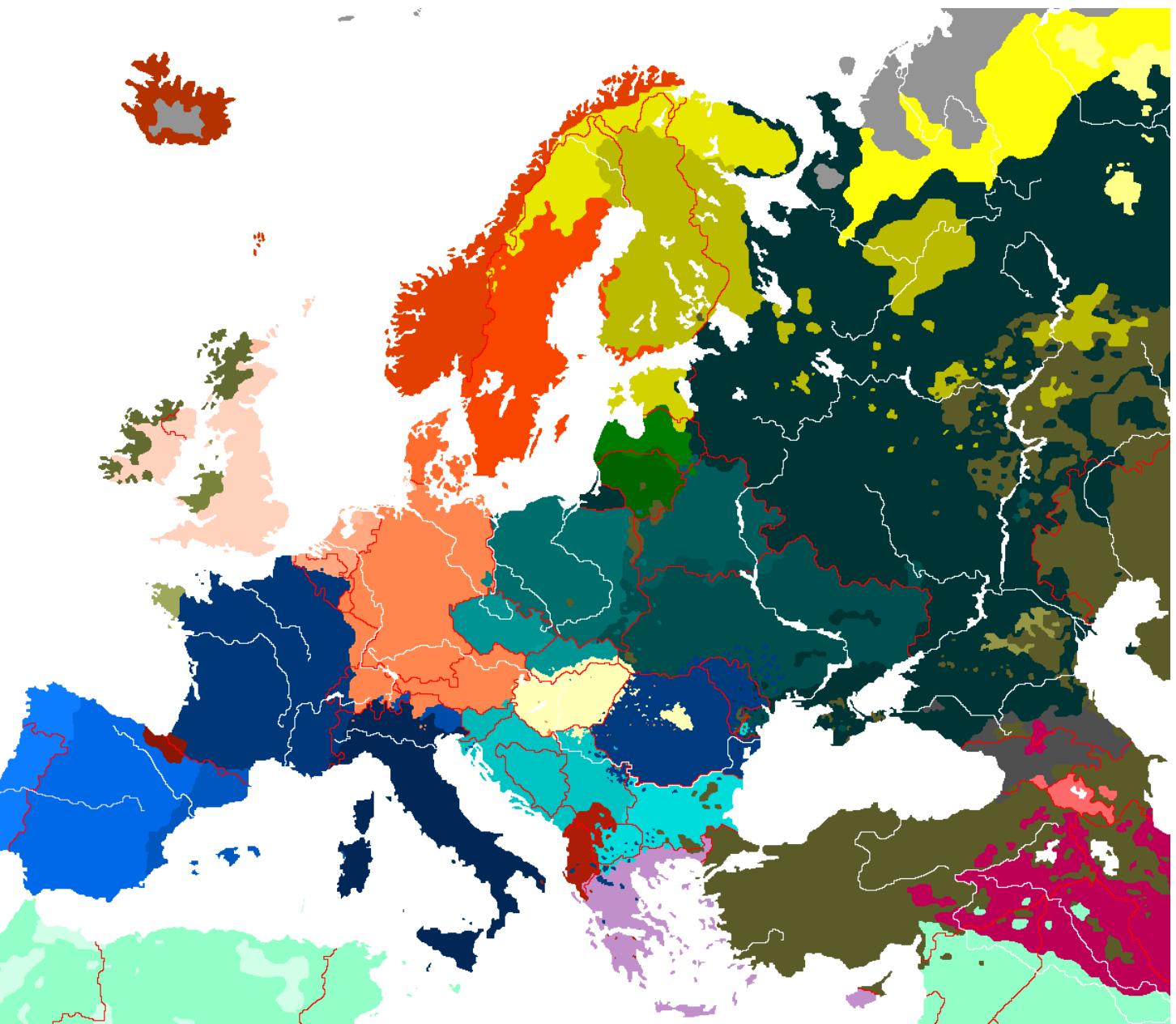
Rapport de bric-à-brac graphique élevé

Effets d'échelle

Choix de données arbitraires

Pourquoi ne pas plutôt utiliser un diagramme en bâtons ou un tableau?





Encodage?

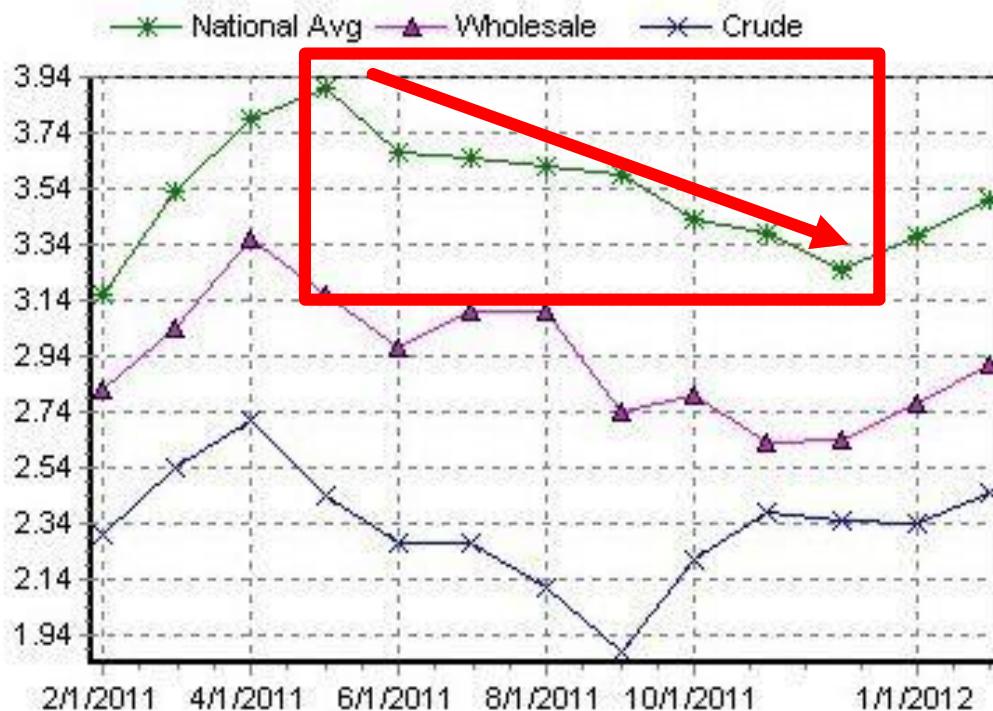
Densité de population?

Langues secondes?

Rivières?

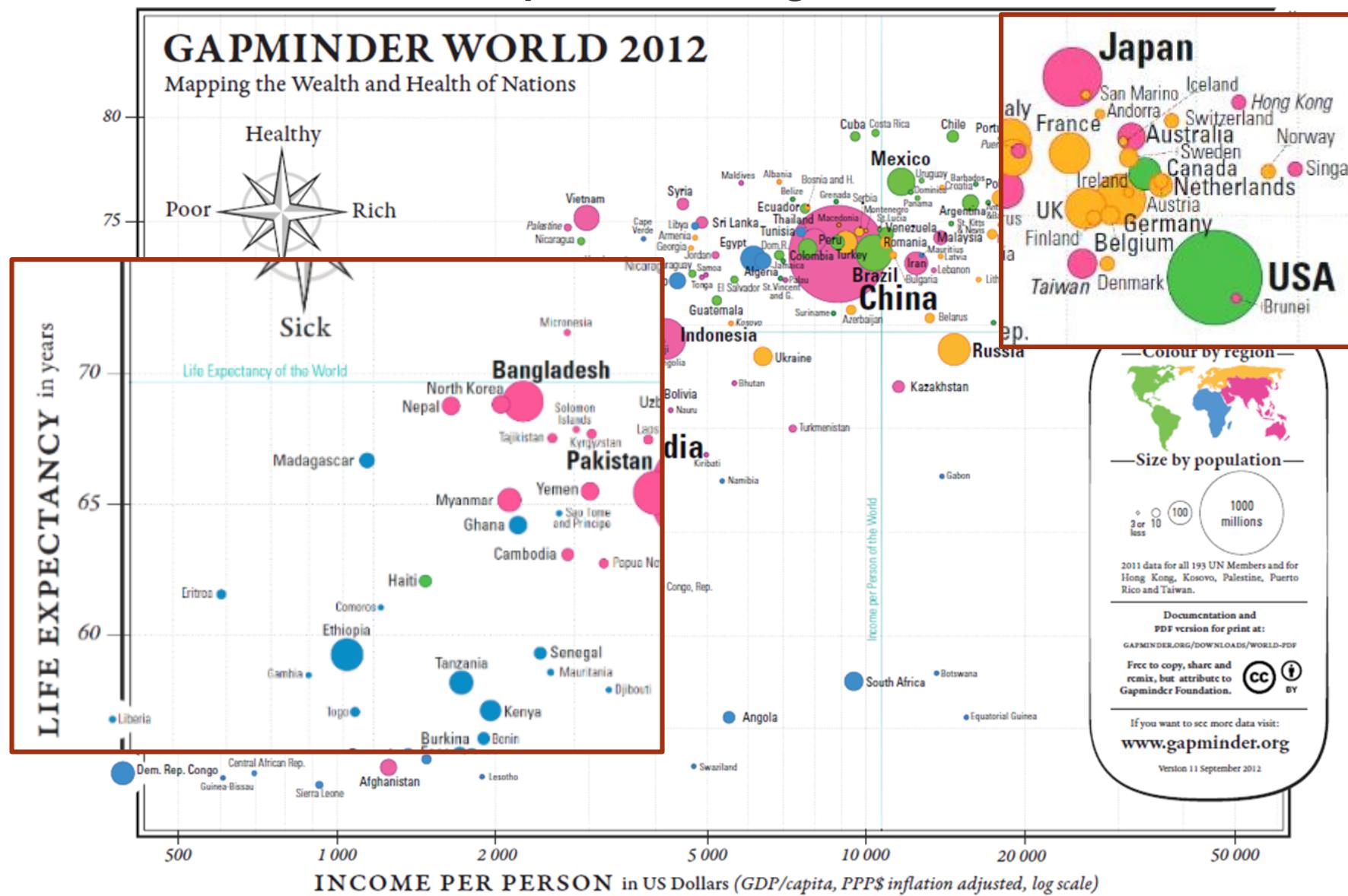
Aucune source

## 12 Month Average for Self-Serve Regular

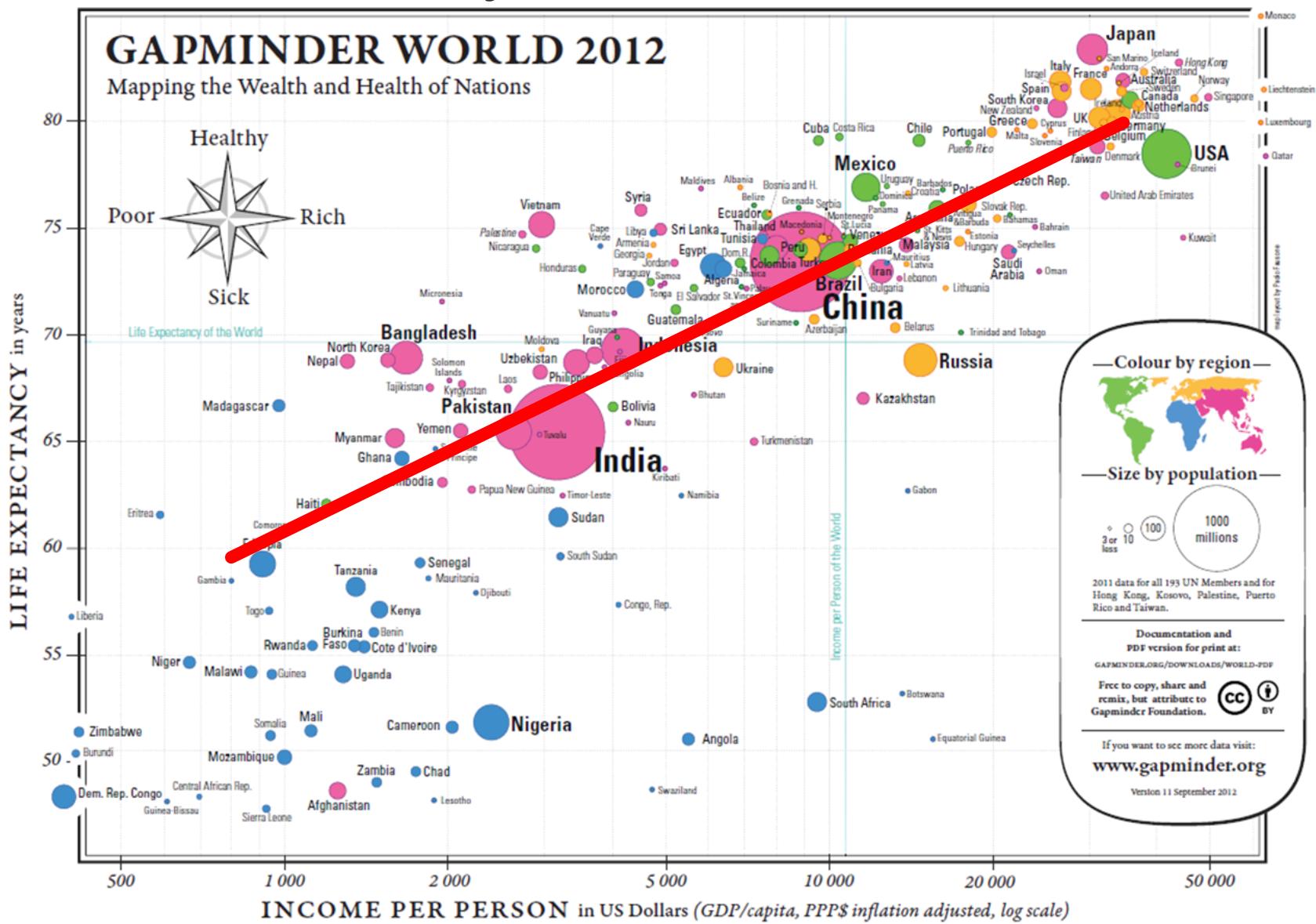


# MATÉRIEL COMPLÉMENTAIRE ET RÉPONSES AUX EXERCICES

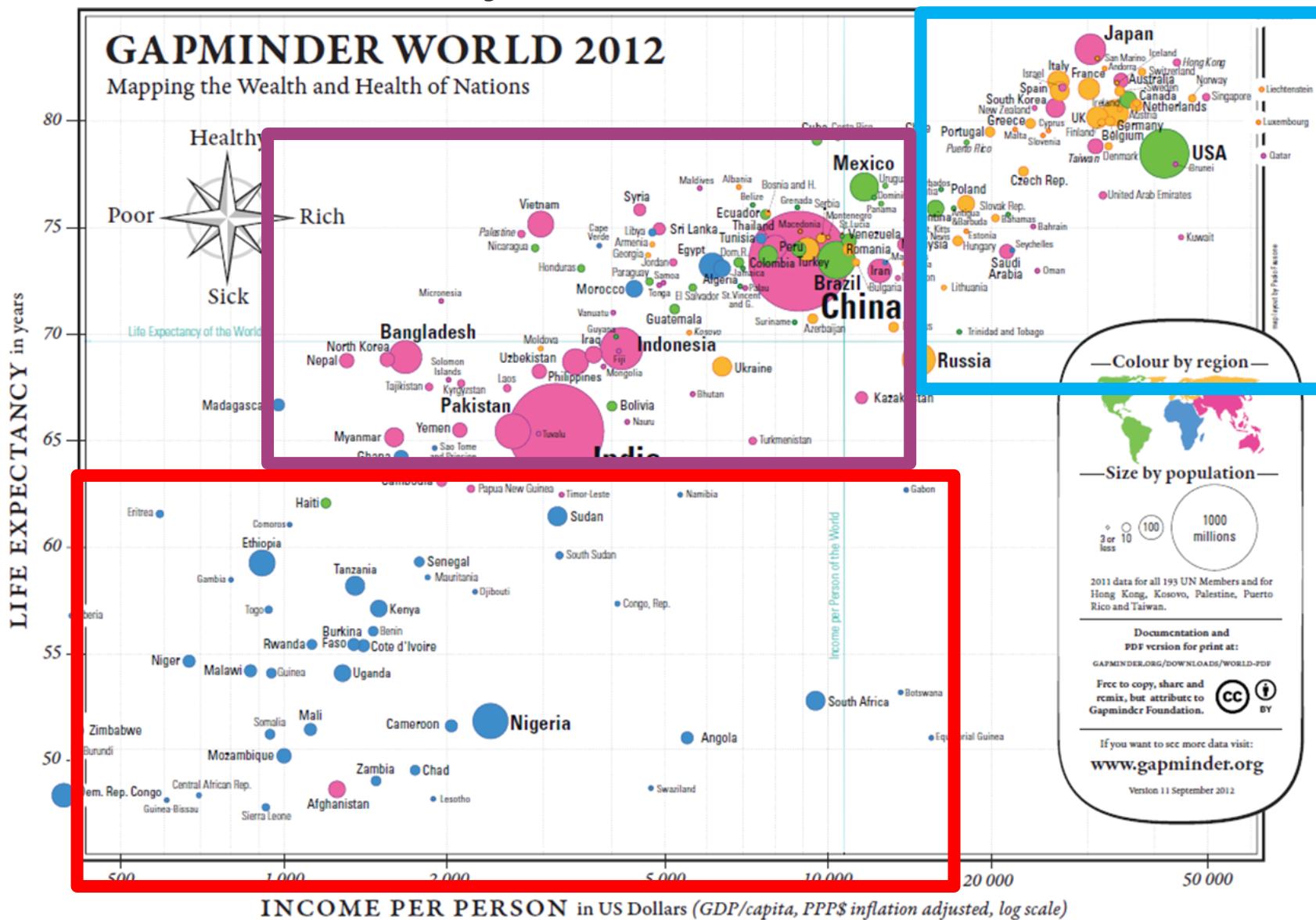
## Comparaisons significatives



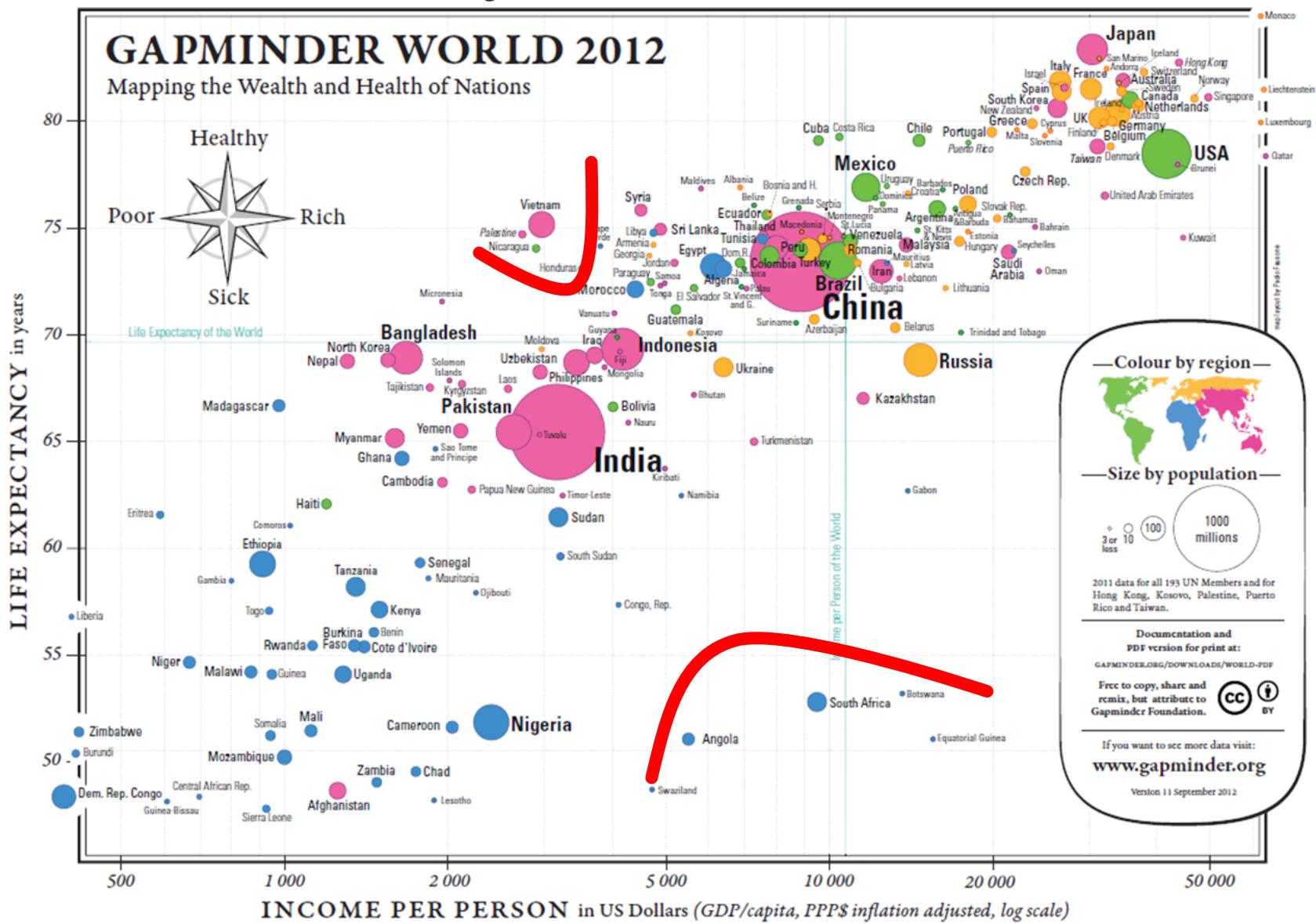
# Structure sous-jacente et relations multivariées



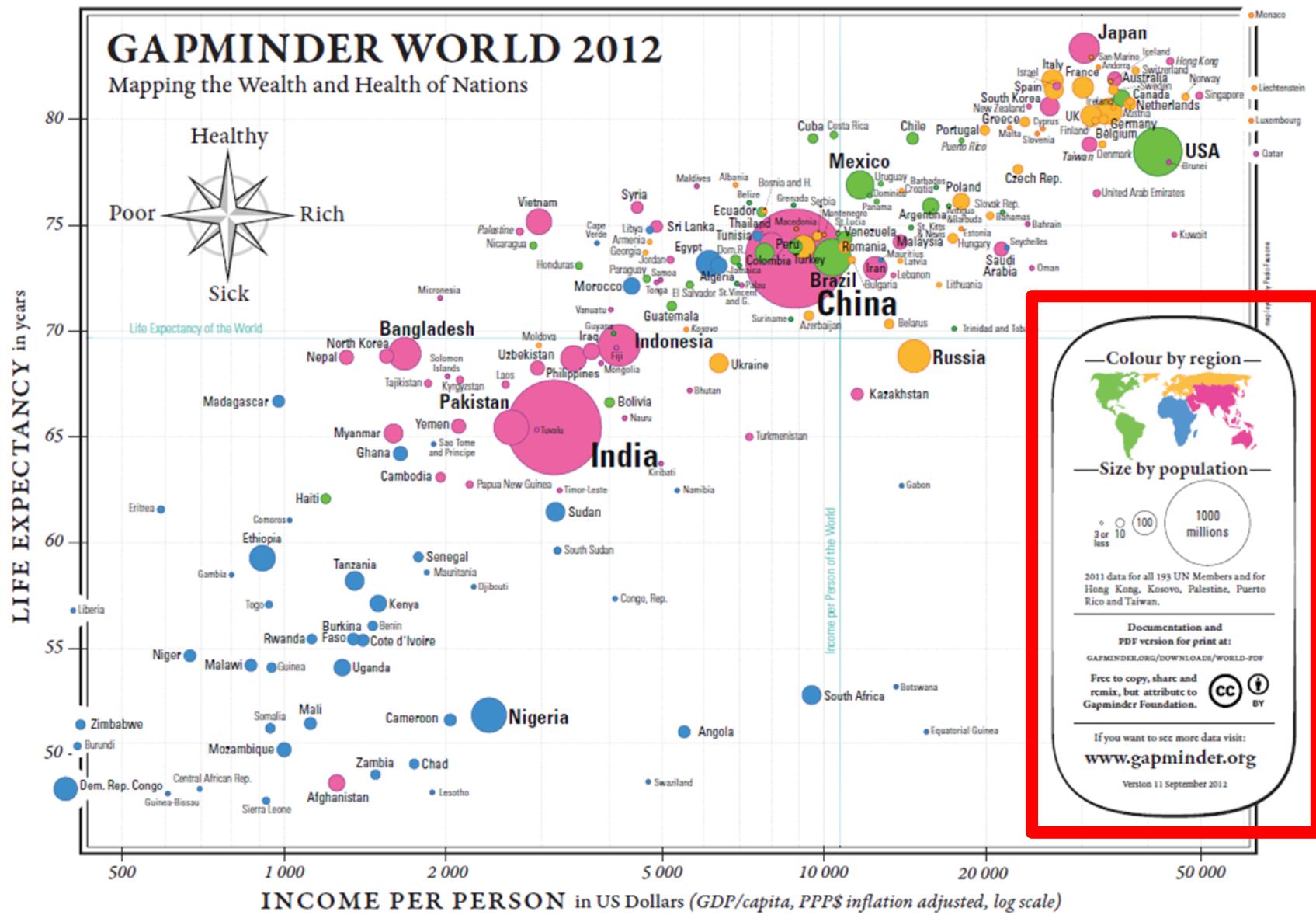
# Structure sous-jacente et relations multivariées



# Structure sous-jacente et relations multivariées



# Documentation



# GRAPHIQUES DU TABLEAU D'HONNEUR ET DU TABLEAU D'HORREUR

Voici notre avis – êtes-vous d'accord?

- Prix de l'essence : mauvais
- Les libéraux démocrates dépassent les travaillistes : mauvais
- Paniers marqués (%) dans la NBA par rapport à la moyenne de la ligue (2015–2016) : bon
- Cuillère ou fourchette : passable mais imparfait
- Diagramme Hertzprung-Russell : bon
- Modèle des causes du cancer du sein : bon
- Distance lexicale entre les langues européennes : passable mais imparfait
- « Dencité » de la population mondiale : bon
- Cartographie des congés de paternité payés : mauvais
- Carte mystère colorée : plutôt mauvais
- Moyenne sur 12 mois pour l'essence ordinaire en libre service : bon

# EXERCICES, LECTURES ET RÉFÉRENCES

EXPLORATION ET VISUALISATION DES DONNÉES

## EXERCICES SUPPLÉMENTAIRES

Exécutez les fichiers Jupyter Notebooks fournis pour explorer la façon de générer des visualisations automatiques.

Lire l'introduction à ggplot2 – ggplot2 est une bibliothèque de visualisations en R bien connue.

D'autres ressources de visualisation sont proposées sur le canal Slack « ressources »!

# RÉFÉRENCES

## Understanding Graphics

Krygier, J., Wood, D., [2016], *Making Maps: A Visual Guide to Map Design for GIS*, Guilford Press

## Interactive Data Visualization, Wikipedia

Is animation an effective tool for data visualization?, NASA

Perception in Visualization, C.G. Healey (très intéressant!)

## Data Physicalizations

Tufte, E. [2001], *The Visual Display of Quantitative Information*, Graphics Press.

Hu, D. [1954], *How to Lie With Statistics*, Norton

Tufte, E. [2008], *Beautiful Evidence*, Graphics Press

## RÉFÉRENCES

Nussbaumer Knaflic, C. [2015], *Storytelling with Data*, Wiley

Cairo, A. [2013], *The Functional Art*, New Riders

Cairo, A. [2016], *The Truthful Art*, New Riders

Meireilles, I. [2014], *Design de l'information*, Parramon France

50 Great Examples of Data Visualization: <http://www.webdesignerdepot.com>

Visualising Data

Nathan Yau, [FlowingData](#)

[Data Visualization](#), Wikipedia

[Misleading Graphs](#), Wikipedia

## RÉFÉRENCES

- Prabhakaran, S., [Top 50 ggplot2 Visualizations](#) (avec liste maîtresse de codes R).
- Miller, M. [2017], [The problem with Interactive graphics](#), Co.Design
- Wickham, H. [2016], *ggplot2: Elegant Graphics for Data Analysis* (2<sup>nd</sup> ed), Springer.
- Gorelik, B., [Data Visualization](#) (blogue).
- Chang, W. [2013], *R Graphics Cookbook*, O'Reilly.
- Wickham, H. [2009], A Layered Grammar of Graphics, *Journal of Computational and Graphical Statistics* 19:3–28.
- Horton, N.J., Kleinman, K. [2016], *Using R and RStudio for Data Management, Statistical Analysis, and Graphics*, 2nd ed., CRC Press.
- Healey, K. [2018], *Data Visualization: A Practical Introduction*.

## RÉFÉRENCES

- Kabacoff, R.I. [2011], R in Action, Second Edition: Data analysis and graphics with R, Live.
- Maindonald, J.H. [2008], Using R for Data Analysis and Graphics: Introduction, Code and Commentary.
- Tyner, S., Briatte, F., Hofmann, H. [2017], Network Visualization with ggplot2, The R Journal, vol. 9(1).
- Broman, K. [2016], Data Visualization with ggplot2.
- Robinson, D., Visualizing Data Using ggplot2, on varianceexplained.org.
- Manipulating, analyzing and exporting data with tidyverse, on datacarpentry.org.
- Wickham, H. [2014], Tidy Data, Journal of Statistical Software, v59, n10.
- Gashim, E., Boily, P. [2018], A ggplot2 Primer, [data-action-lab.com](http://data-action-lab.com)