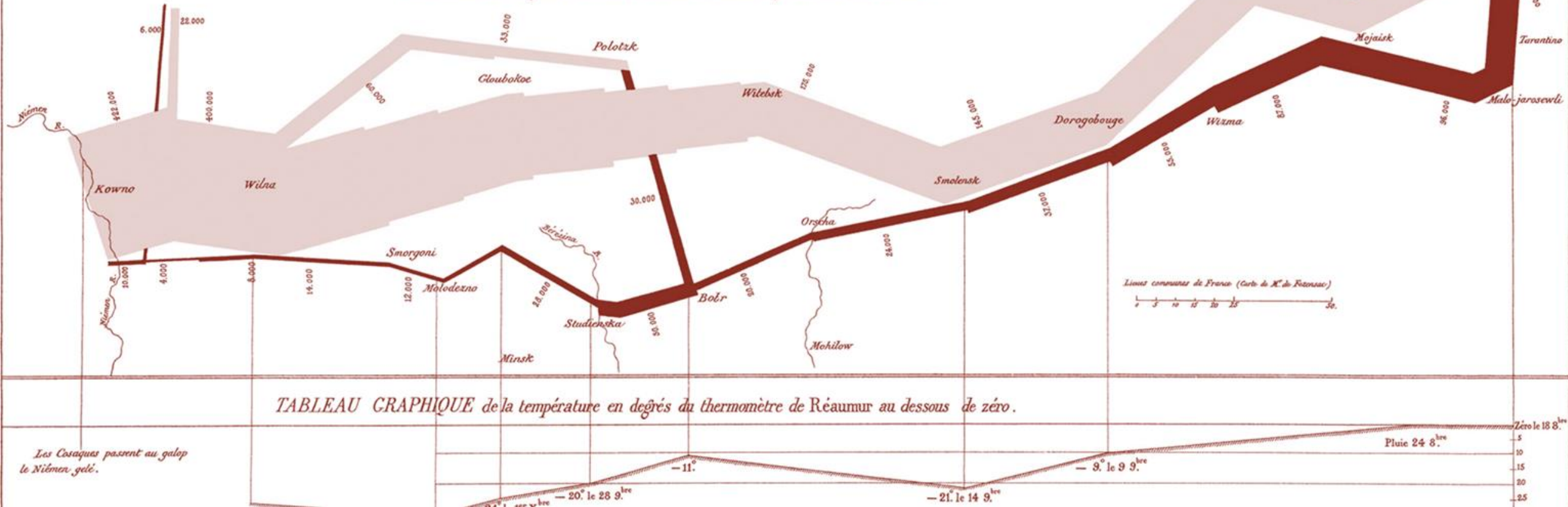


Les concepts de la visualisation des données

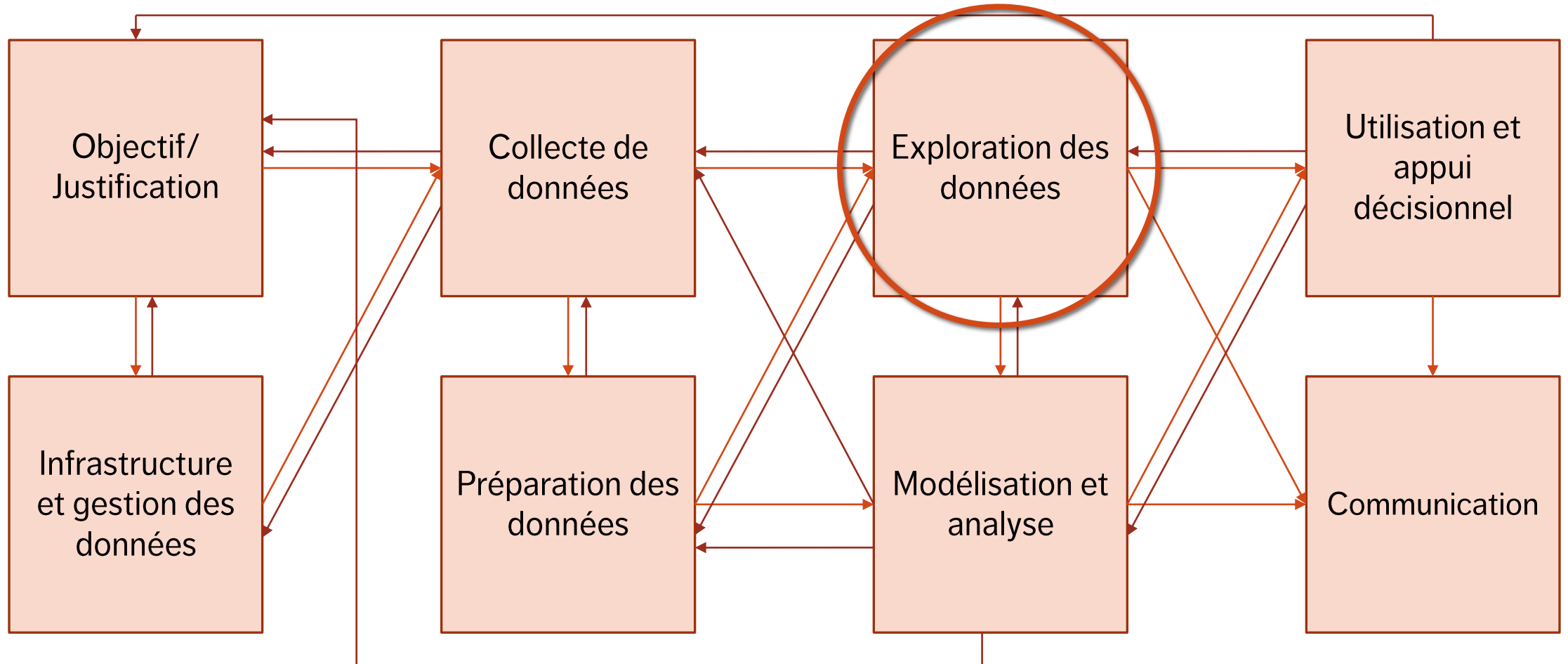
LA VISUALISATION DES DONNÉES ET LES TABLEAUX DE BORD

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



1. L'analyse exploratoire des données

Le processus d'analyse (désordonné)



Quelques questions de base

Quel **système** vos données représentent-elles - **objets, attributs, relations** ?

Comment représente-t-il ce système - c-à-d quel est le **modèle de données** ?

Qui a créé cet ensemble de données ? Quand ? Dans quel but ?

Supposons un fichier “plat” - que représentent les lignes et les colonnes ?

Disposez-vous de suffisamment d'informations (e.g., des **métadonnées**) pour répondre à ces questions ? Où pouvez-vous trouver plus d'informations ?

Les résumés non visuels

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

```

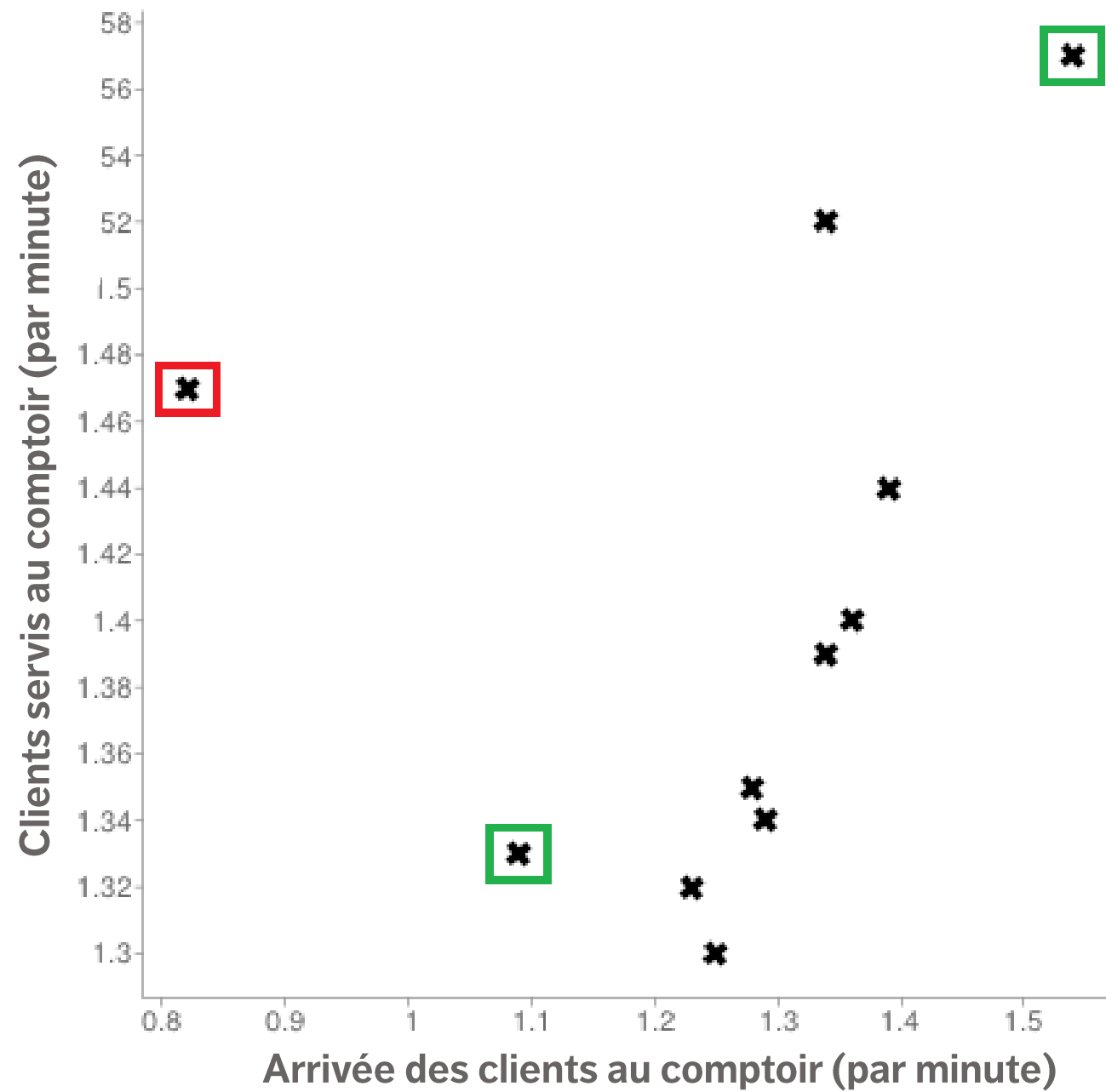
season
Length:340
Class :character
Mode :character
autumn spring summer winter
      80      84      86      90

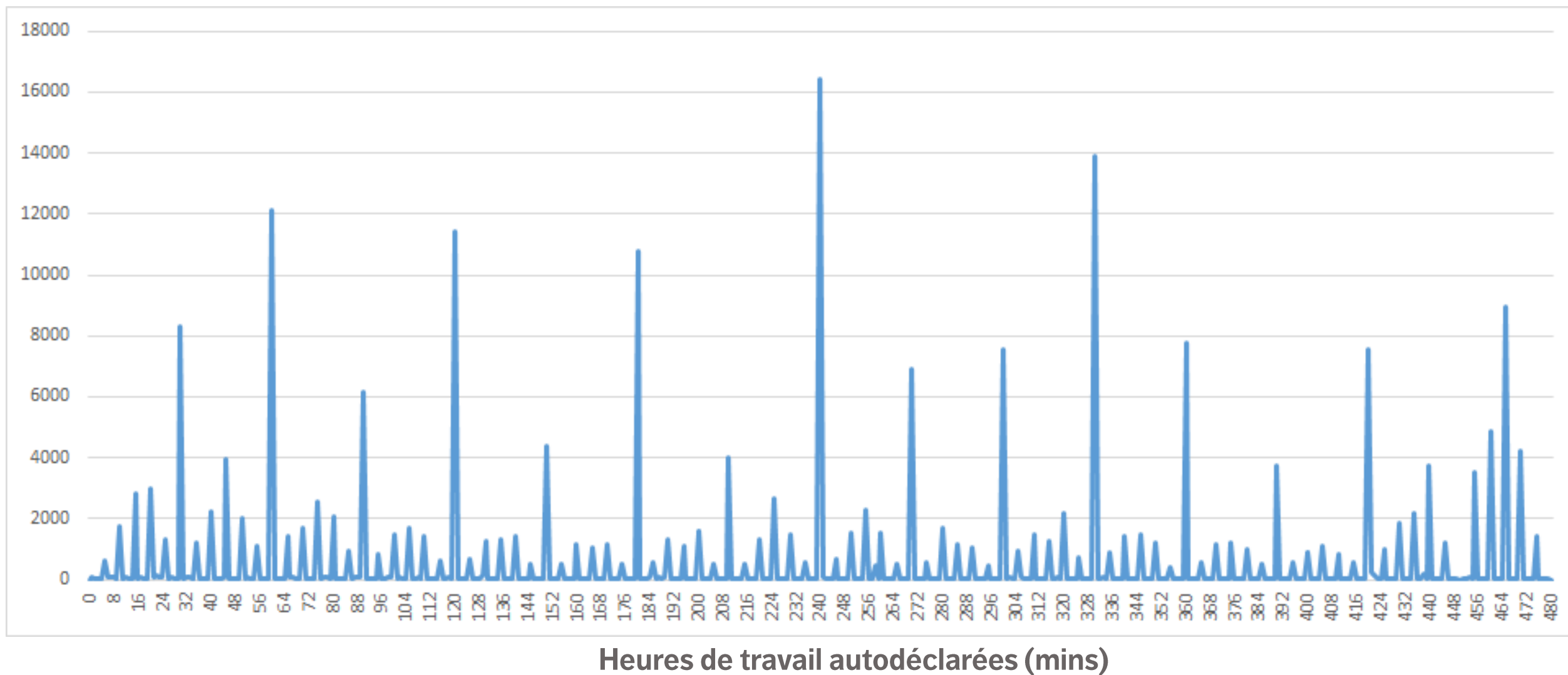
```

L'utilisation avant l'analyse

La visualisation des données peut être utilisée pour préparer l'étape de l'analyse :

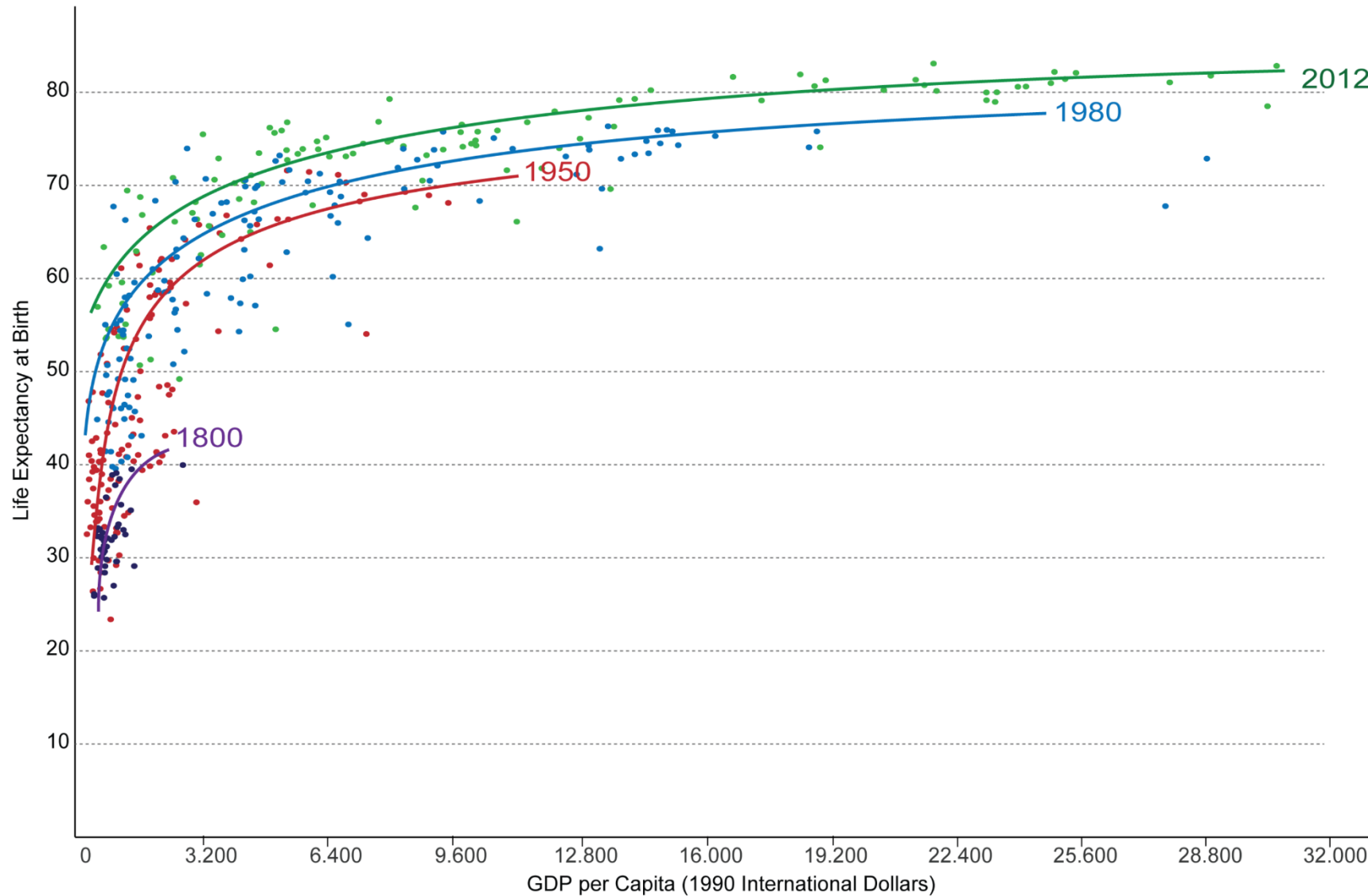
- **détecter les entrées anormales**
entrées non valides, valeurs manquantes, valeurs aberrantes
- **façonner les transformations des données**
“binning”, normalisation, transformations Box-Cox, transformations de type PCA
- **se faire une idée des données**
analyse des données en tant que forme d'art, analyse exploratoire
- **identification de la structure cachée des données**
regroupement, associations, modèles informant les étapes suivantes





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

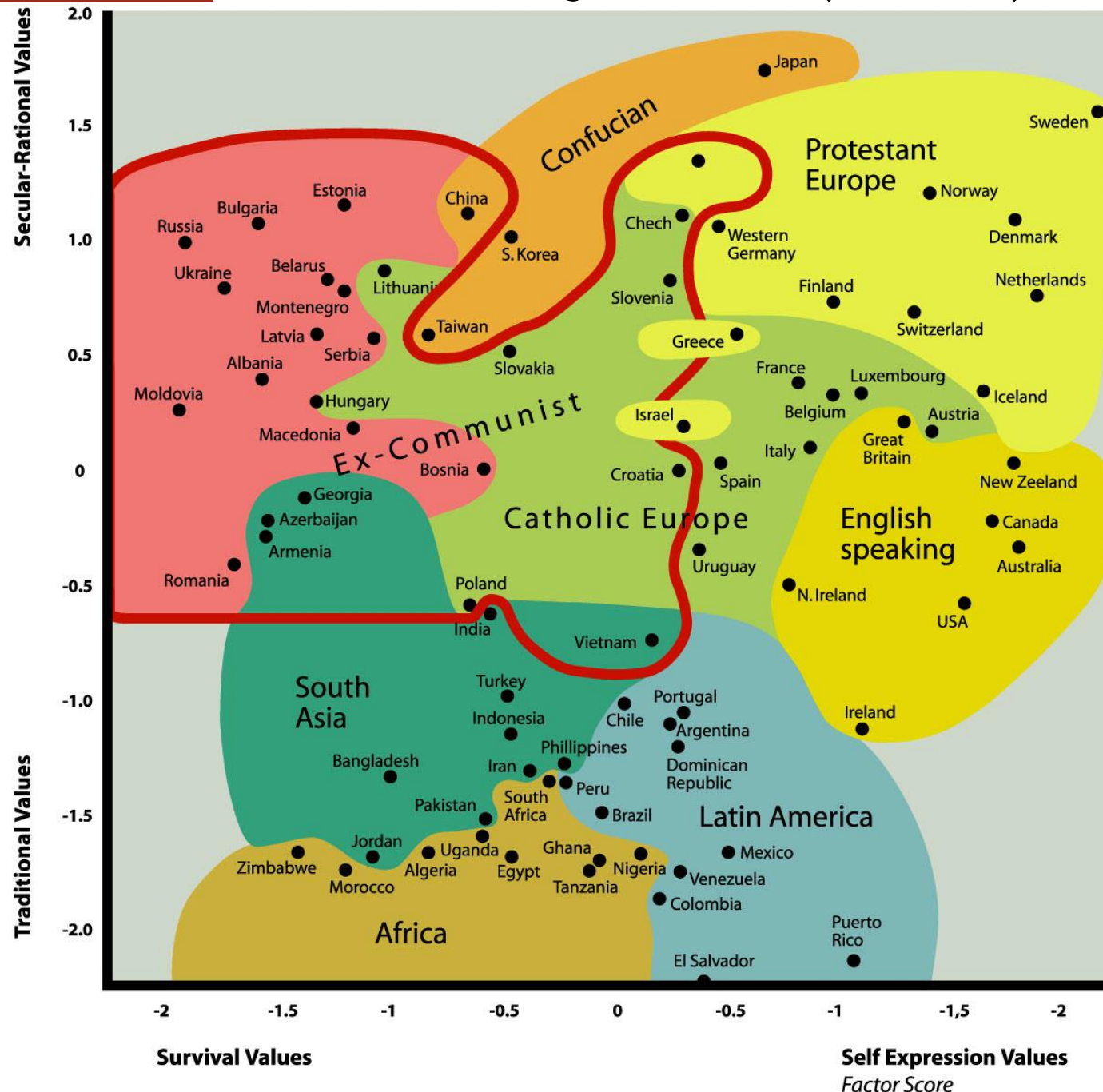
GDP per capita is measured in International Dollars. This is a currency that would buy a comparable amount of goods and services a U.S. dollar would buy in the United States in 1990. Therefore incomes are comparable across countries and across time.



Ce graphique montre la corrélation entre l'espérance de vie et le PIB par habitant.

Les pays dont le PIB est plus élevé ont une espérance de vie plus élevée, en général.

La relation semble suivre une tendance logarithmique : l'augmentation de l'espérance de vie par unité d'augmentation du PIB diminue à mesure que le PIB par habitant augmente.



Les valeurs traditionnelles

importance de la religion, des liens parents-enfants, de la déférence envers l'autorité et des valeurs familiales traditionnelles.

Valeurs laïques et rationnelles

Moins d'importance accordée à la religion, aux valeurs familiales traditionnelles et à l'autorité.

Valeurs de survie

accent mis sur la sécurité économique et physique.

Valeurs d'expression personnelle

priorité élevée à la protection de l'environnement, tolérance croissante à l'égard des étrangers, des gays et lesbiennes et de l'égalité des sexes

Quelques graphiques d'exploration

Texte et tableaux

Cartes à tapis/Lignes de nombres

Histogrammes/Chartes à barres

Boxplots/Graphiques à moustache

Graphiques en ligne

Diagrammes de dispersion

Les lignes de nombres (“rugs”)

Les lacunes dans la ligne numérique : l’**absence** de ces valeurs numériques dans les données.

Rappelez-vous : ceci est (peut-être) différent de l'ordre dans lequel les valeurs apparaissent dans l'ensemble de données – puisqu'il s'agit d'une ligne de nombres, elle montre où les valeurs tombent numériquement.

Si certaines valeurs sont **identiques**, elles sont superposées.

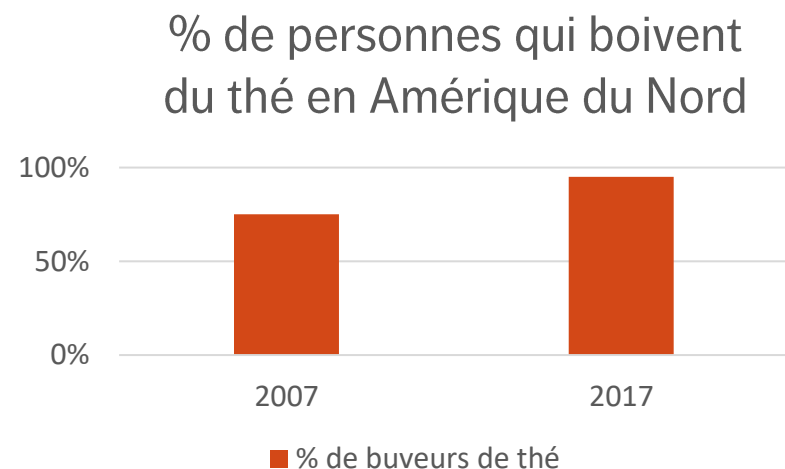


Du texte simple

Un ou deux chiffres sur lesquels se concentrer.

Bon pour "planter le décor".

Attire l'attention sur une zone du rapport.



95% de la population
boit du thé aujourd'hui, par rapport à
75% en 2007

Les tables et tableaux

Les tableaux interagissent avec notre système **verbal**, ce qui signifie que nous les **lisons** :

- utilisés pour comparer les valeurs
- le public cherche les rangées

Le design de la table doit **se fondre** dans le décor

- les données doivent ressortir, pas les bordures
- tableaux/données denses : utilisez des couleurs de ligne qui alternent

Nom	L'année dernière	Cette année
Bob	20	30
Fred	30	40
George	10	15

Nom	L'année dernière	Cette année
Bob	20	30
Fred	30	40
George	10	15

Les tables “heatmap”

Tirez parti de la couleur pour transmettre l'ampleur des quantités

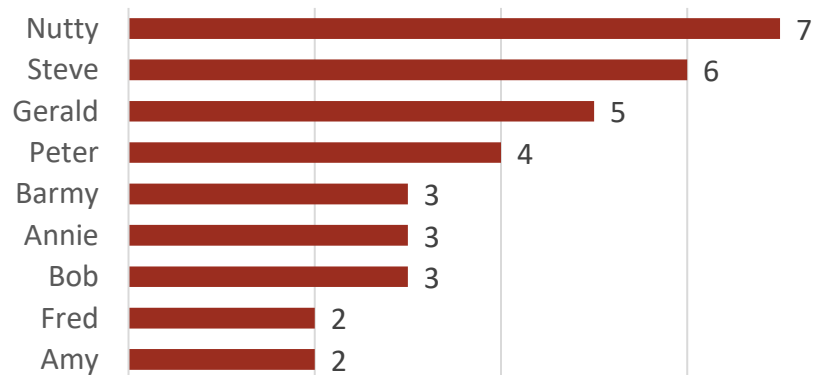
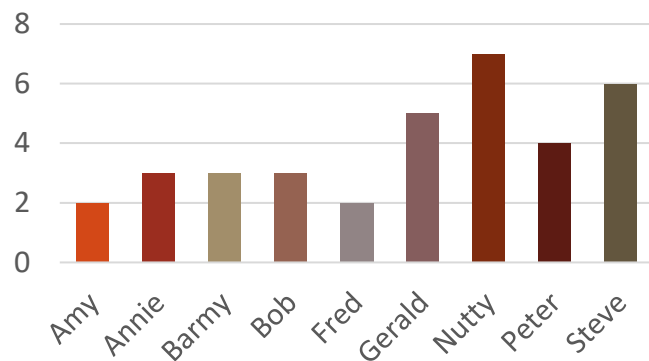
- utilisez la **saturation d'une seule couleur** plutôt que la différenciation (différentes couleurs)
- avec une légende (blanc = faible, bleu = élevé), les chiffres peuvent être supprimés sans altérer le message

	Last Year	This Year	Next Year	Optimum
George	20	20	20	20
Peter	40	35	30	25
John	10	10	5	5
Sandra	25	30	35	40

	Last Year	This Year	Next Year	Optimum
George	20	20	20	20
Peter	40	35	30	25
John	10	10	5	5
Sandra	25	30	35	40

	Last Year	This Year	Next Year	Optimum
George				
Peter				
John				
Sandra				

Les diagrammes à barres



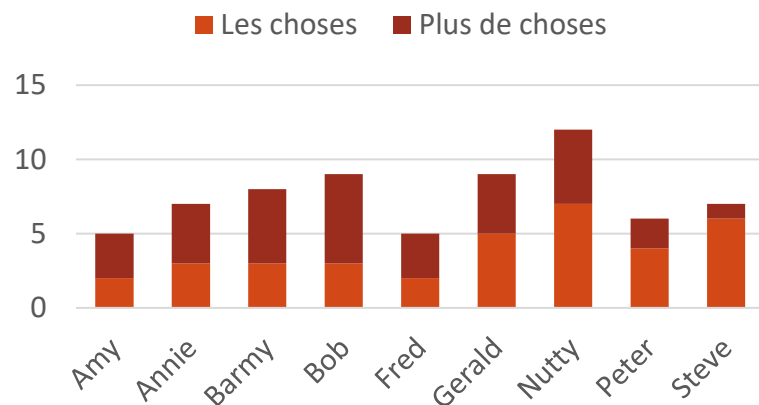
Très polyvalent et utile : utilisez TOUJOURS (?) une ligne de référence au zéro.

Utilisez soit l'axe du graphique OU des étiquettes pour les données (l'axe pour les déclarations générales, les étiquettes de données pour plus de détails).

Les graphiques horizontaux sont apparemment **plus faciles à lire** (selon de nombreuses études).

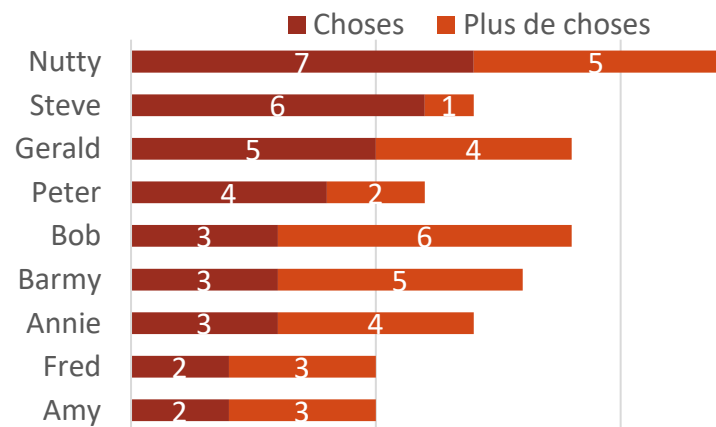
Pensez à l'ordre des catégories.

Les diagrammes à barres empilées



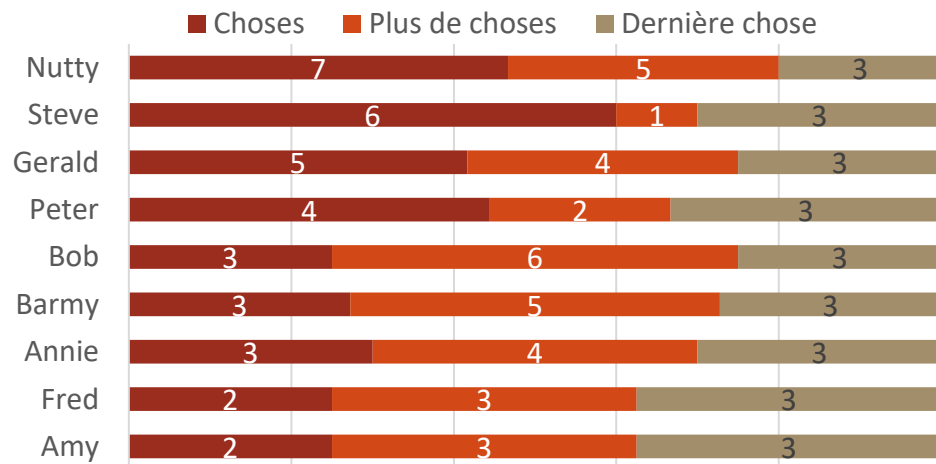
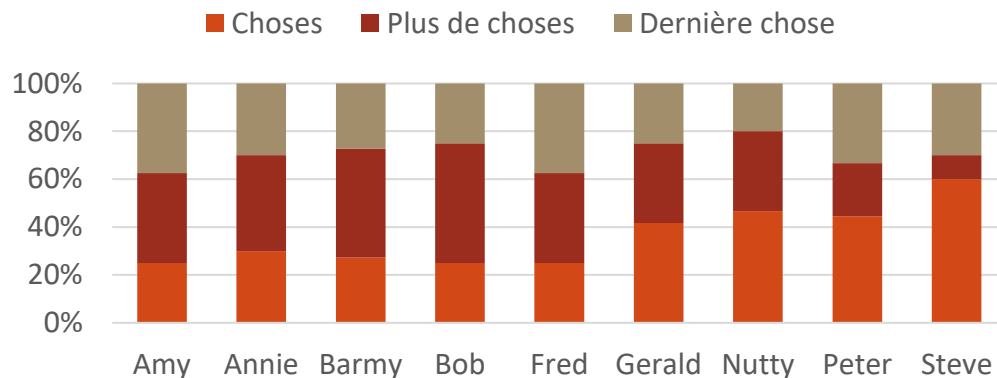
Conçus pour **comparer les totaux**, mais peuvent rapidement devenir “**écrasant**”.

Difficile de trier / ordonner.



Le filtrage est assez compliqué (sur quoi faut-il cliquer et comment le graphique réagit-il lorsque l'on clique sur le filtre ?)

Les diagrammes à barres à 100%



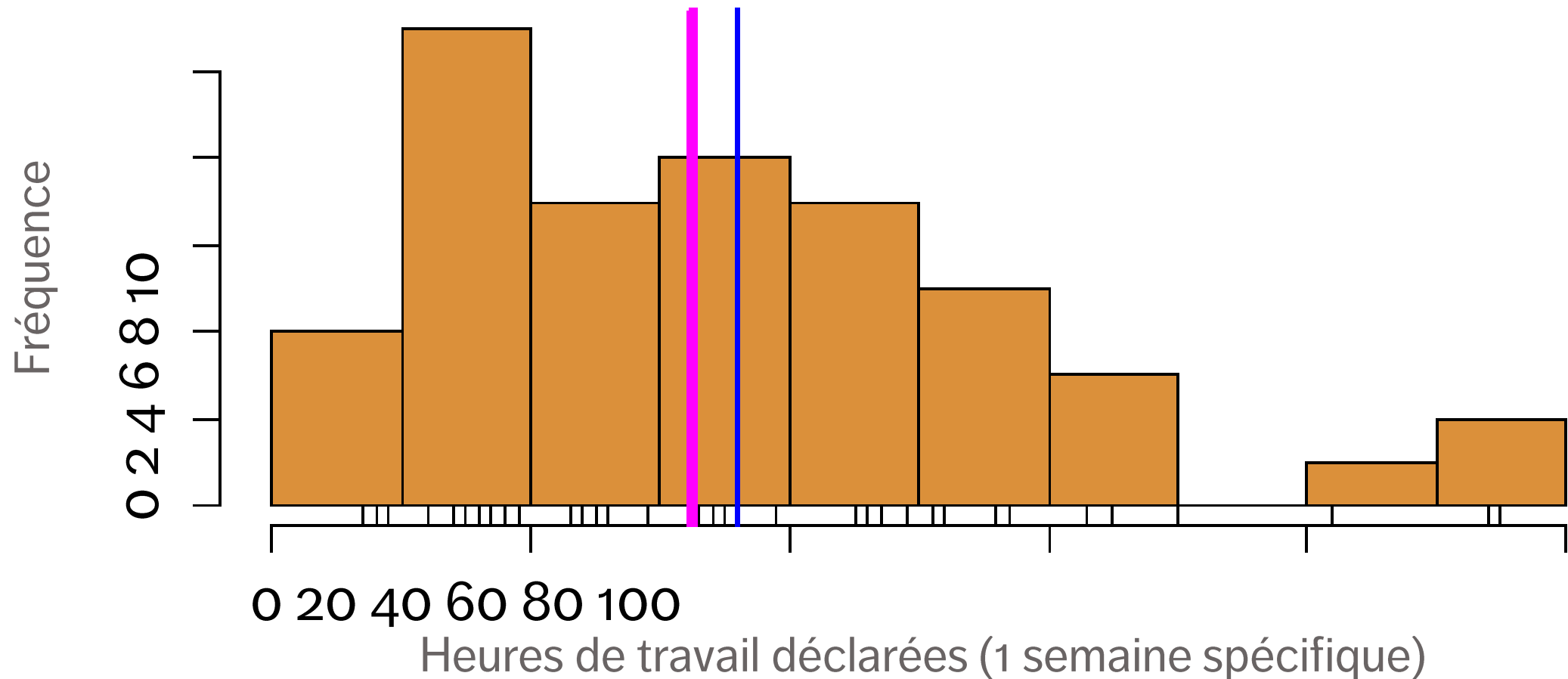
Fonctionnent bien pour visualiser des **portions** d'un tout sur une échelle allant du négatif au positif.

Ligne de base cohérente aux extrêmes.

Aucune mesure relative à l'**ampleur** des données.

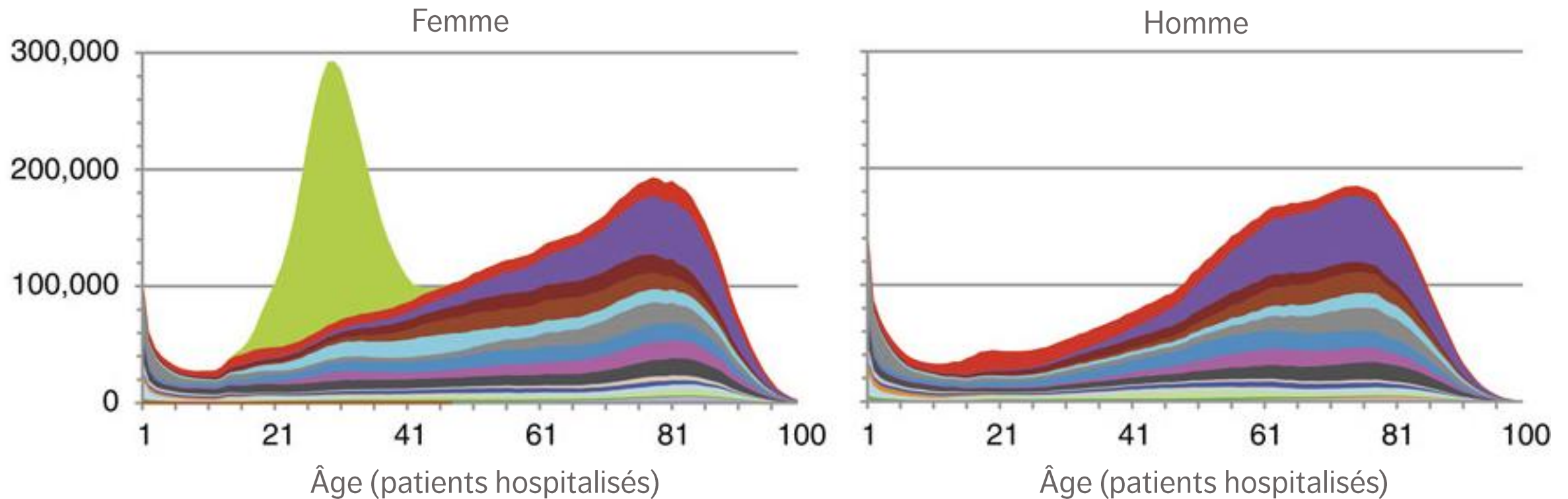
La recherche montre que l'horizontal est plus facile à lire que le vertical.

Les histogrammes

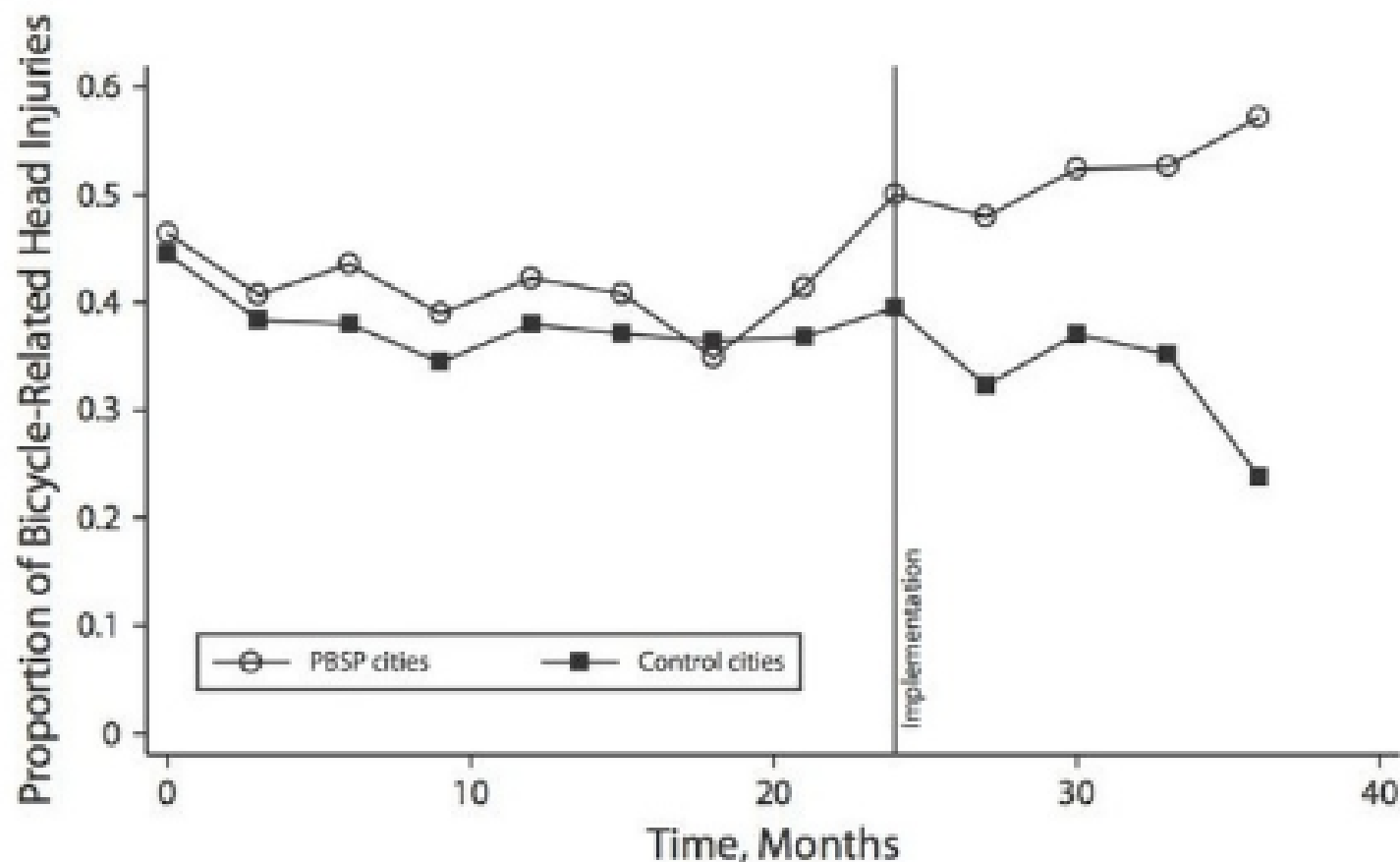


Les histogrammes empilés

Nombre de diagnostics (patients hospitalisés)

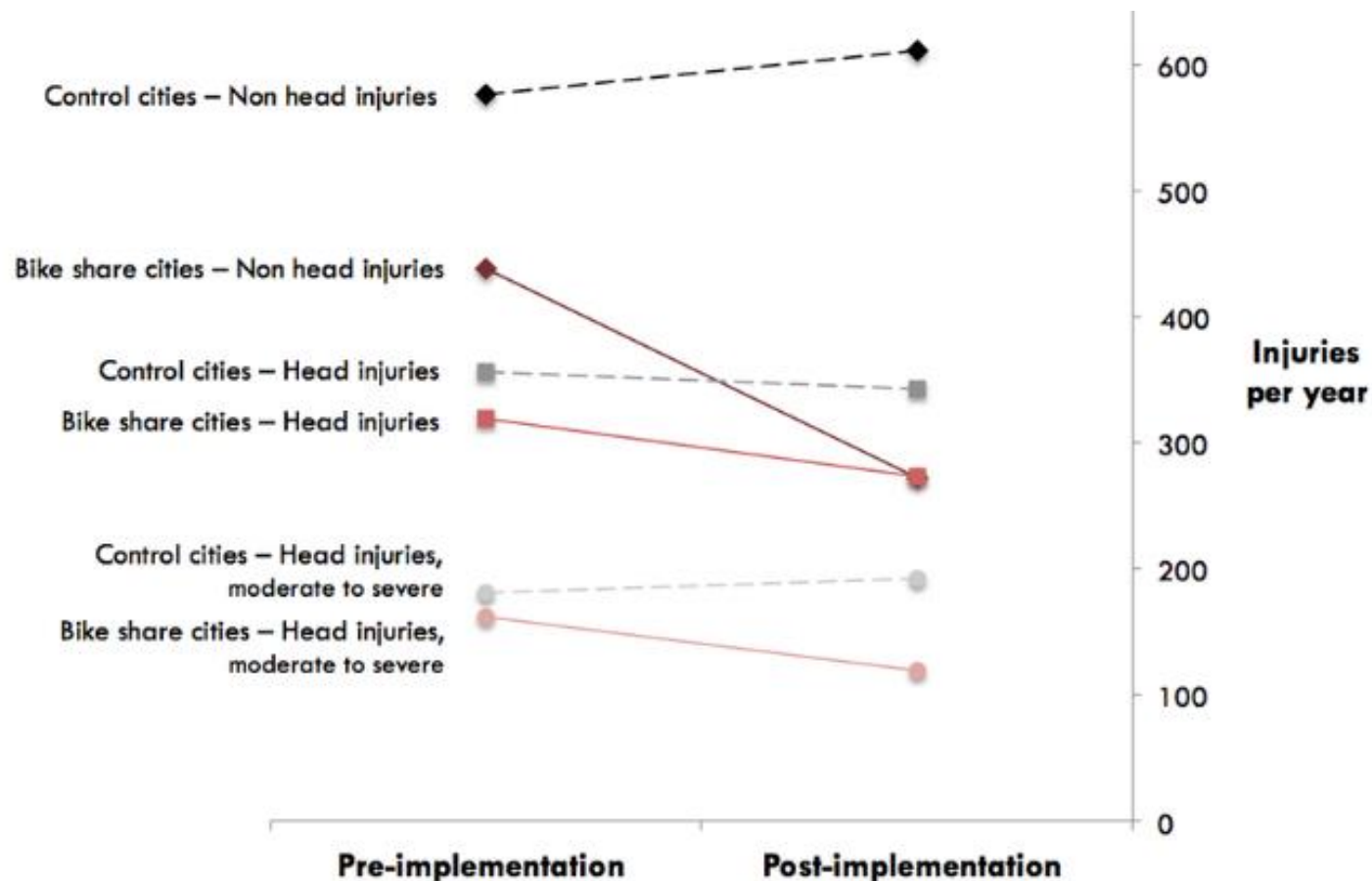


Les graphiques en ligne



Proportion de toutes les blessures liées à la bicyclette qui ont été classées comme des blessures à la tête parmi les villes ayant un programme de vélos en libre-service et les villes témoins, centrées sur la date d'intervention.

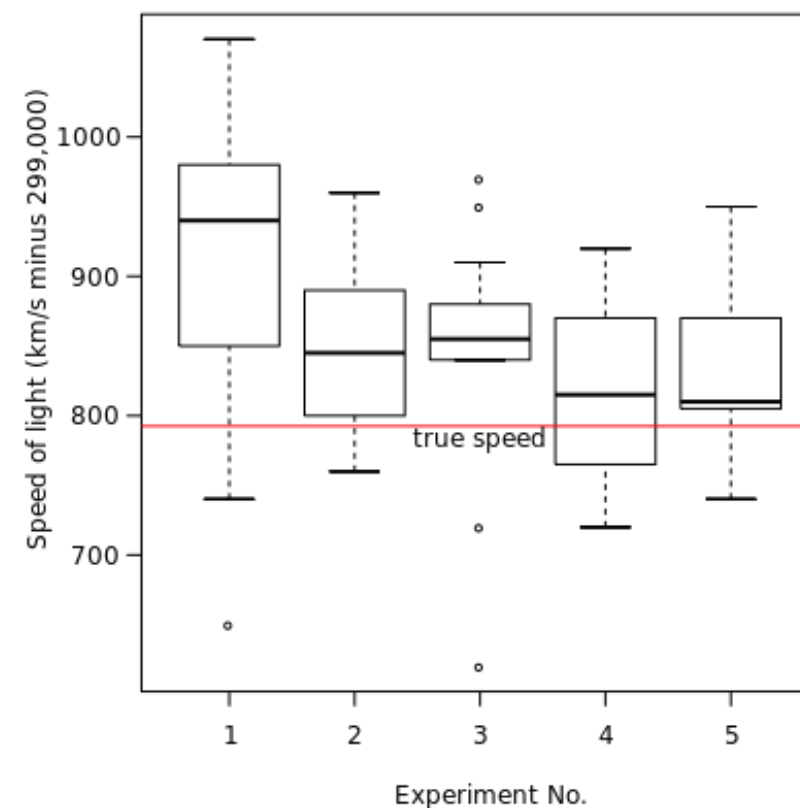
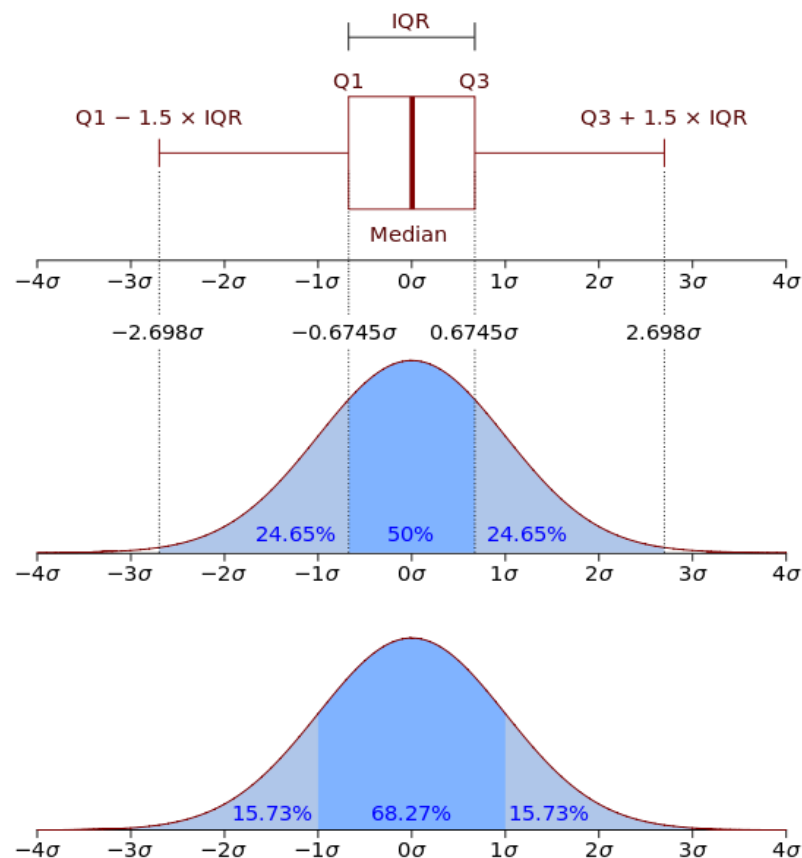
Les graphiques en ligne



Les données d'une nouvelle étude montrent une baisse de toutes les blessures, y compris les traumatismes crâniens, après la mise en place du système de vélos en libre-service.

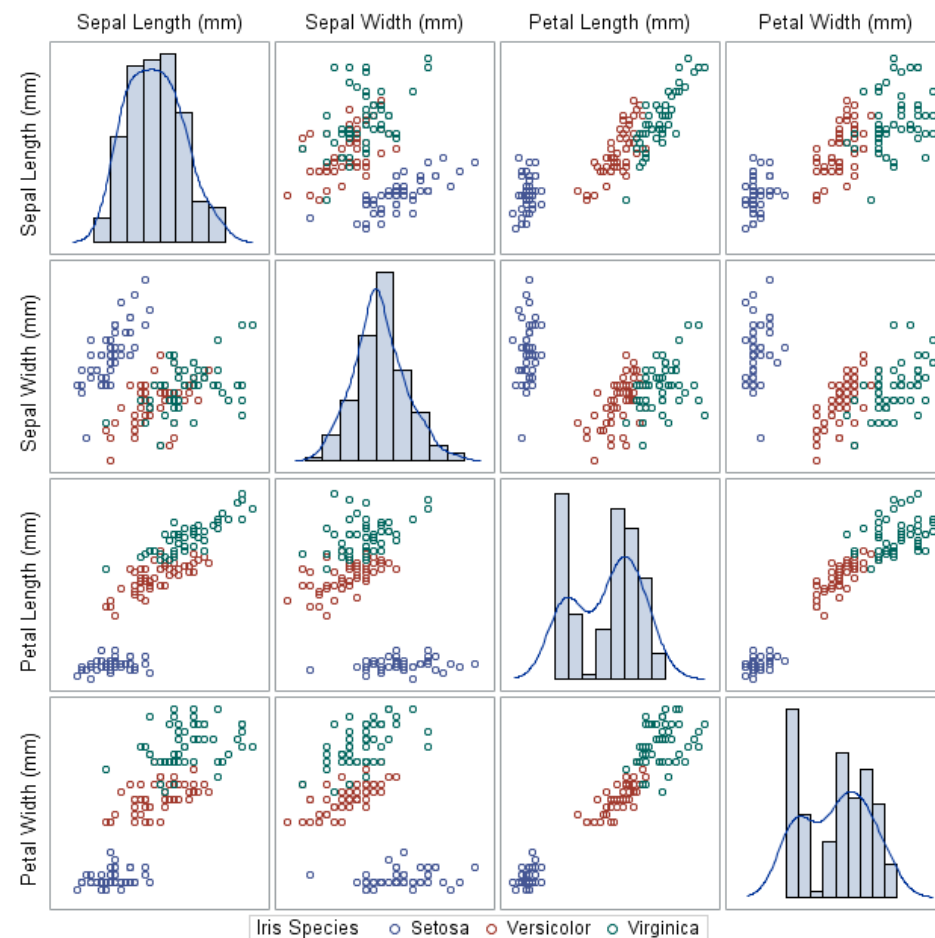
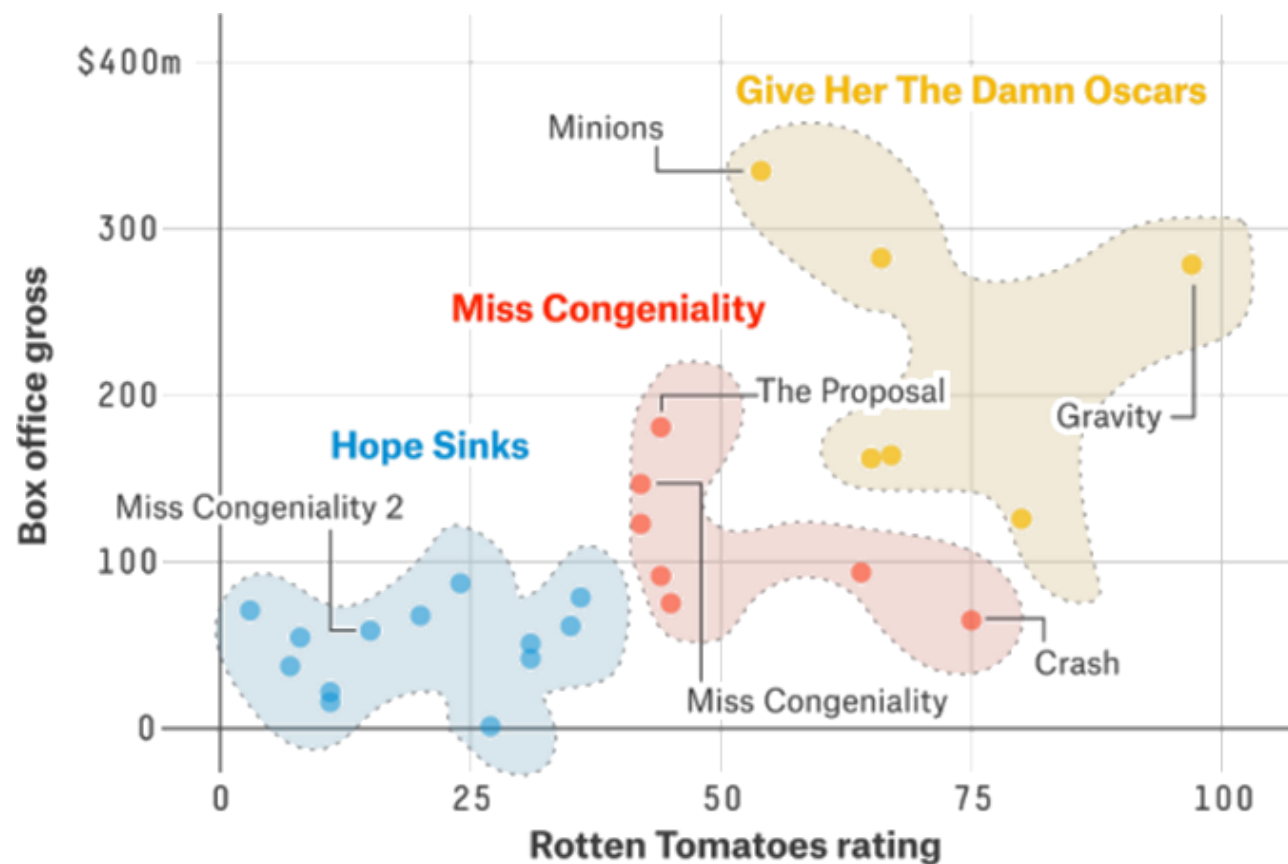
Comme les traumatismes crâniens diminuent moins que les autres blessures, ils représentent aujourd'hui une proportion plus importante de toutes les blessures.

Les boîtes à moustaches (“boxplots”)



Expérience de Michelson-Morley afin de déterminer la vitesse de la lumière

Les diagrammes de dispersion



Lectures conseillées

L'analyse exploratoire des données

Data Understanding, Data Analysis, Data Science **Volume 2: Fundamentals of Data Insight**

18. Data Exploration and Data Visualization

18.1 Data and Charts

- Pre-Analysis Uses

The Practice of Data Visualization **Part I: A Data Exploration Primer**

2. Data Exploration and Visualization

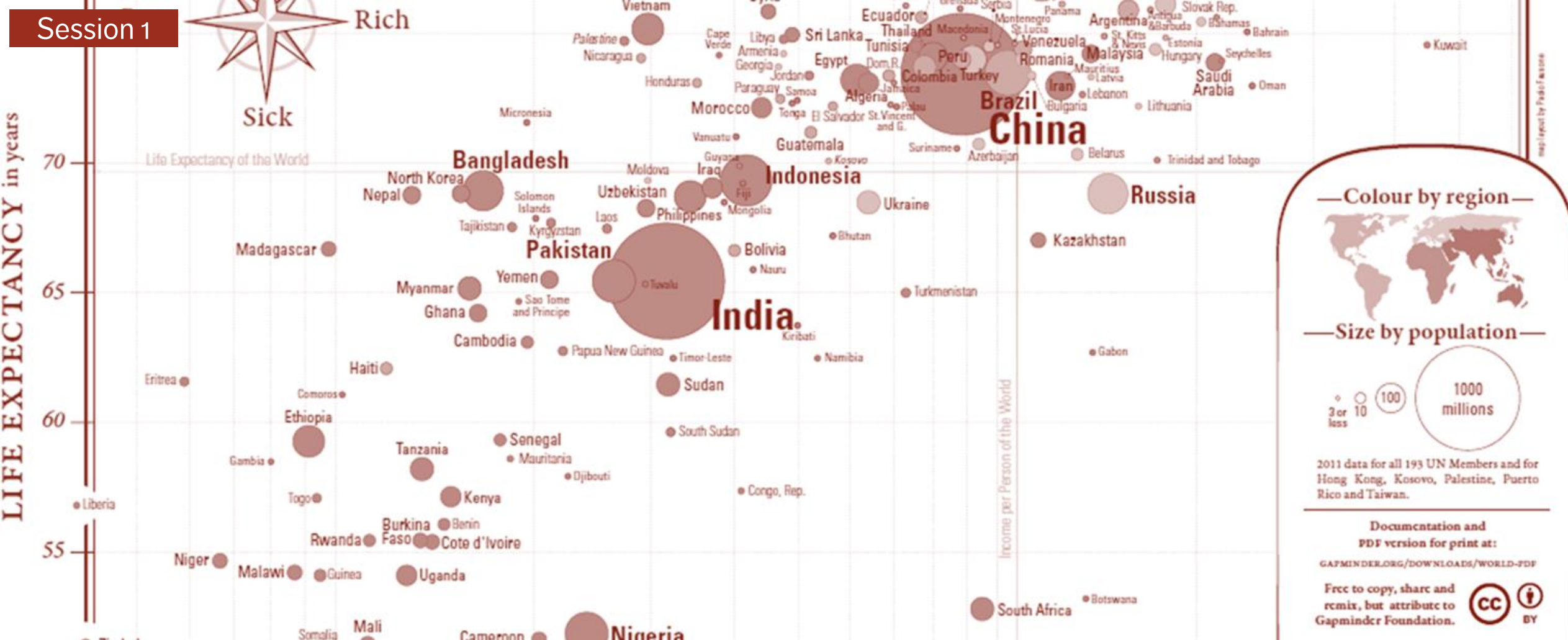
2.1 Exploratory Data Analysis

2.2 Workhorse Data Visualizations

Exercices

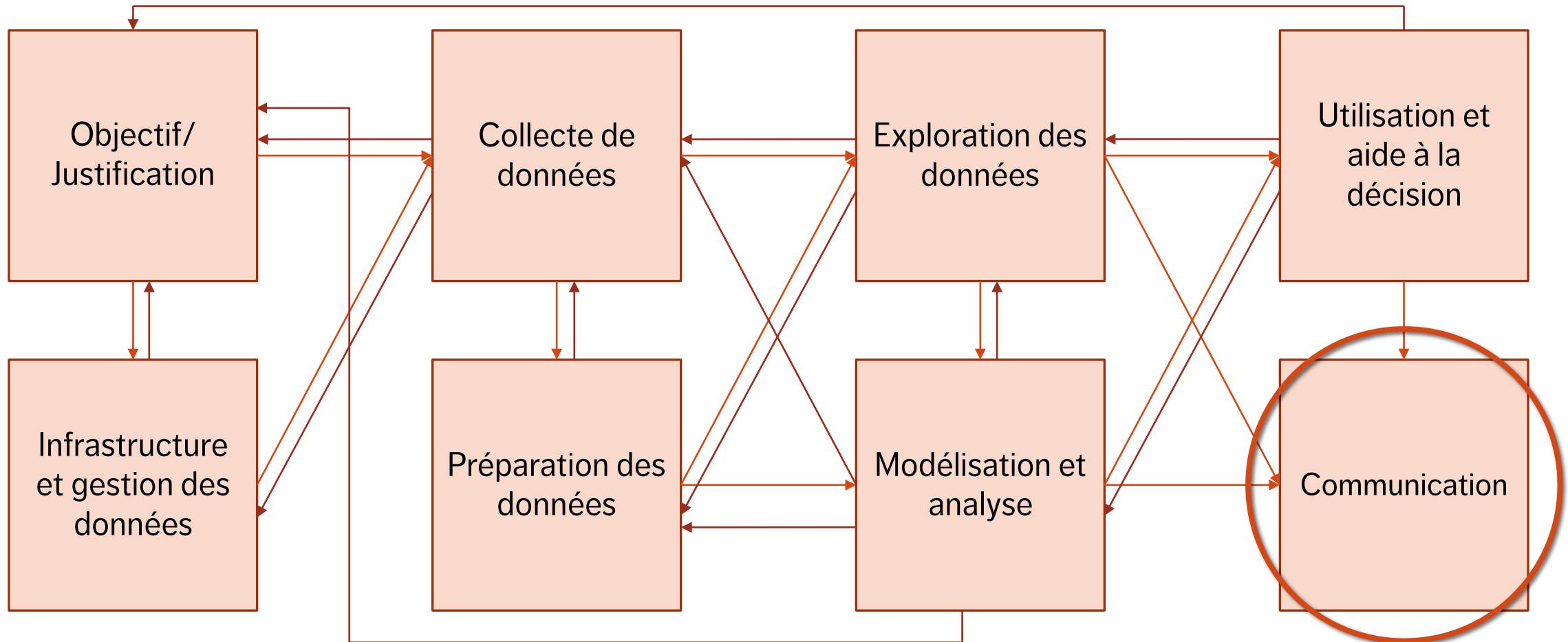
L'analyse exploratoire des données

1. Trouvez des exemples de présentations de données que vous considérez comme particulièrement perspicaces et/ou puissantes. Discutez de leurs forces/faiblesses.
2. Trouvez des exemples de présentations de données que vous considérez comme particulièrement trompeuses et/ou inutiles. Discutez de leurs forces/faiblesses.
3. Comment pensez-vous que les nouvelles technologies (par exemple, la réalité virtuelle ou augmentée, l'impression 3D, l'informatique portable) influenceront les présentations de données ?



2. La communication et la visualisation des données

Le processus d'analyse (désordonné)

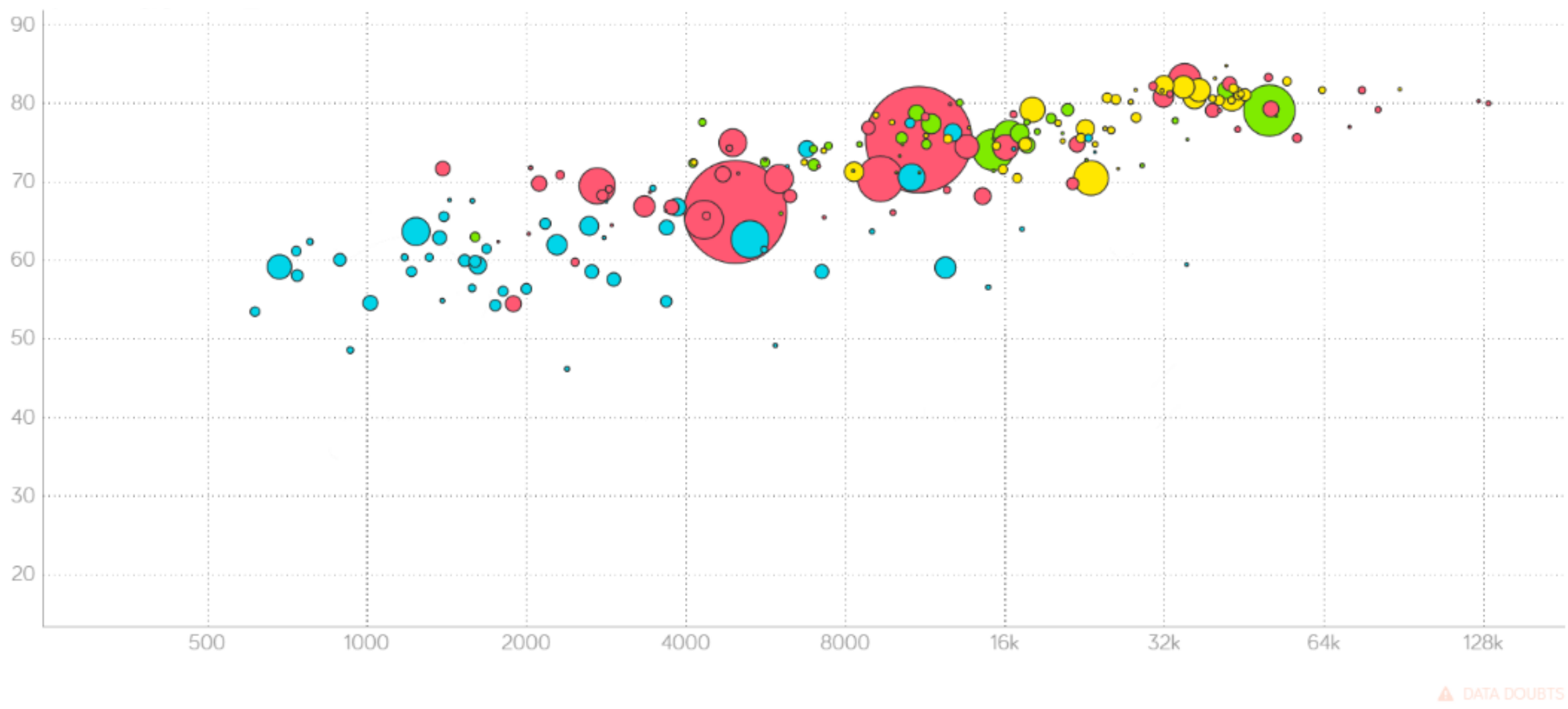


Les principes du design analytique

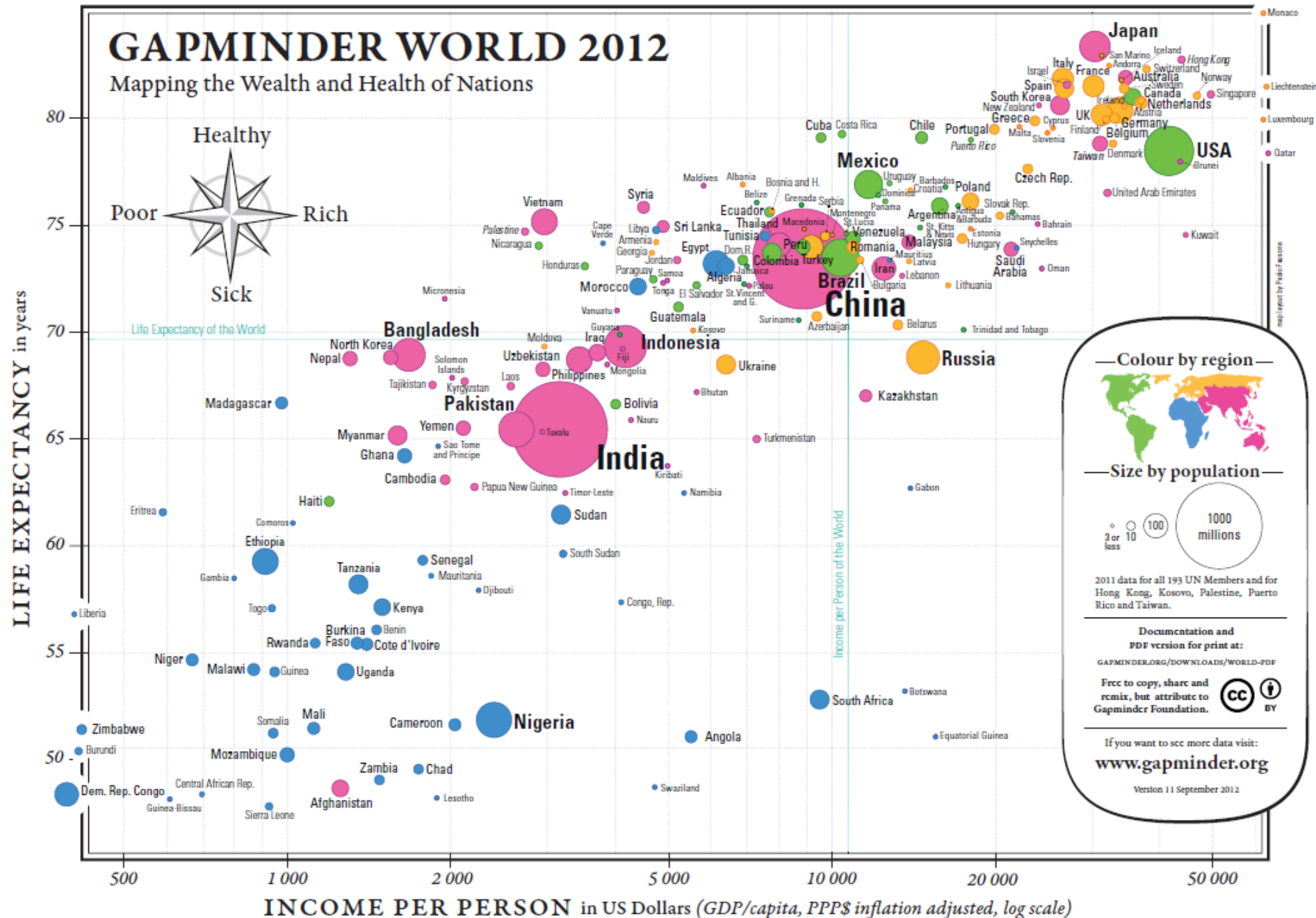
Le **raisonnement** et la **communication de** nos pensées sont intimement liés.

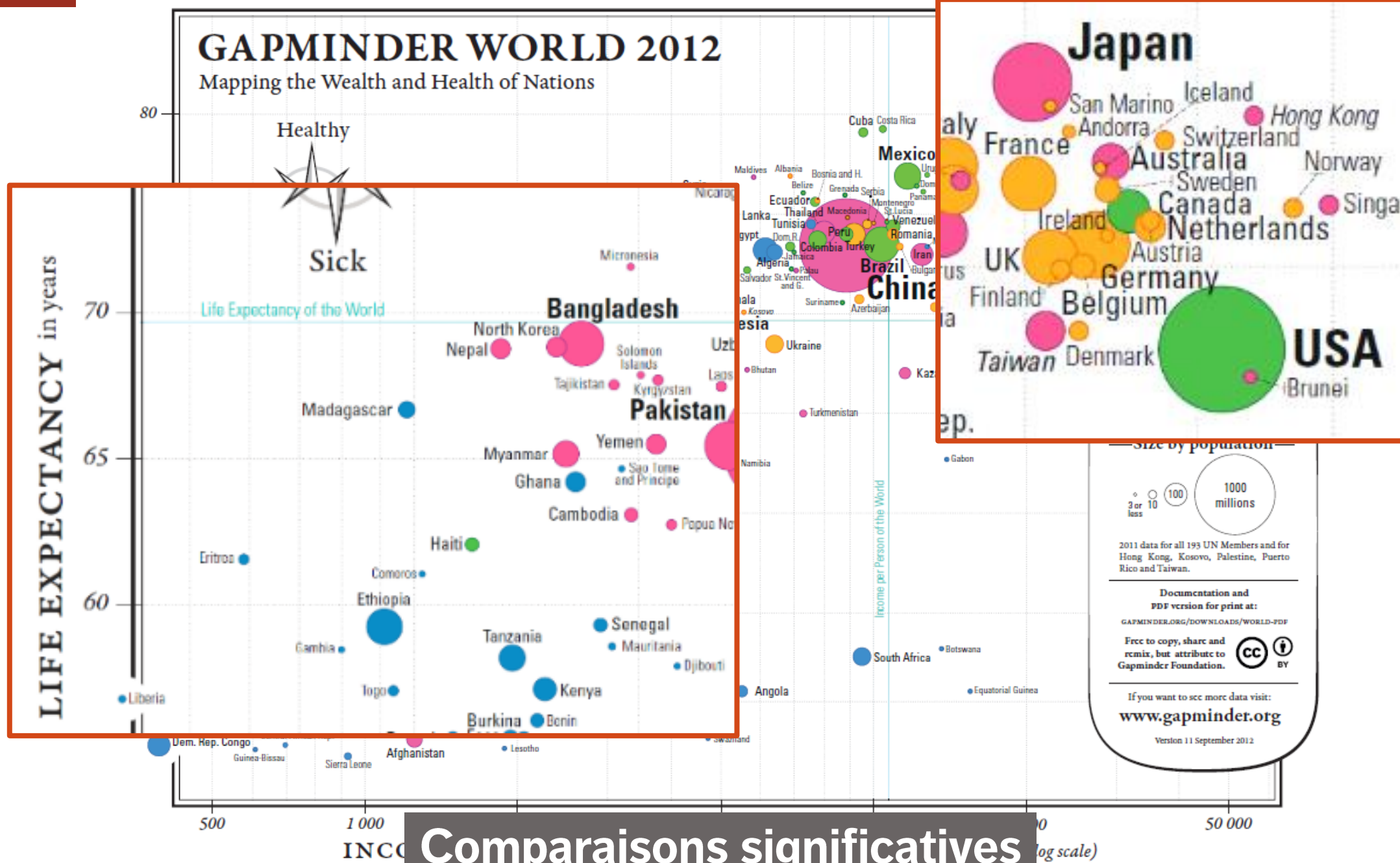
Il existe une **symétrie** dans la présentation visuelle des preuves. Le public devrait rechercher exactement ce que les producteurs devraient fournir :

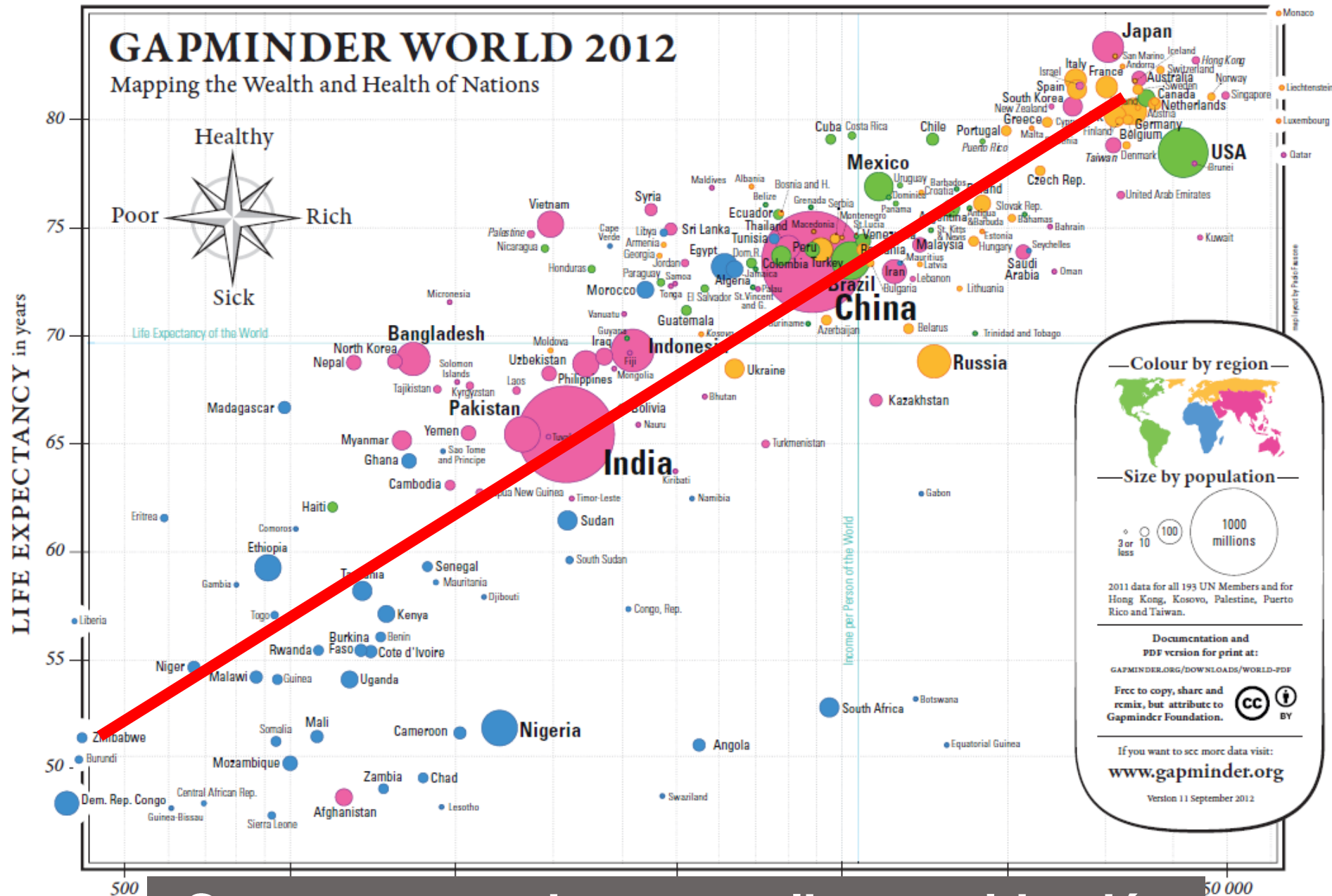
- des comparaisons significatives
- des réseaux de causalité potentiels et la structure sous-jacente
- des liens multivariés
- des données intégrées et pertinentes
- une documentation honnête
- l'accent est mis sur le contenu

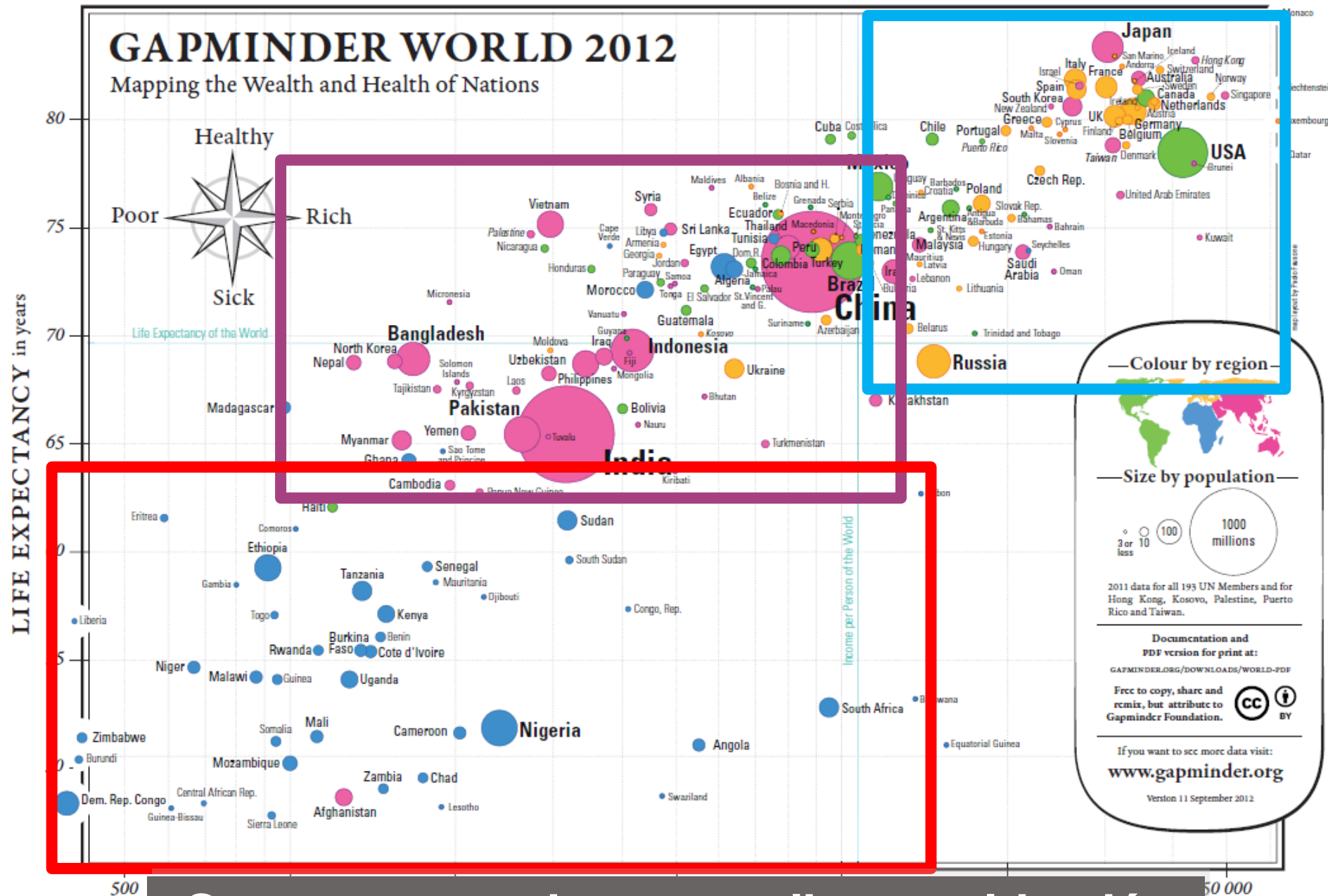


Données non intégrées

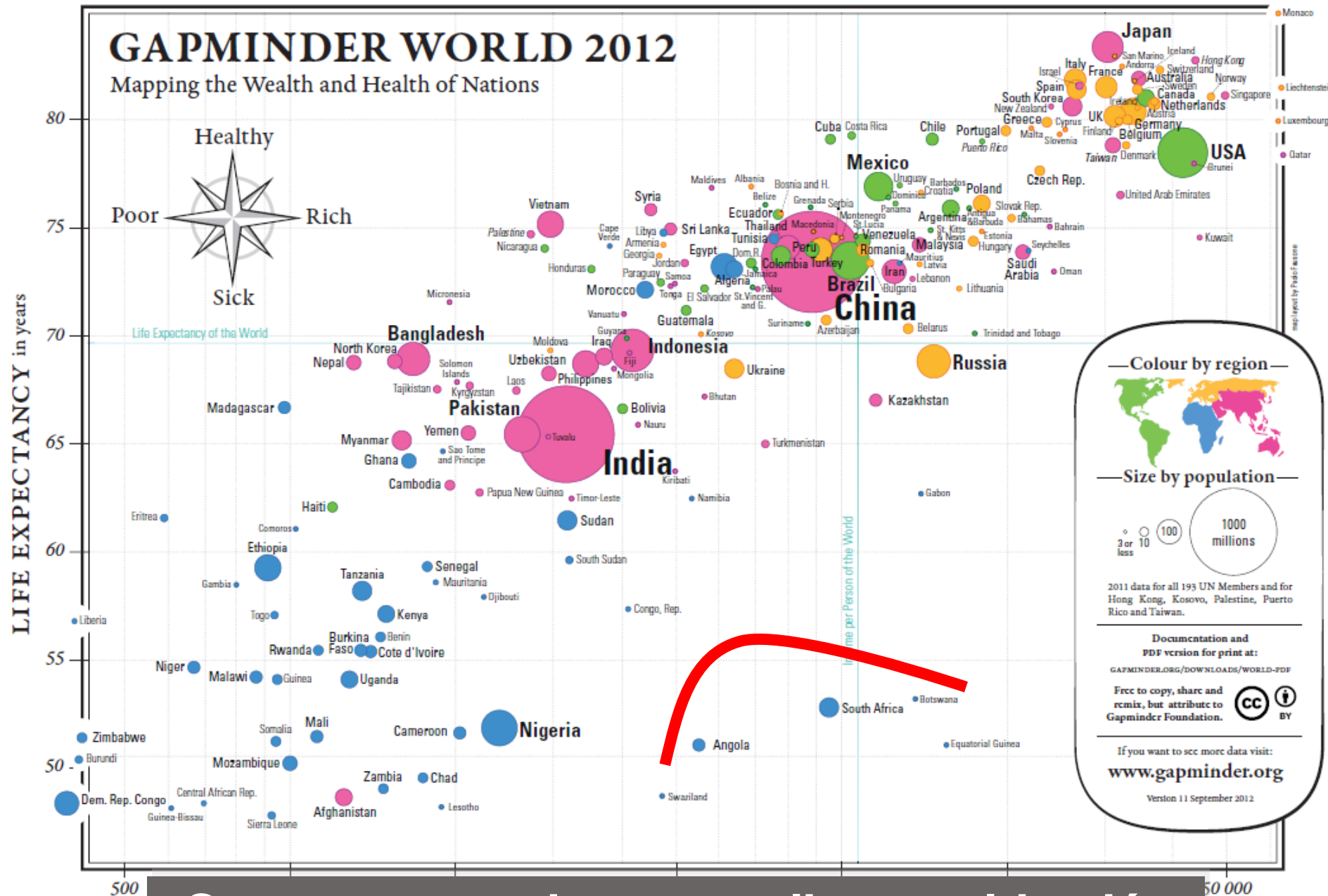




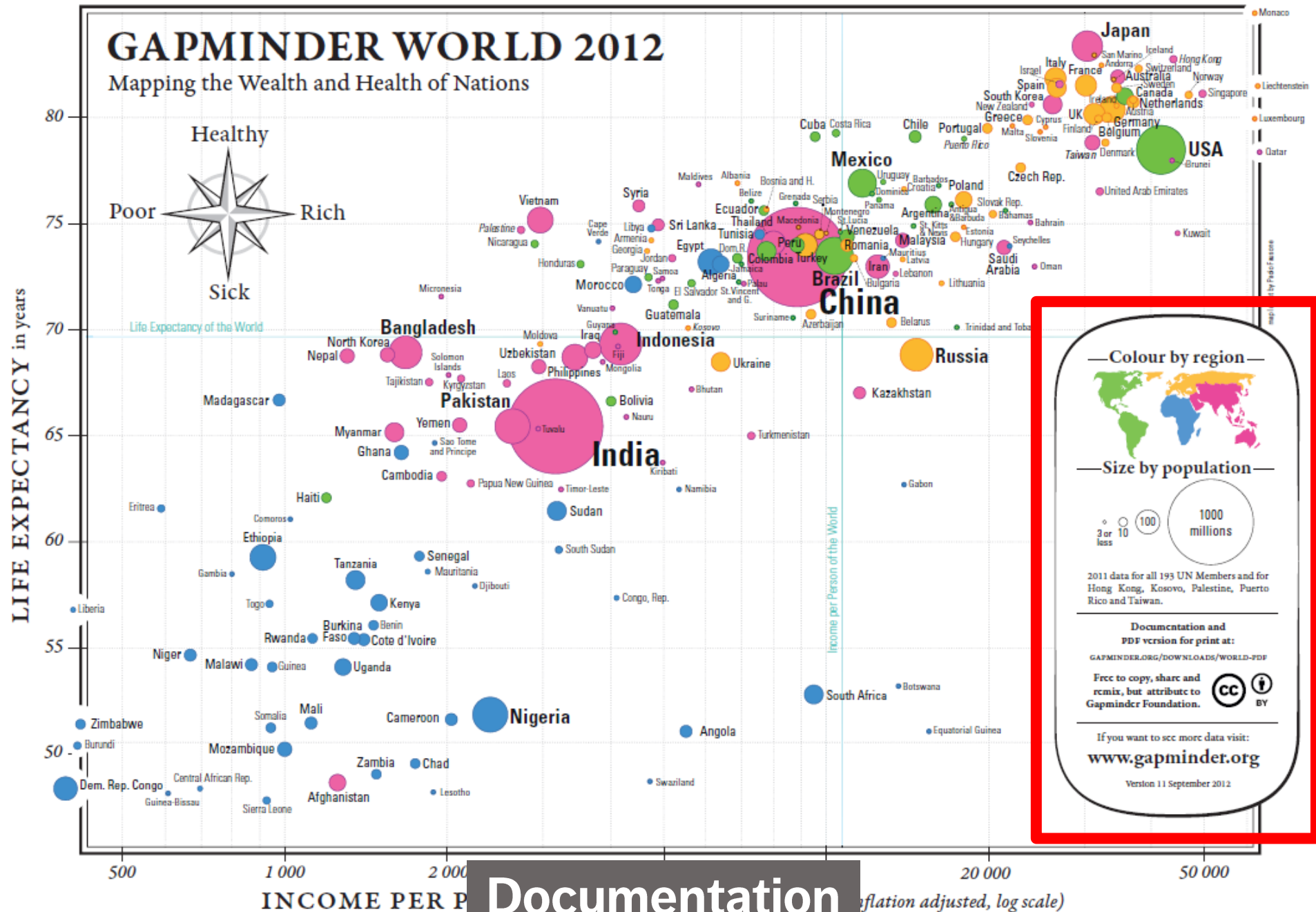




Structure sous-jacente et liens multivariés



Structure sous-jacente et liens multivariés



Les règles fondamentales

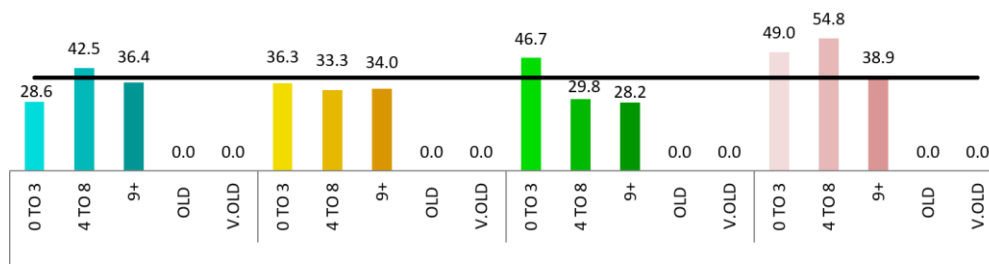
1. Vérifiez les données

aberrations, pics, anomalies

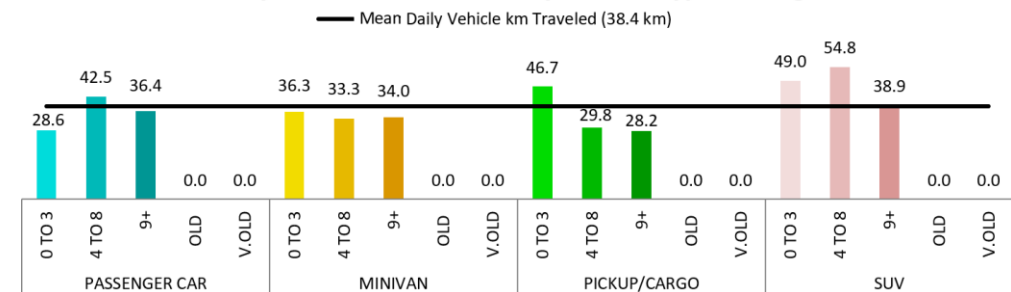
2. Expliquez le codage

ne supposez pas que le public sait ce que tout signifie

Daily Vkt by Type and Age



Daily Vehicle km Traveled by Vehicle Type and Age



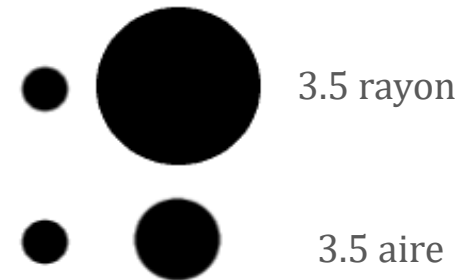
3. Étiqueter les axes

Il est important de connaître l'échelle

Les règles fondamentales

4. Inclure des unités

éliminer le besoin de deviner



5. Contrôlez votre géométrie

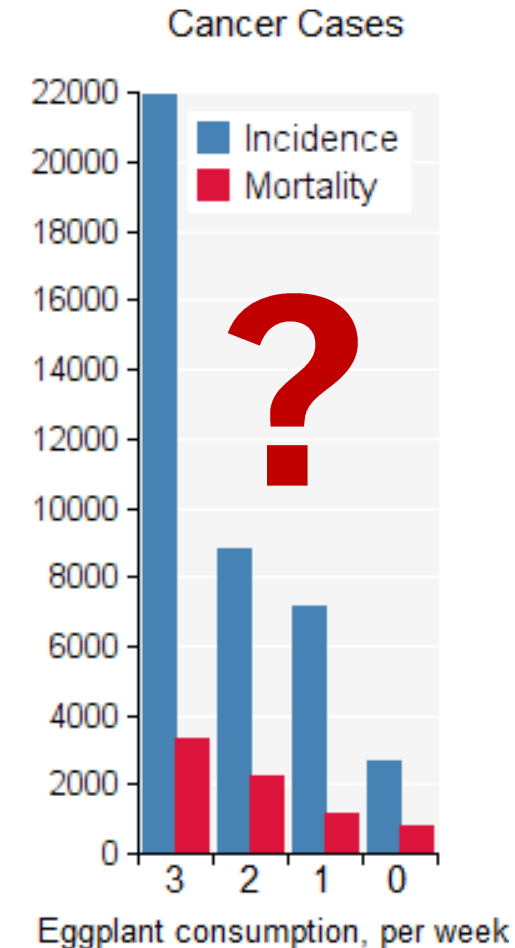
les cercles et les formes 2D sont dimensionnés par la surface, les diagrammes à barres par la longueur

6. Incluez vos sources

protégez-vous, laissez ceux qui veulent creuser davantage le faire

7. Tenez compte de votre public

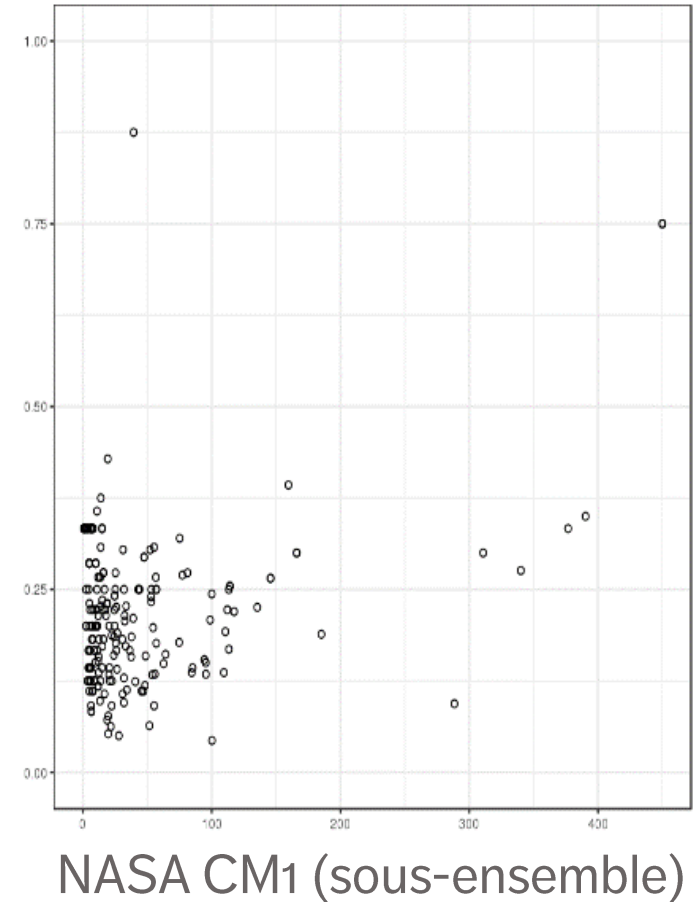
un poster peut être verbose, une présentation, minimaliste



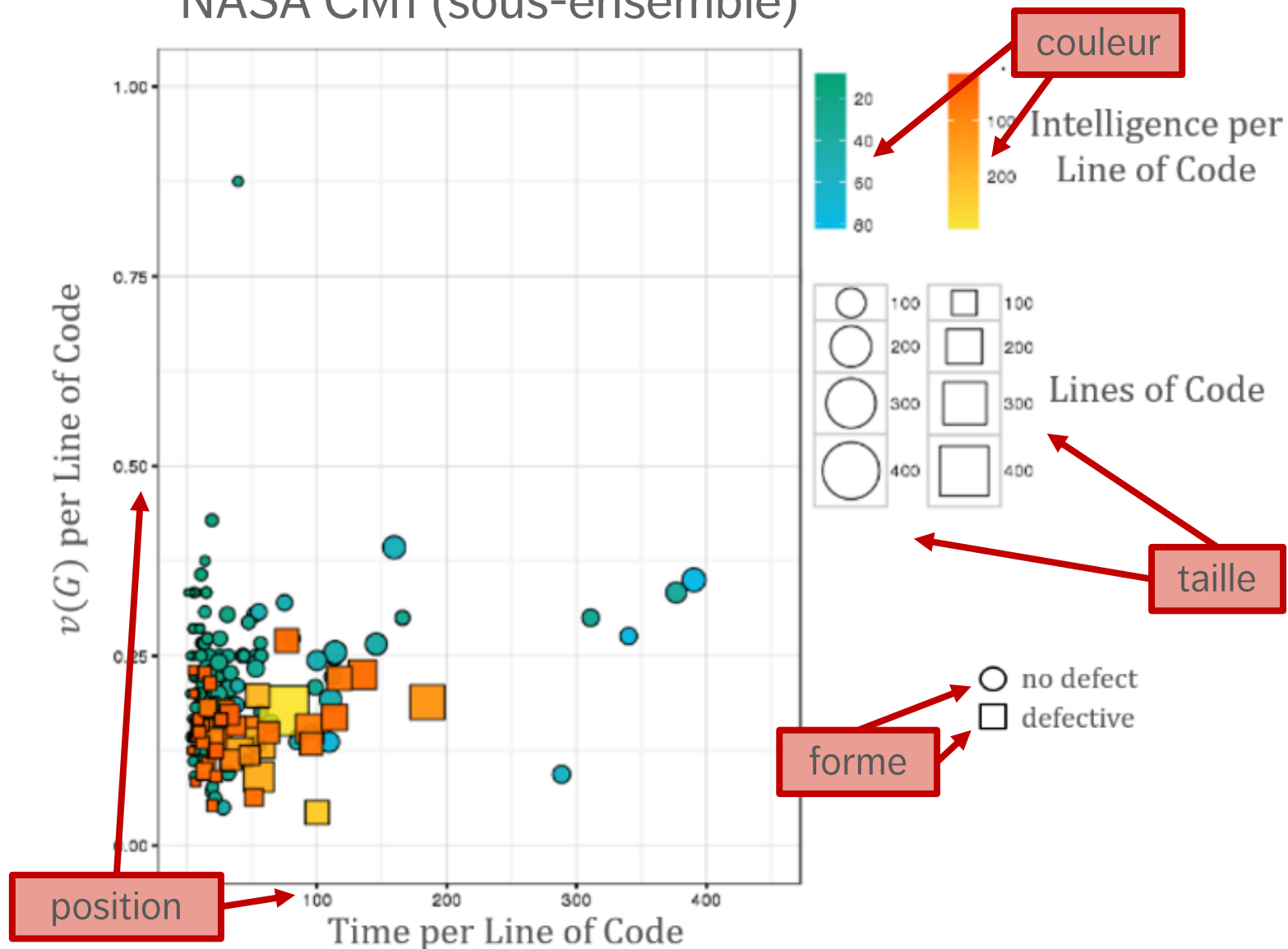
Les représentations multivariées

2 variables peuvent être représentées par la position dans le plan. Les facteurs supplémentaires peuvent être représentés par :

- taille
- couleur
- valeur
- texture
- orientation des lignes
- forme
- (mouvement ?)

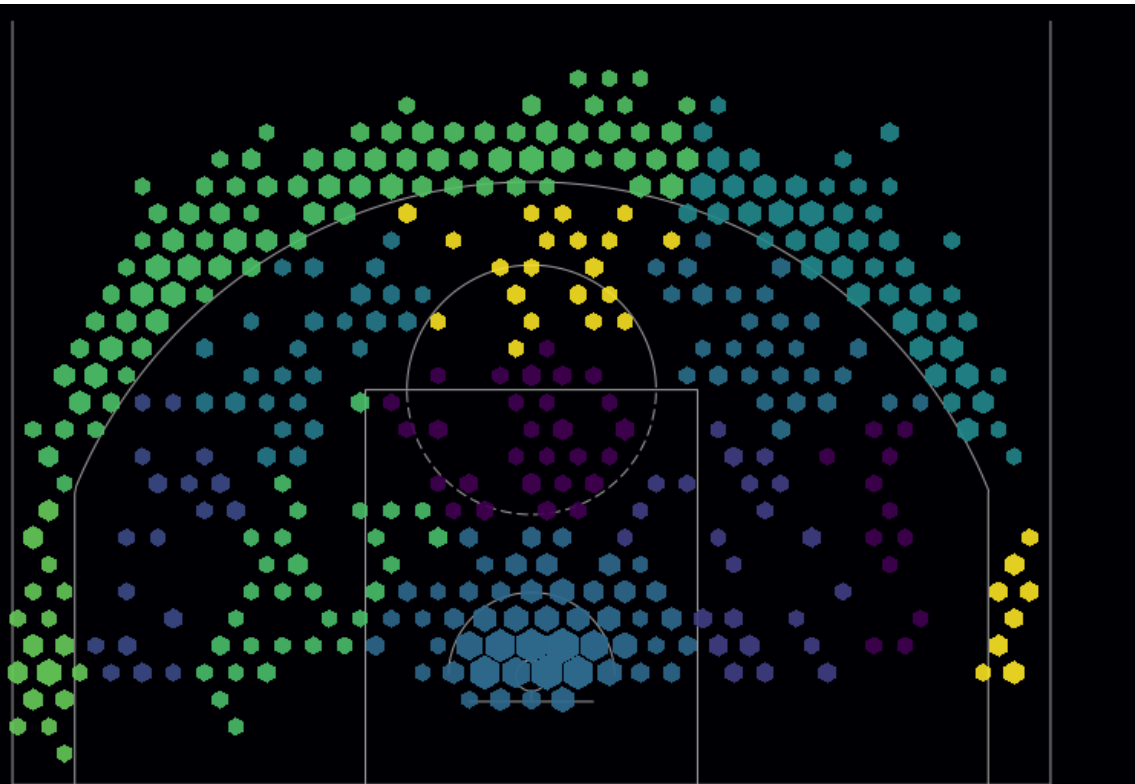


NASA CM1 (sous-ensemble)

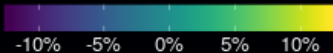


FG% contre la moyenne de la NBA ('15-'16)

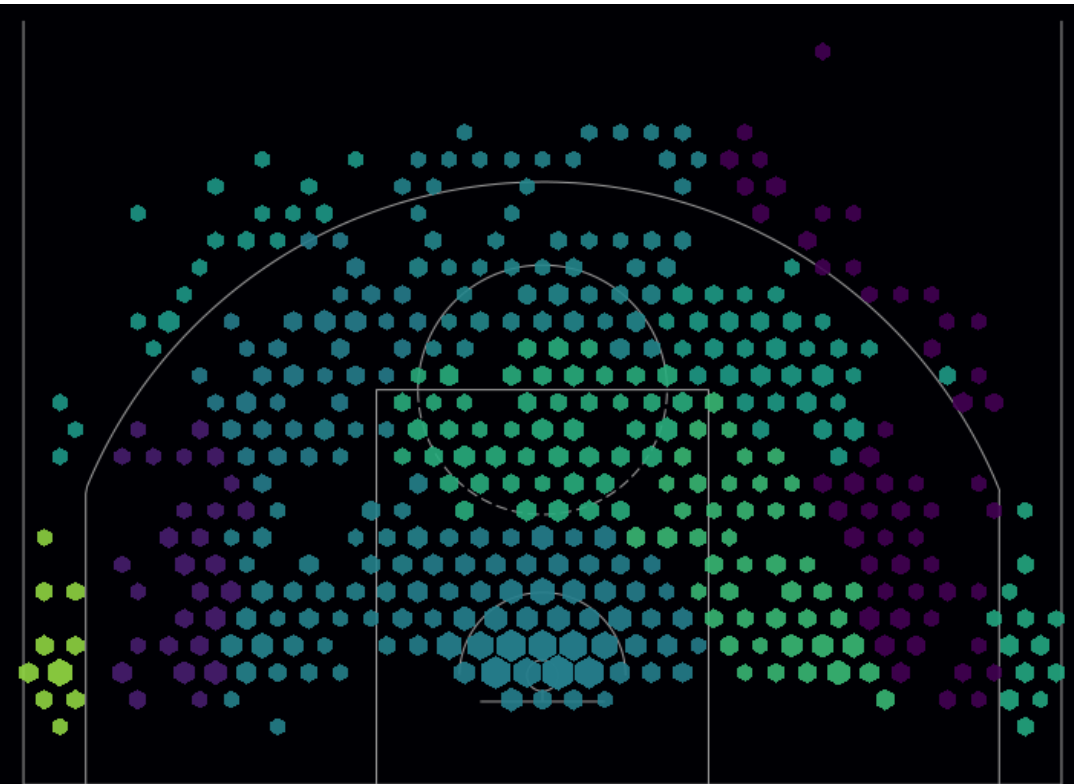
Kyle Lowry



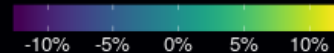
FG% vs. League Avg



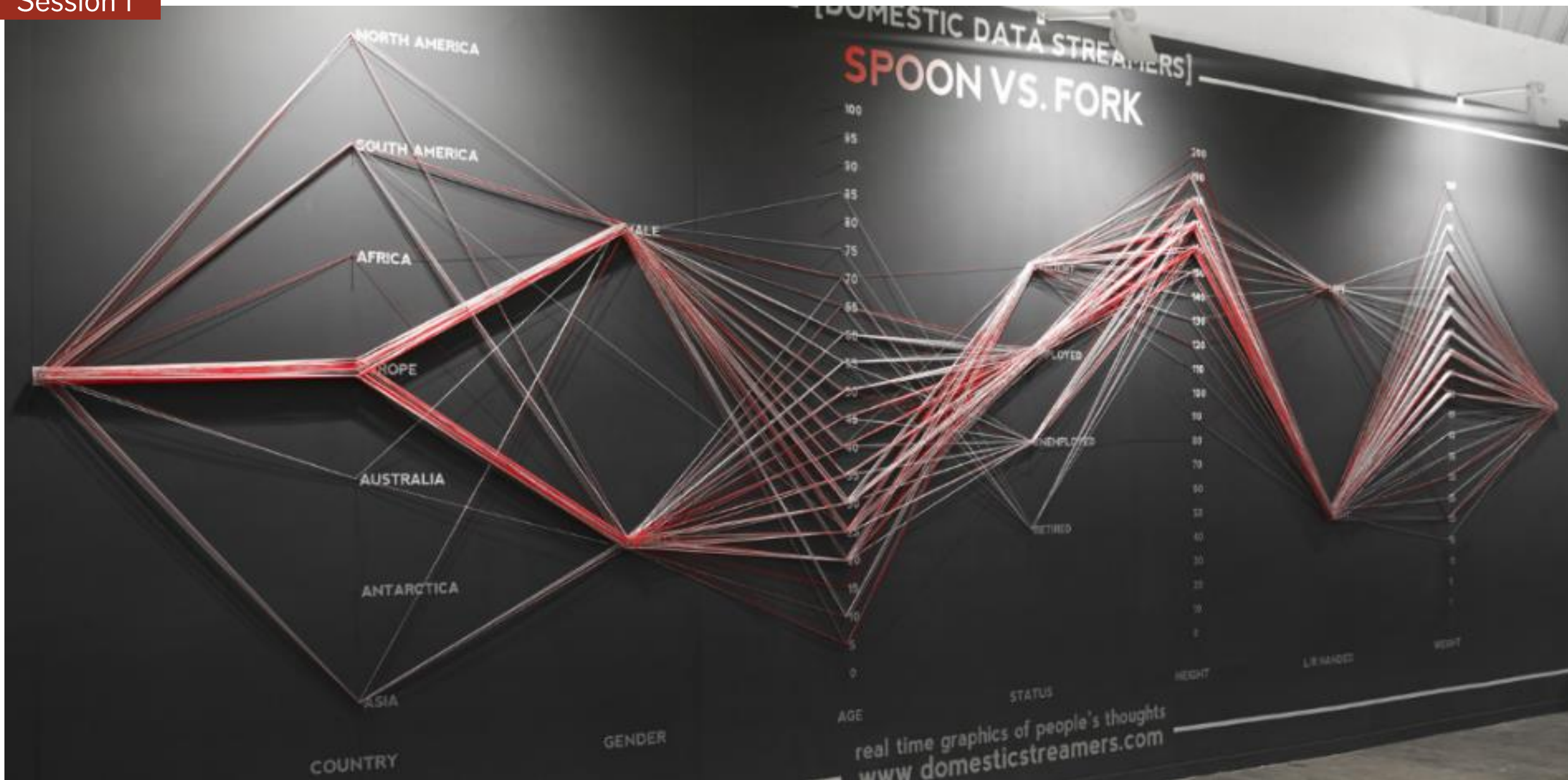
DeMar DeRozan



FG% vs. League Avg

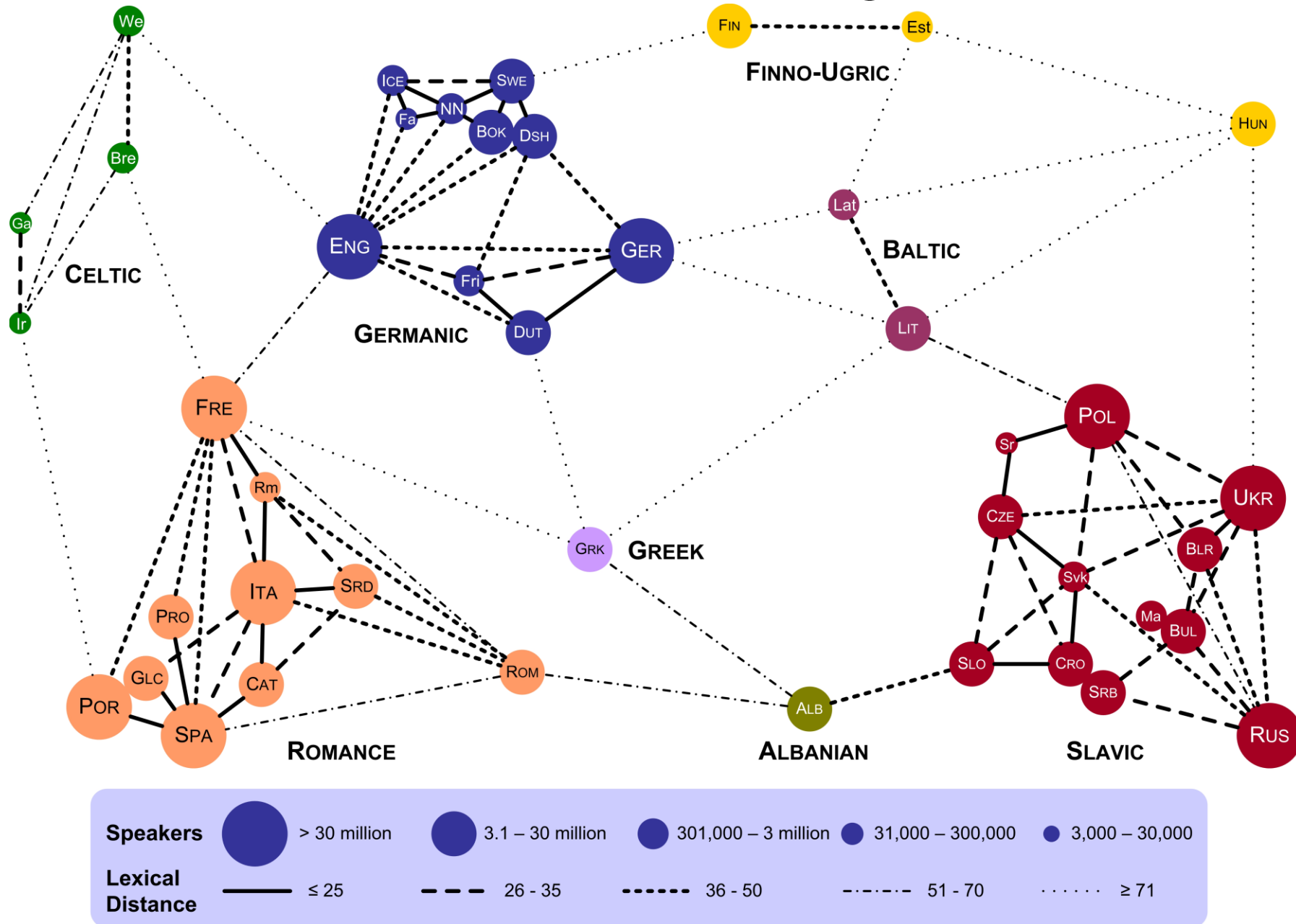


Quelles comparaisons pouvez-vous faire ? Comprenez-vous l'encodage ? Le contexte ?

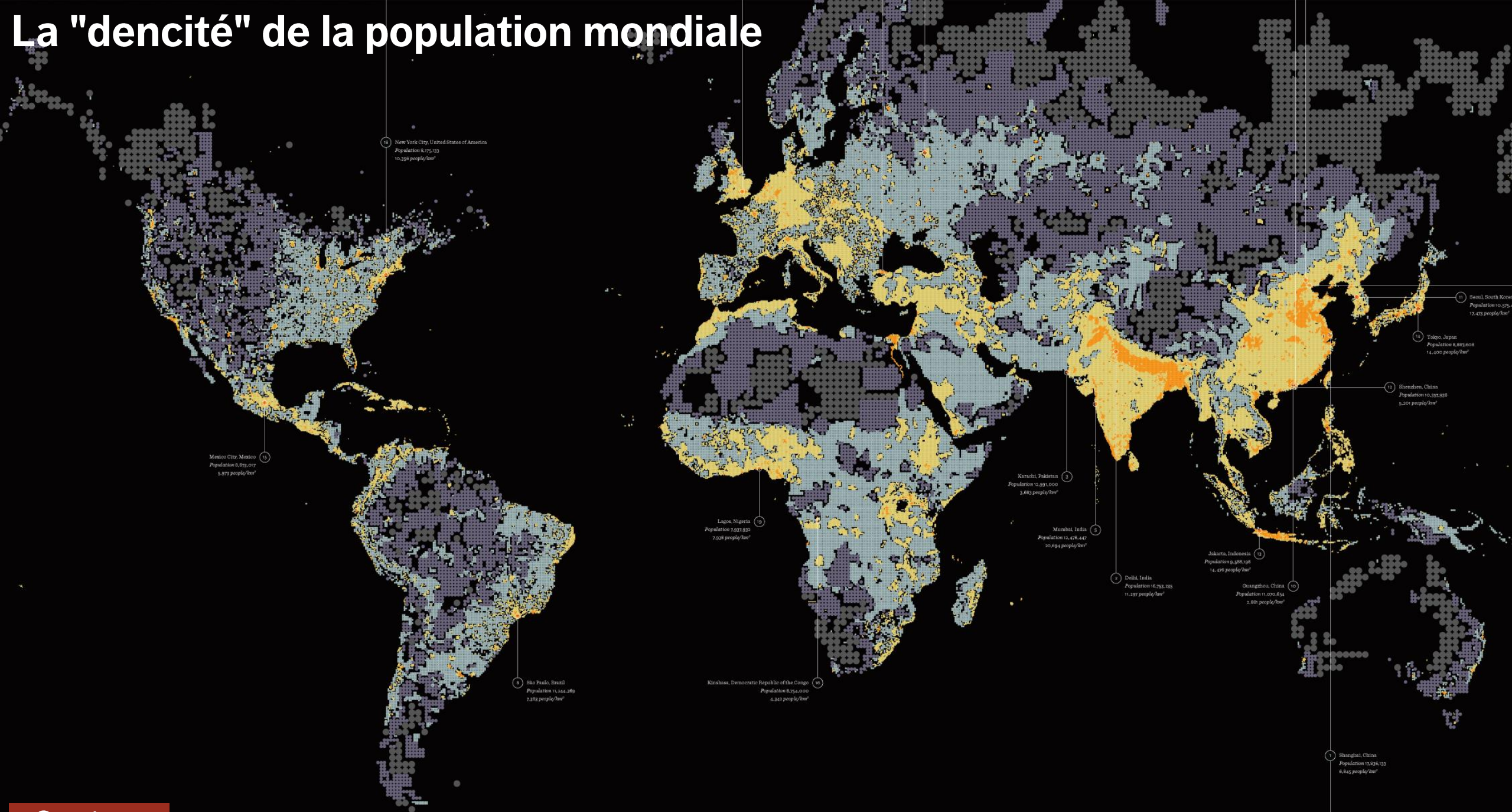


La collecte des données pose-t-elle des problèmes ? Où pensez-vous que cet événement a eu lieu ?
La question de la cuillère/de la fourchette est-elle un faux-fuyant ?

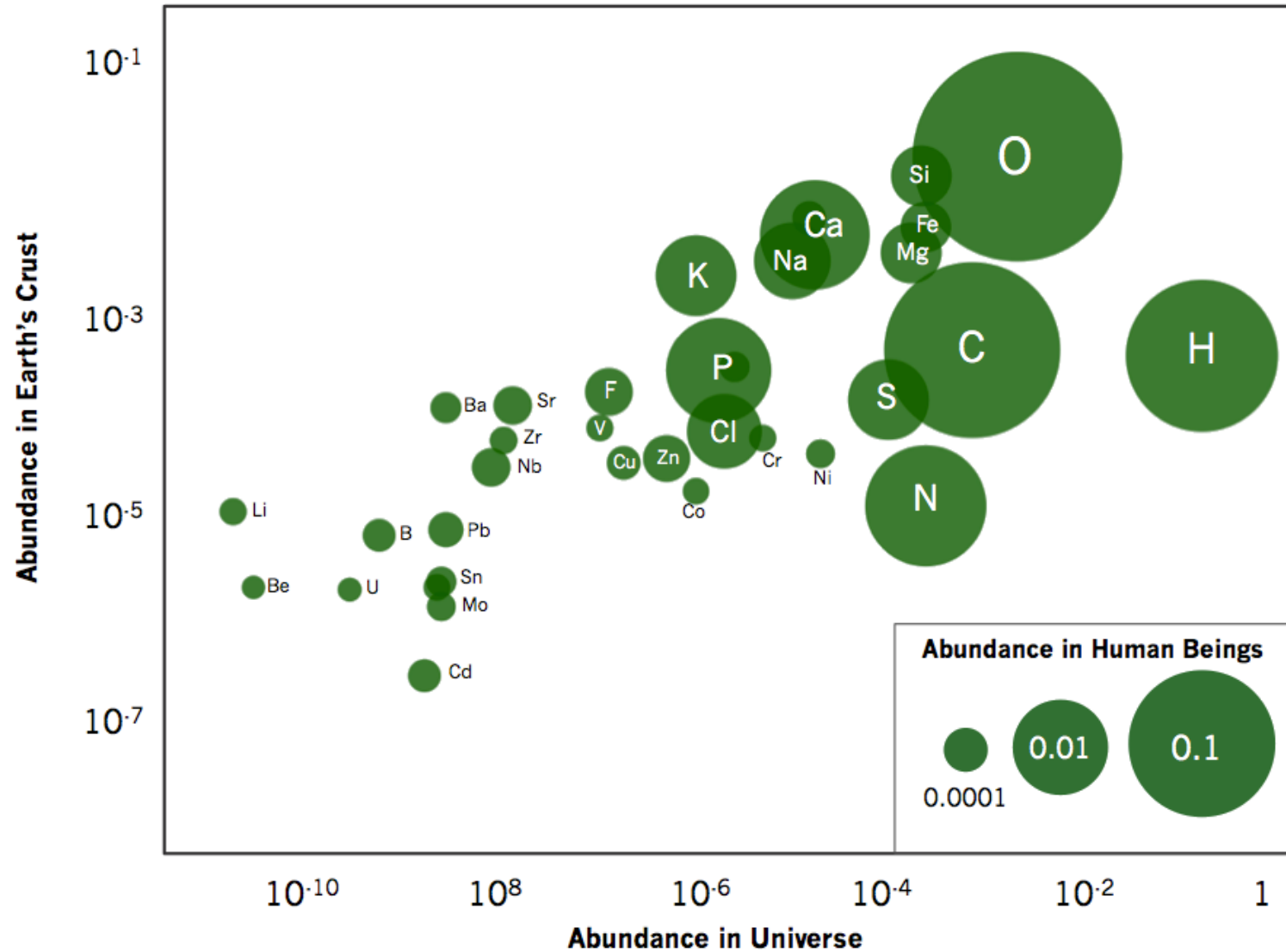
Distance lexicale entre les langues d'Europe



La "dencité" de la population mondiale



Abondance des éléments chimiques



Lectures conseillées

La communication et la visualisation
des données

Data Understanding, Data Analysis, Data Science **Volume 2: Fundamentals of Data Insight**

18. Data Exploration and Data Visualization

18.1 Data and Charts

- Presenting Results
 - Multivariate Elements in Charts
-

The Practice of Data Visualization **Part I: A Data Exploration Primer**

2. Data Exploration and Visualization

2.3 Representing Multivariate Observations

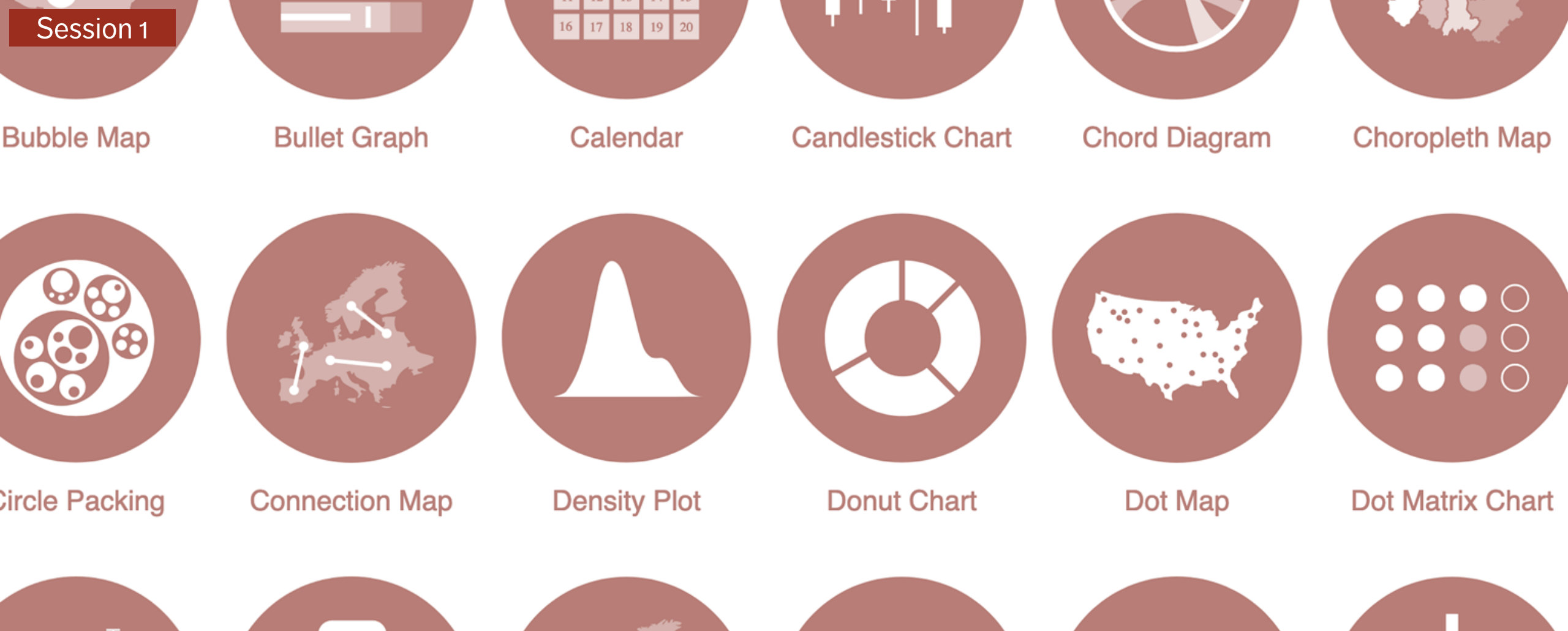
2.4 Communicating Analysis Results

3. Hall-of-Fame / Hall-of-Shame

Exercices

La communication et la visualisation
des données

1. Étudiez les graphiques présentés dans cette section (ou tout autre graphique de votre choix) en utilisant les principes de la conception analytique. Les règles de base sont-elles respectées ?
2. Quels éléments graphiques/de conception pourraient être ajoutés pour améliorer les présentations ?



3. Un catalogue de visualisations

Data comparison charts

Data reduction charts

Comparison

Composition

Distribution

Evolution

Relationship

Profiling

Bars



Pie



Histogram



Line



Scatterplot



Grouped bars



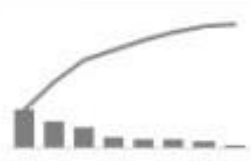
Dot plot



Bullet



Pareto



ID Scatterplot



Horizon



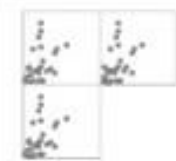
Connected Scatterplot



Cycle plot



Scatterplot matrix



ID Scatterplot



Heat map



Multidimensional Pie



Boxplot



Step



Bubble



Reorderable matrix



Horizon



Slope



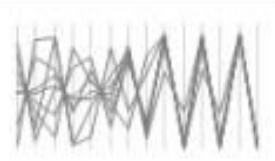
Alert



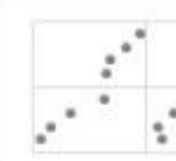
Connected Scatterplot



Parallel Plot



Trellis



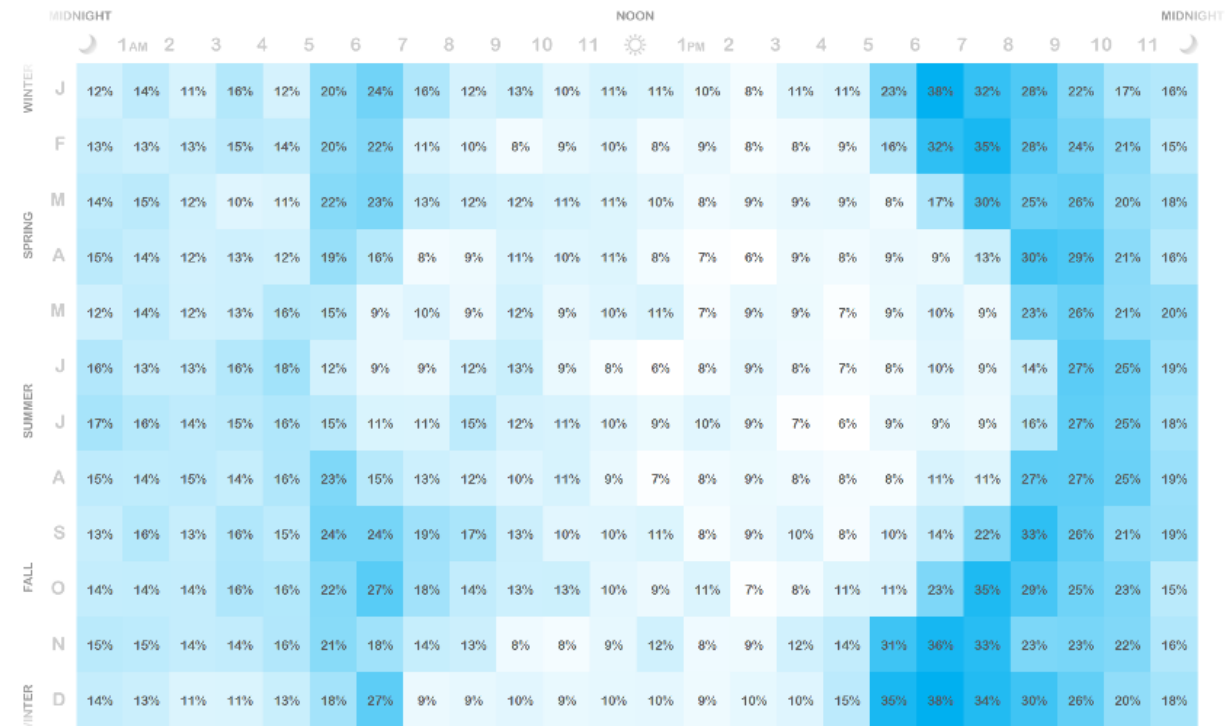
Les cartes thermiques (“heatmaps”)

L'horizon du risque – piétons

Le taux d'incidents de circulation mortels impliquant des piétons, à chaque heure de la journée, tout au long des saisons de l'année.

Le décalage saisonnier de notre soleil couchant trace une arche de risque élevé - un écho de la courbe de la Terre, elle-même (Note : ???).

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

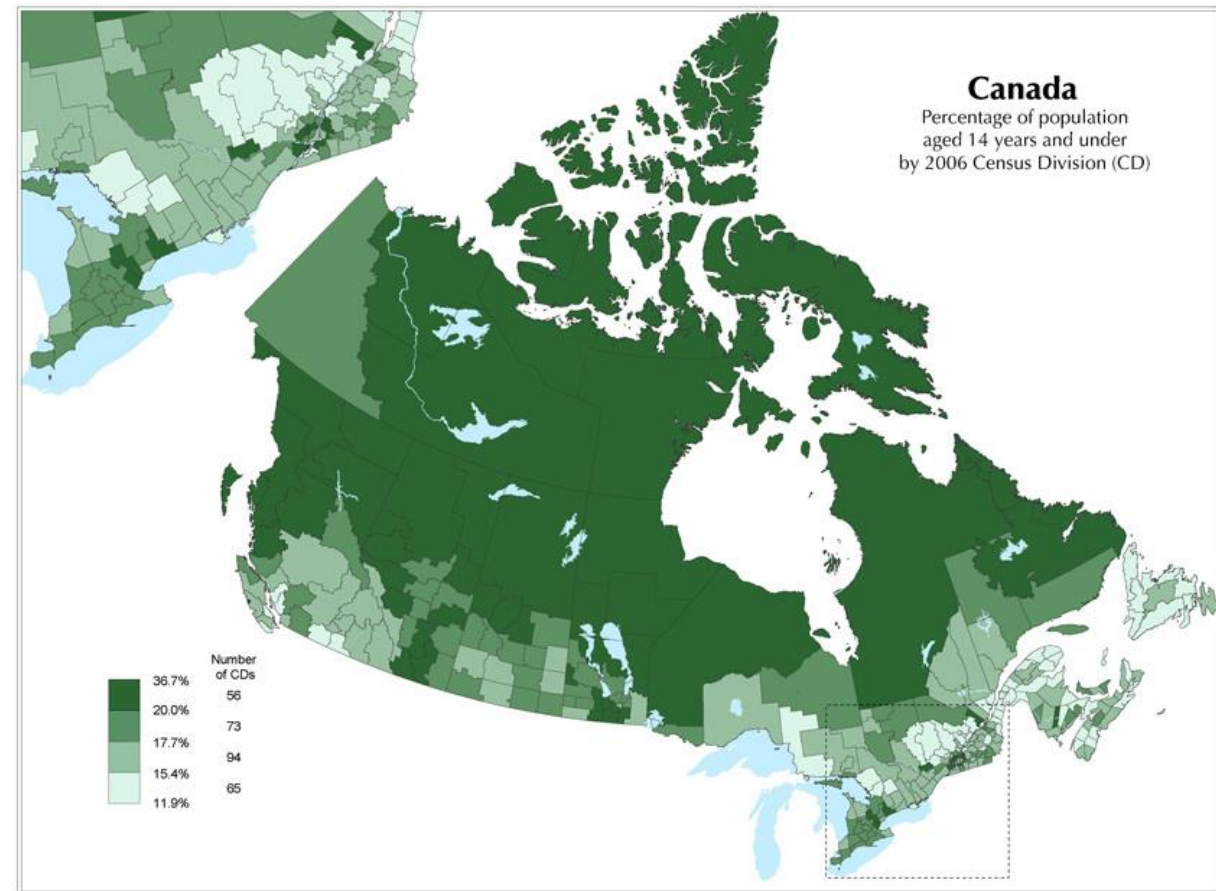
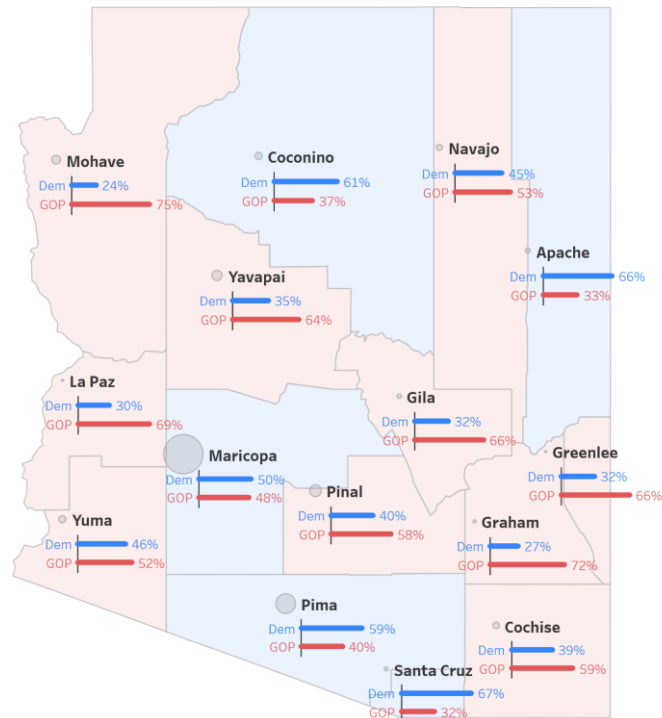


Les cartes thermiques (choroplèthes)

2020 Arizona Presidential Election Results by County

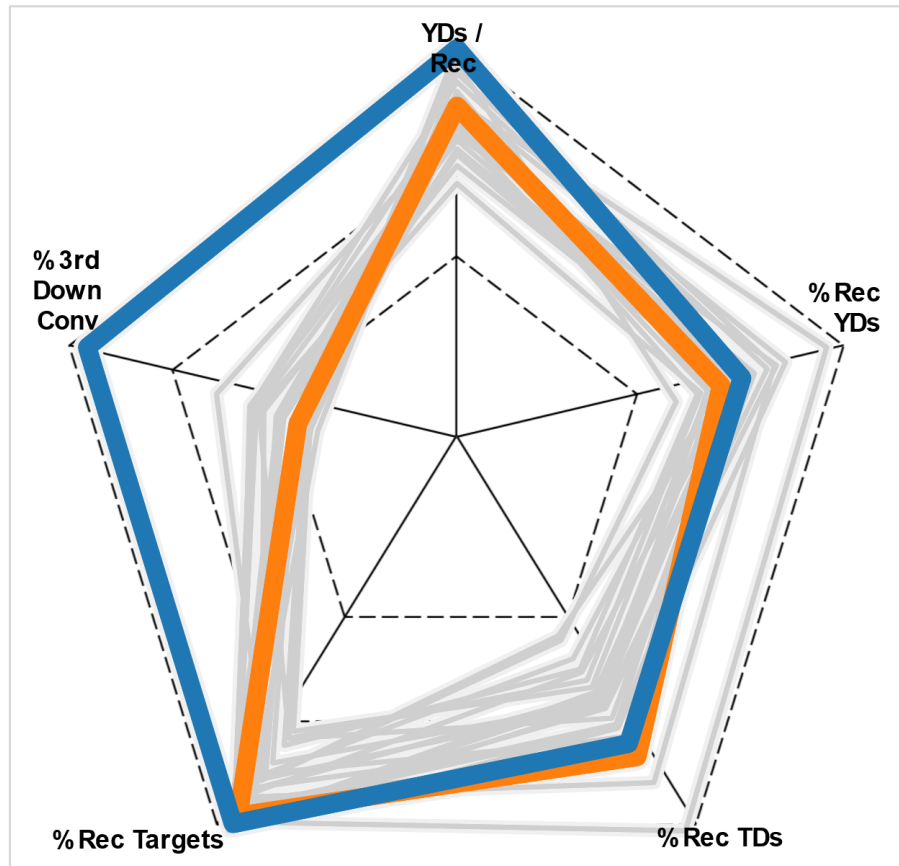
County color = 2020 winner. Blue = Dem. Red = GOP. Source NY Times

Change Year —●— 2020

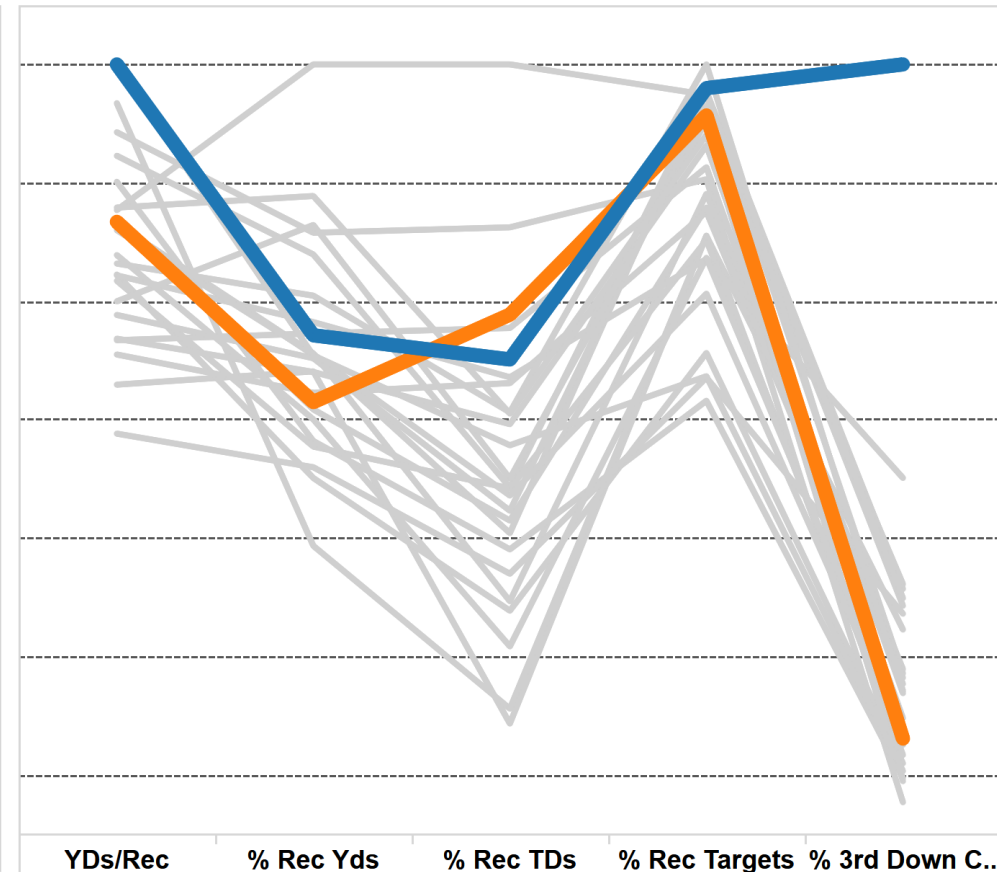


Les coordonnées parallèles/radiales

Radar Chart



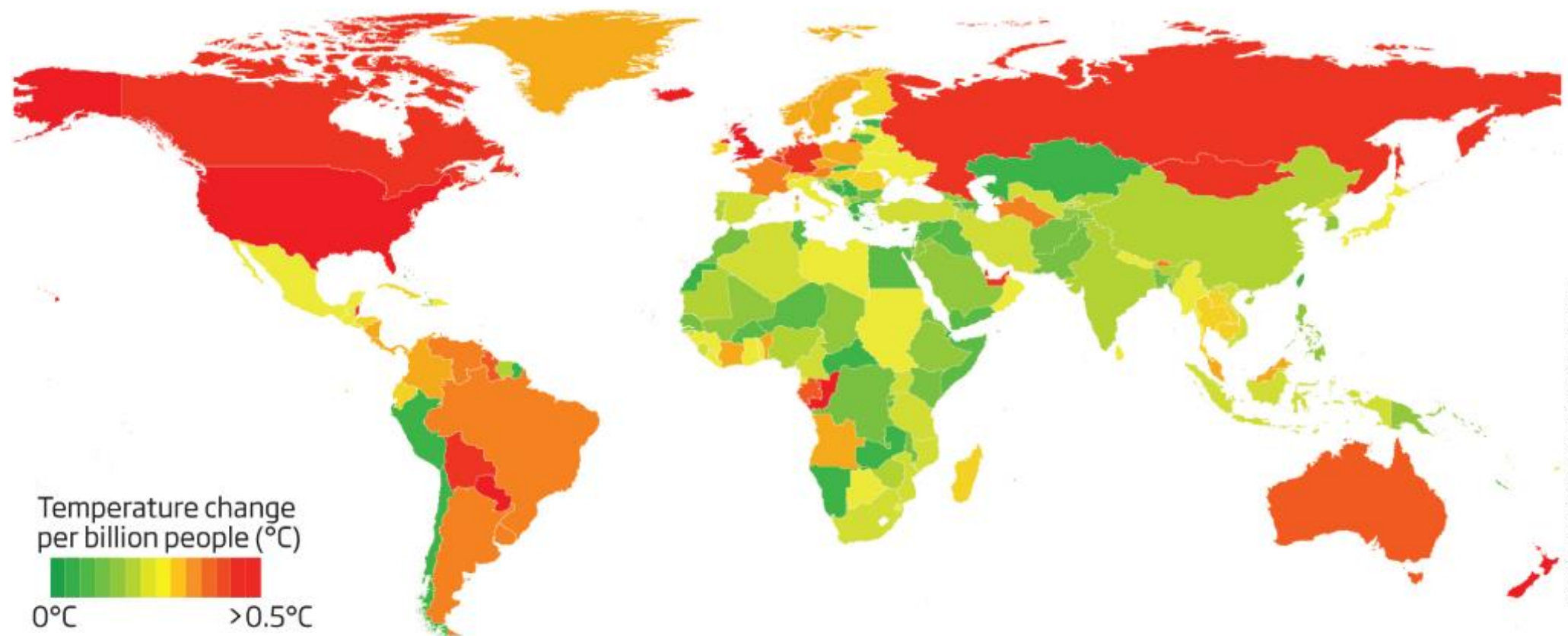
Parallel Coordinate



Les cartes géographiques

Global warming culprits, judged by population

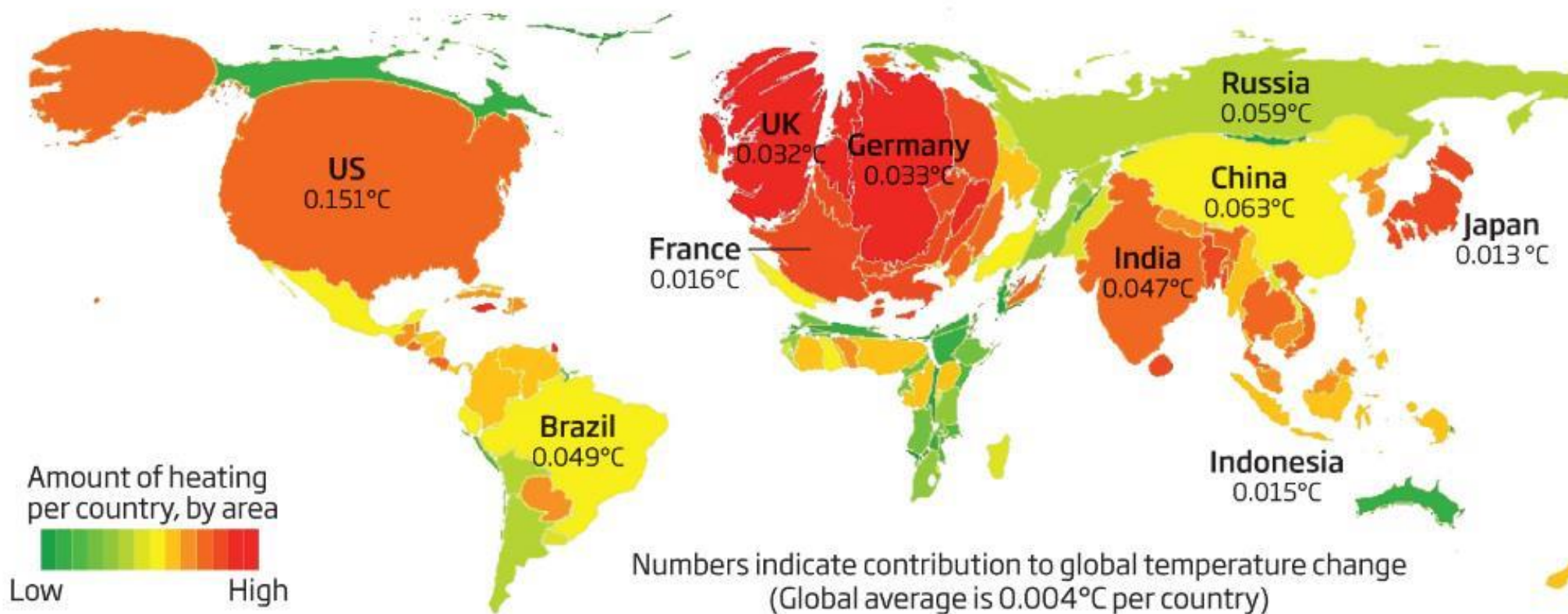
Countries that have caused more global warming per billion people are coloured red and low-emitters are dark green



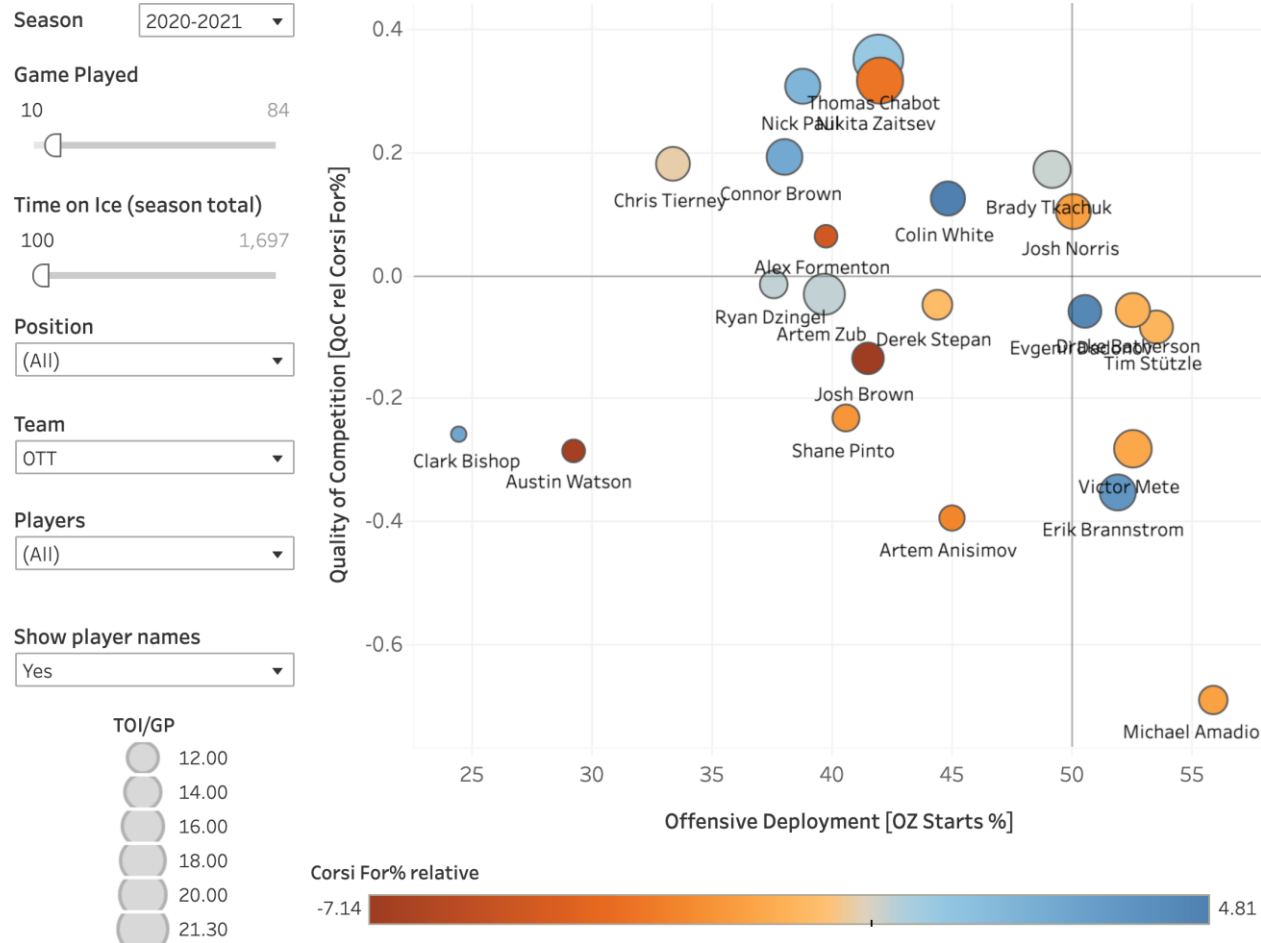
Les cartes géographiques déformées

Global warming culprits, judged by size

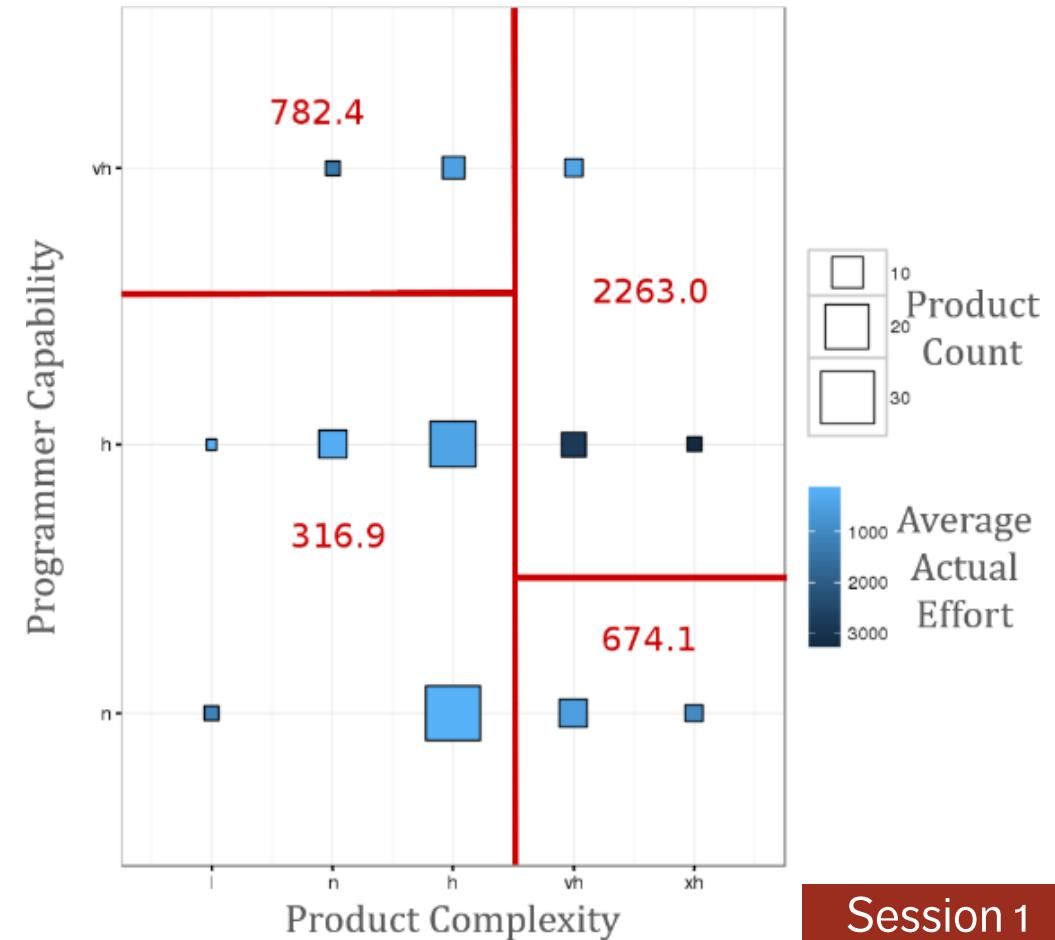
Countries that have caused disproportionately more global warming than their area would suggest are shown swollen, while low-emitters in relation to their size are shrunk



Les graphiques à bulles

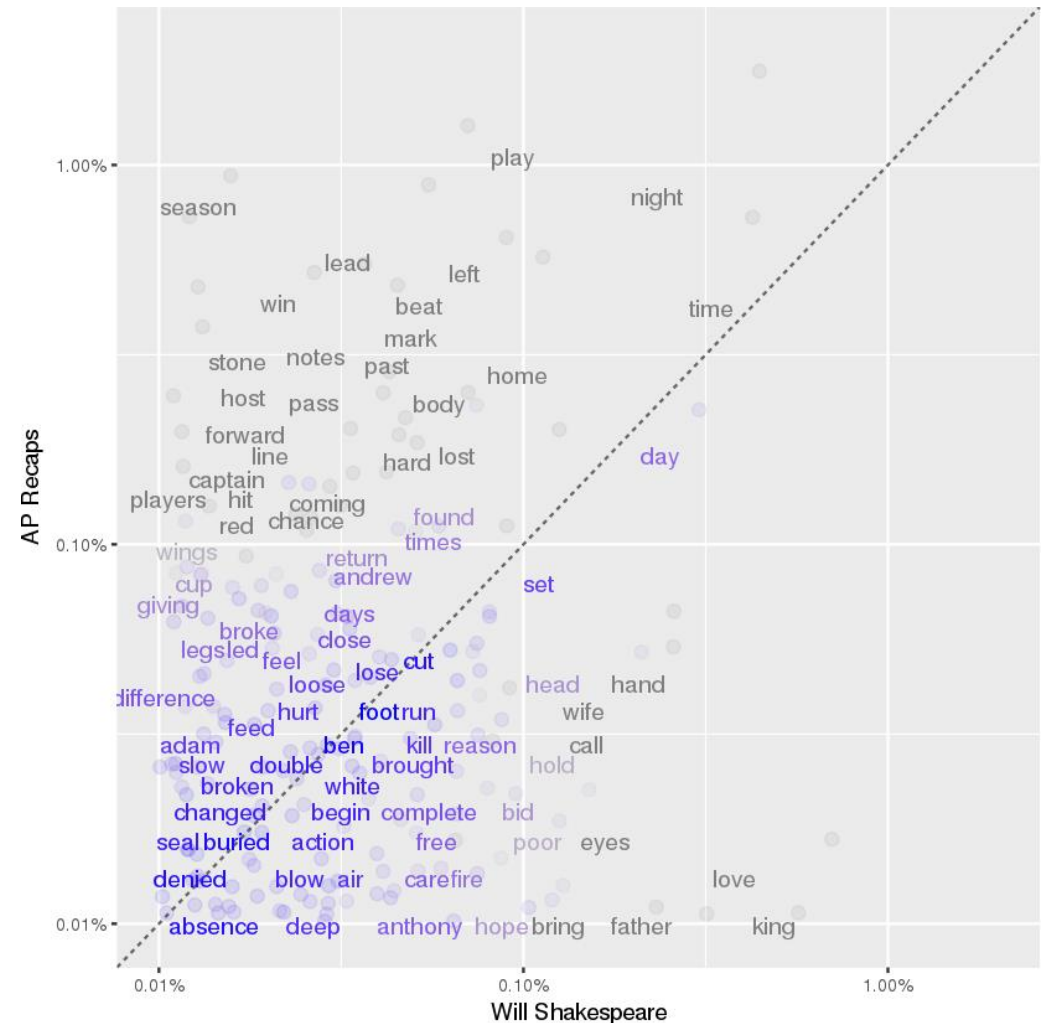
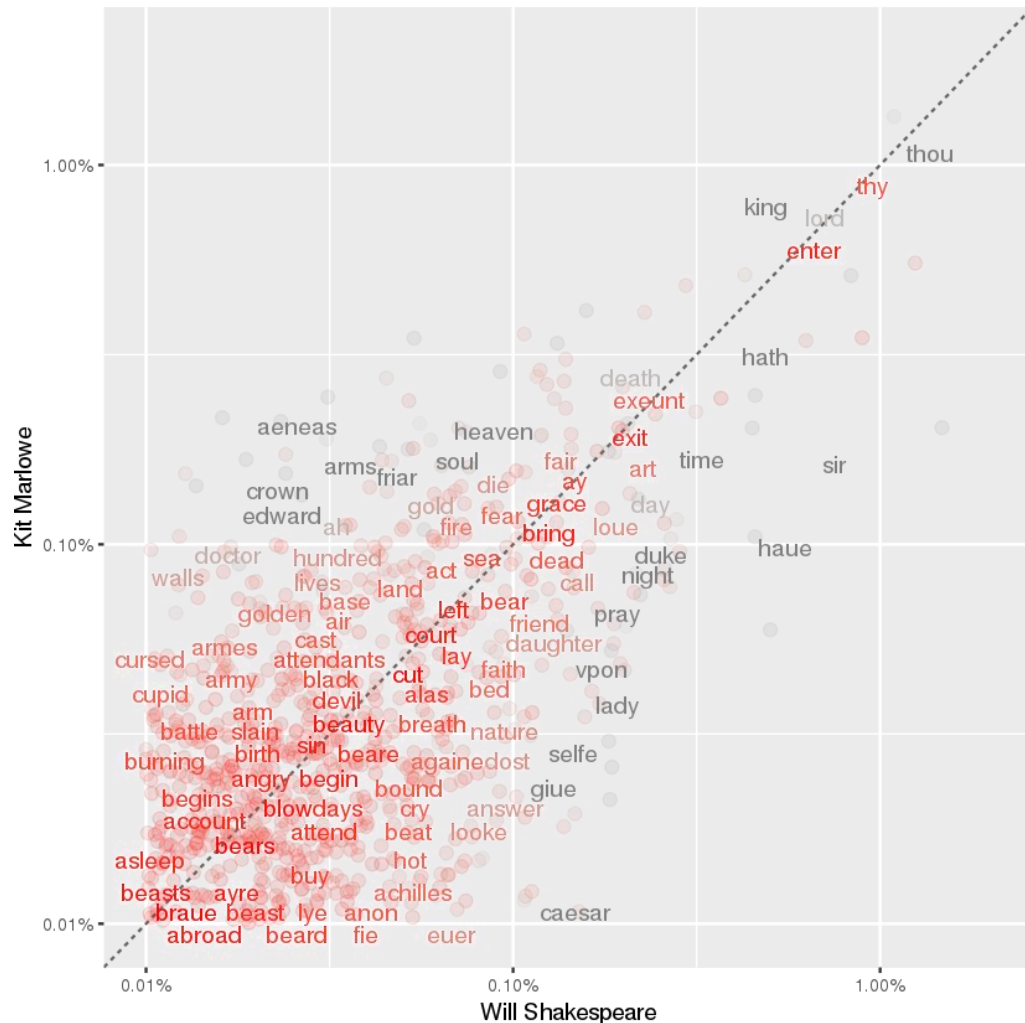


Ensemble de données
COCOMO de la NASA

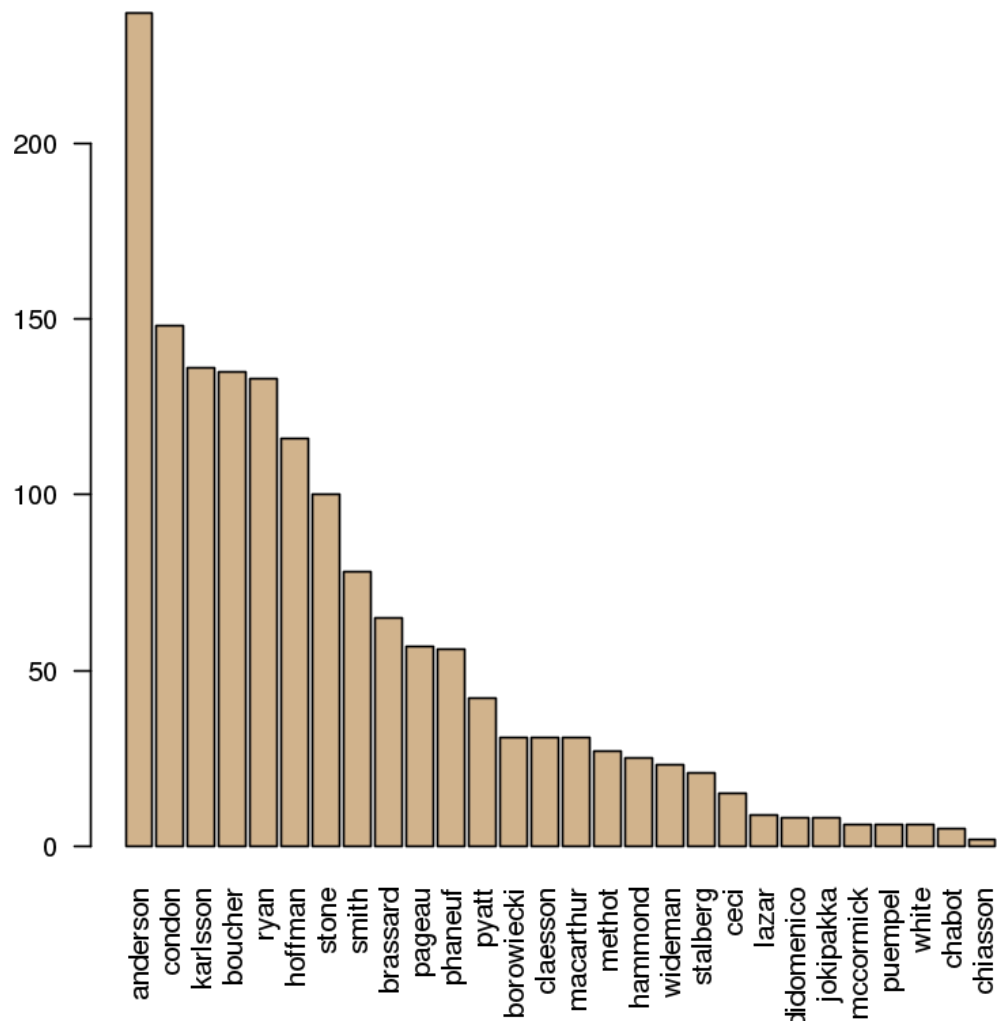


Session 1

La visualisation et la représentation de texte

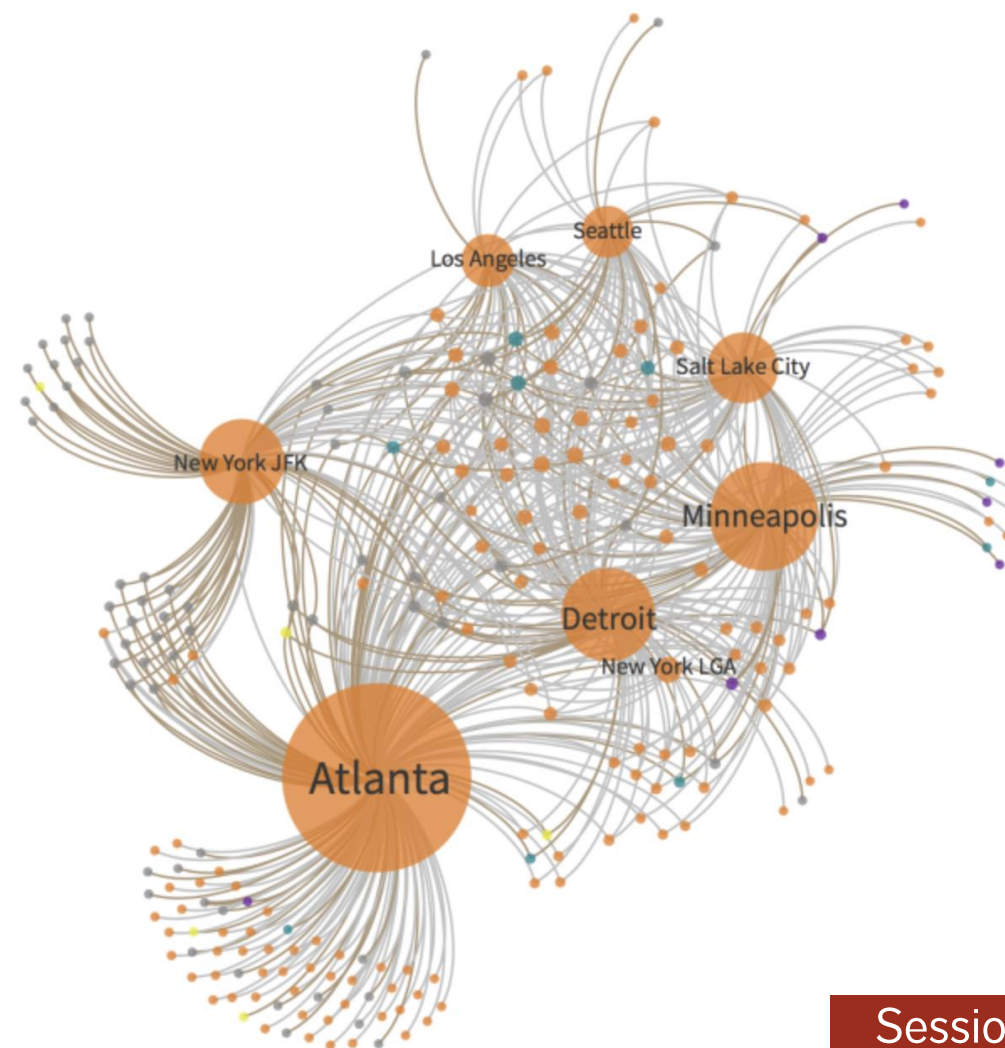
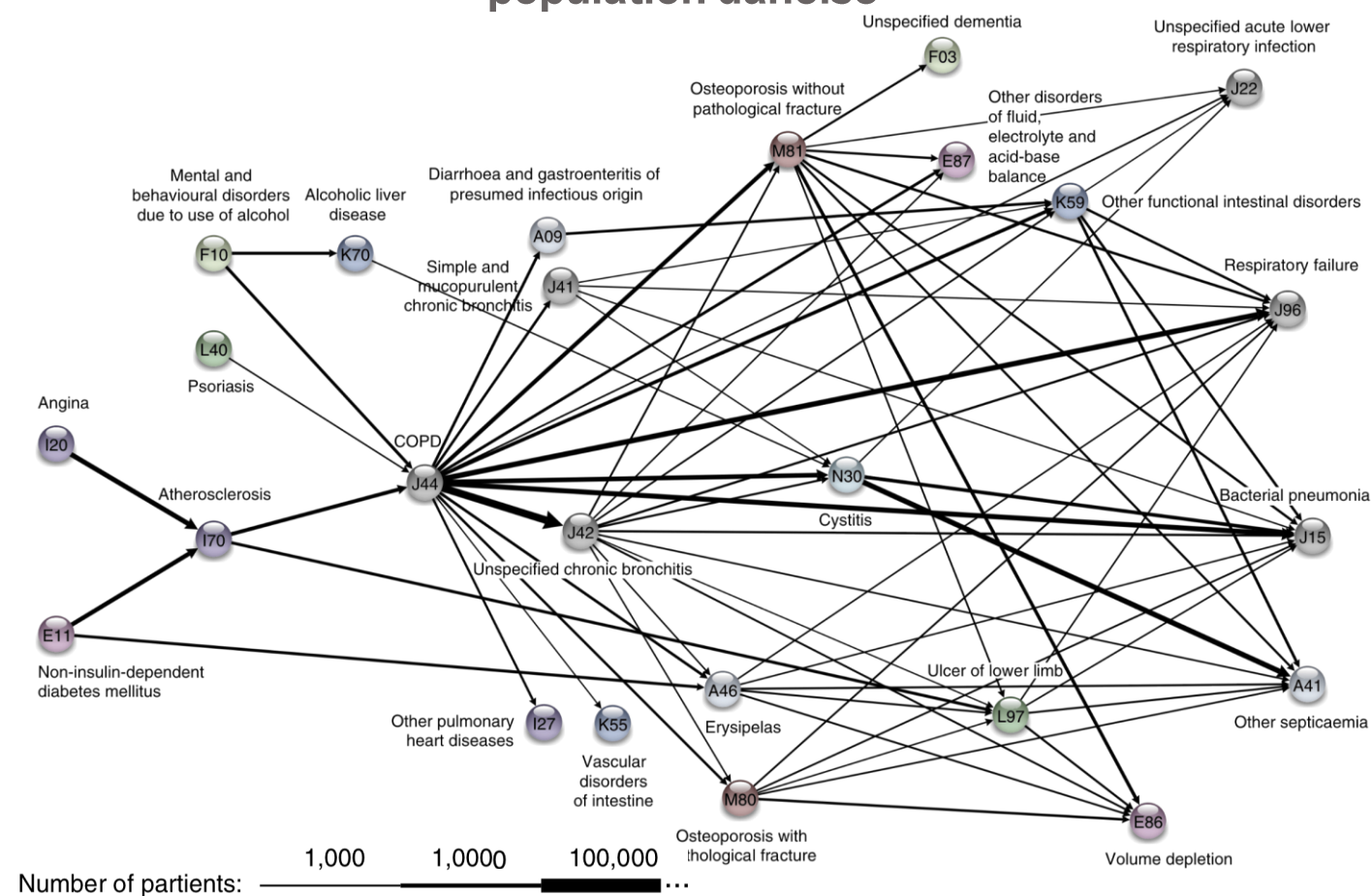


La visualisation et la représentation de texte



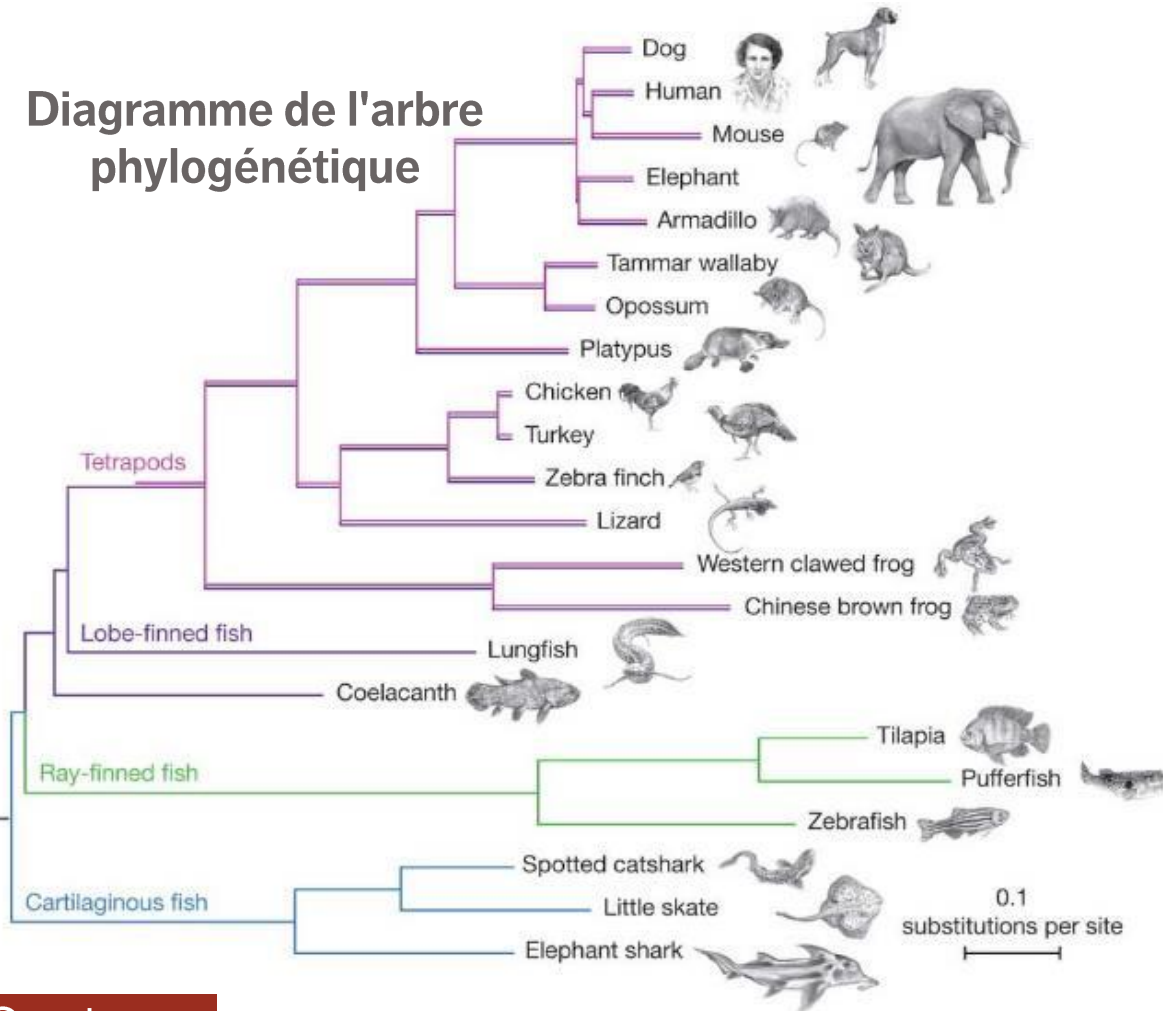
Les diagrammes de réseau

Parcours de diagnostic autour de la BPCO dans la population danoise

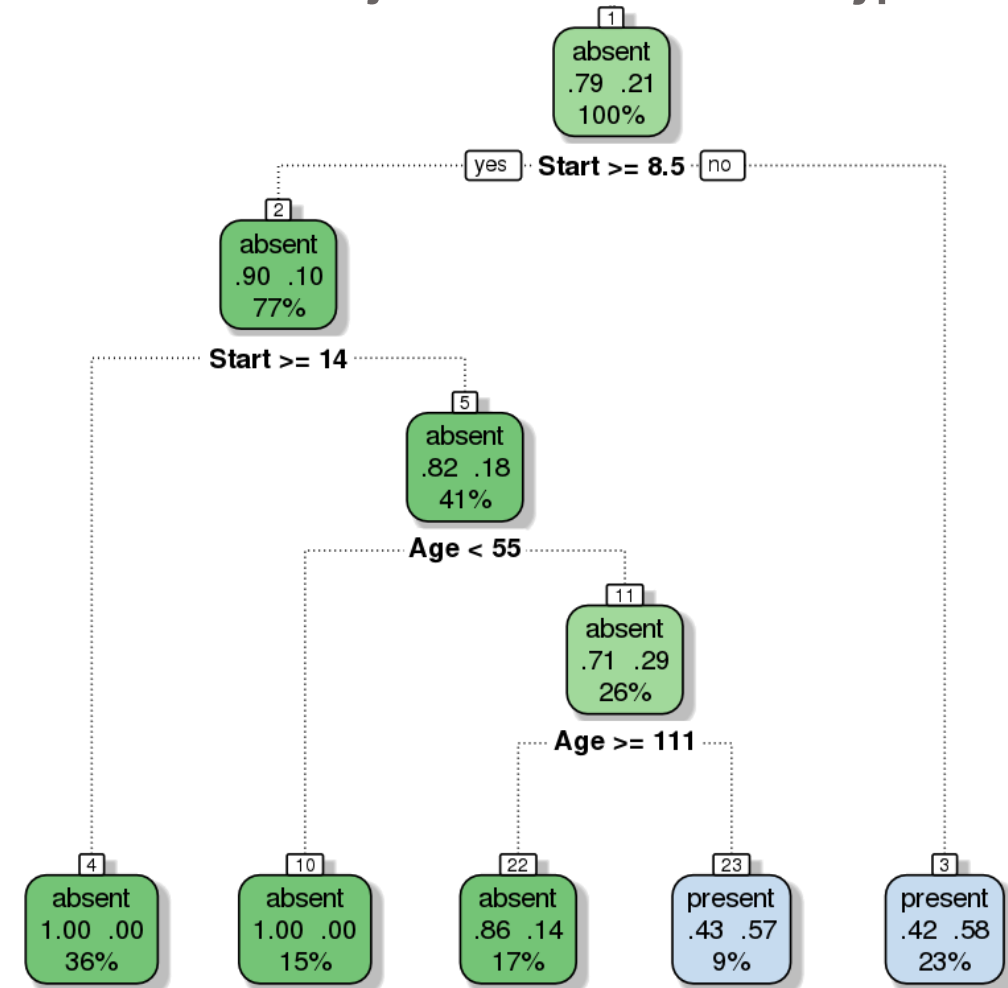


Les arbres décisionnels et les dendrogrammes

Diagramme de l'arbre phylogénétique



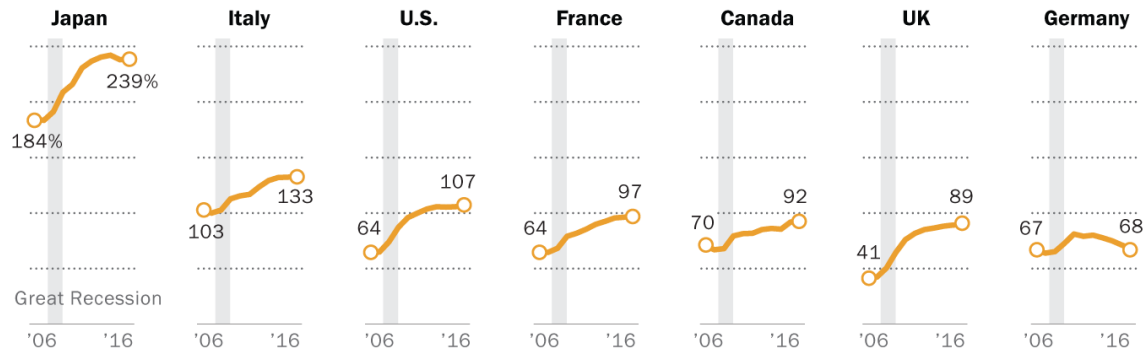
Classification du jeu de données sur la cyphose



Les petits multiples

After Great Recession, debt increased substantially in most G-7 economies

Total gross debt as a share of GDP in the Group of Seven nations

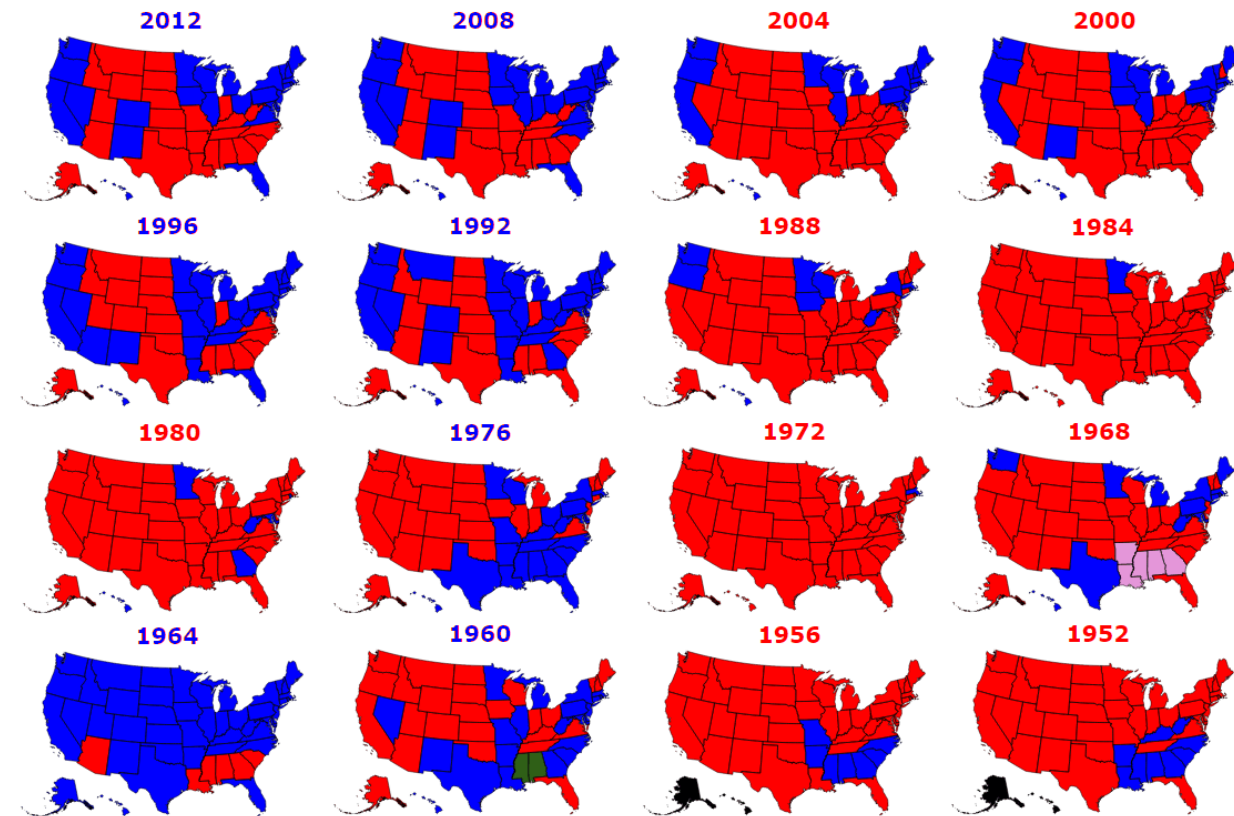


Note: Gross debt represents total liabilities of all levels and units of government — national, state/provincial and local — less liabilities held by other levels or units of government, unless otherwise noted by source.

Source: The International Monetary Fund, World Economic Outlook, accessed Sept 7, 2017.

PEW RESEARCH CENTER

Résultats du collège électoral américain 1952 - 2012



Les “sparklines”

TOTAL

Hôpital n° 1

Hôpital n° 2

Hôpital n° 3

Hôpital n° 4

Hôpital n° 5

Hôpital n° 6

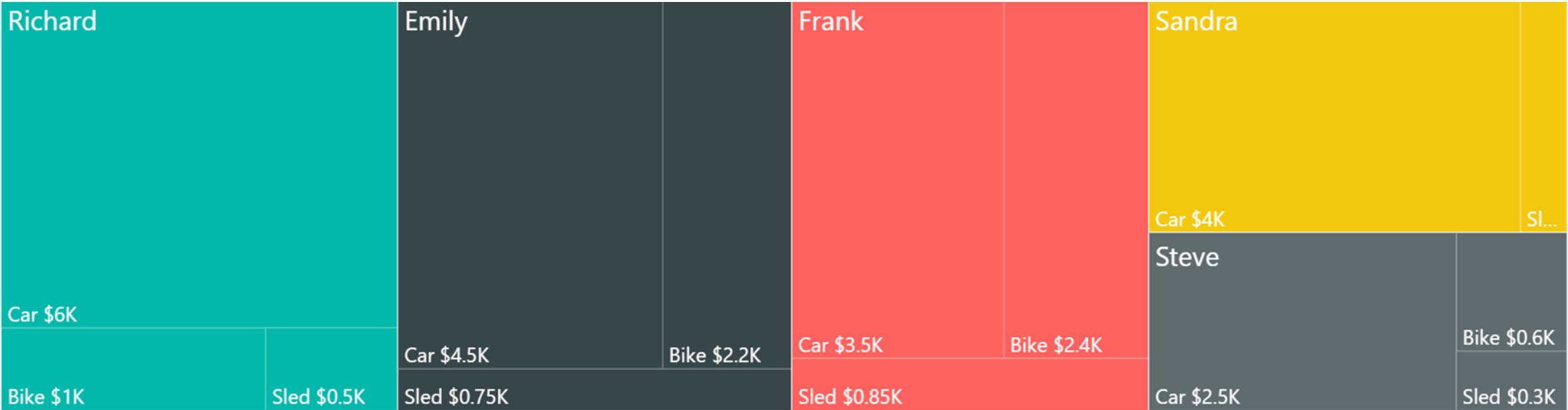
Hôpital n° 7

Hôpital n° 8

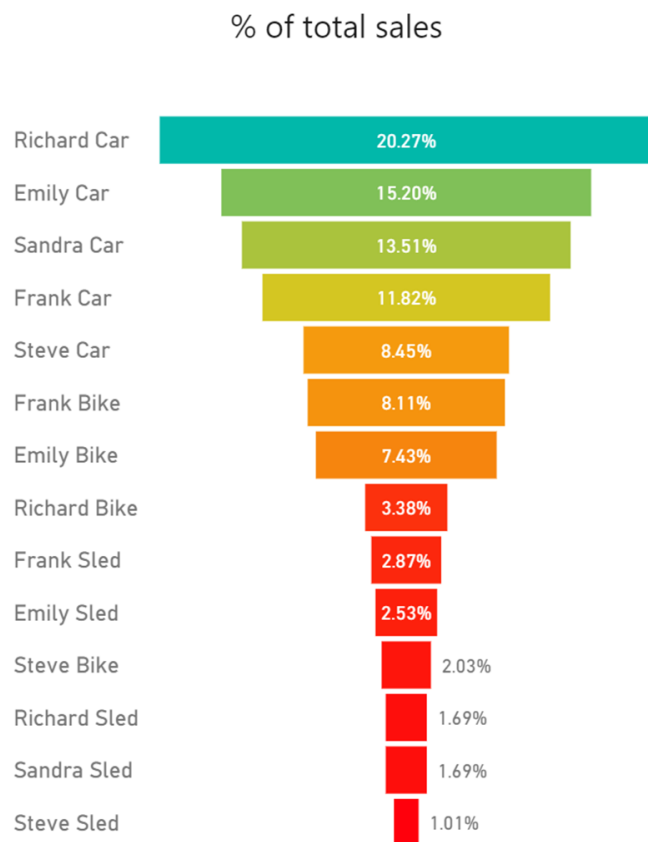
...

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

Les “treemaps”



Les graphiques en entonnoir

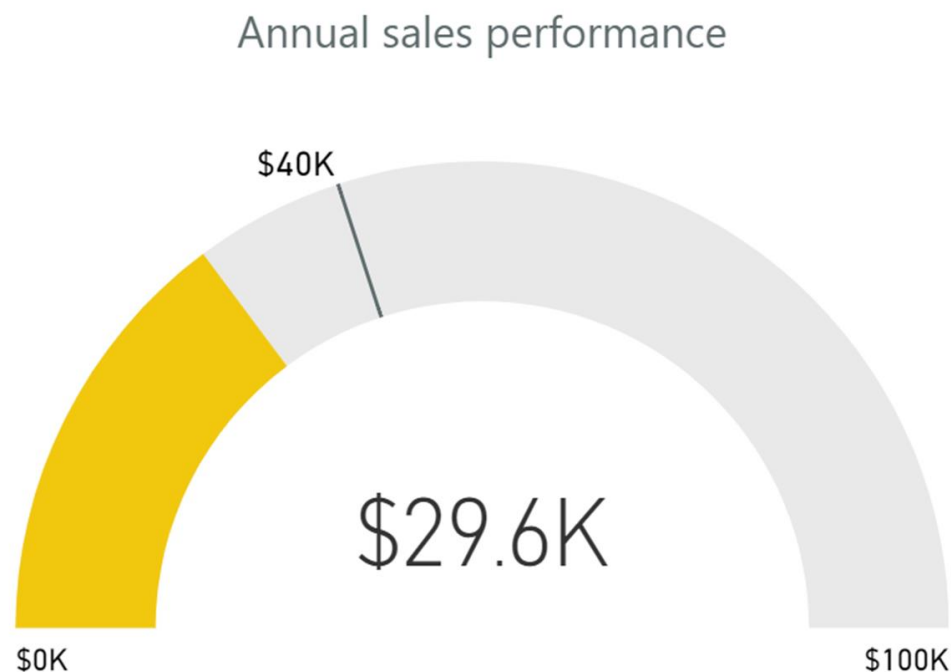


Ils représentent généralement des **proportions décroissantes**, jusqu'à un total de 100% (souvent).

Si ce n'est pas l'ordre par défaut, les utilisateurs doivent trier de haut en bas manuellement.

TRÈS utiles pour aider le public à classer rapidement les éléments par ordre de priorité sans avoir à filtrer activement.

Les graphiques de jauge



Souvent utilisés comme élément du tableau de bord (avec ou sans aiguille).

Affichent des mesures à valeur unique menant à un objectif / indicateur de rendement clé.

Idéal pour montrer le progrès (un “fad” de gestion, cependant...)

Affiche des informations qui peuvent être rapidement **scannées** et **comprises**.

Les visualisations interactives et animées

L'animation **n'améliore pas toujours** une visualisation. Quels aperçus l'interactivité peut-elle apporter ? Cela dépend des données, et de la méthode de visualisation.

Exemples :

- [The Clubs That Connect the World Cup](#), NY Times, 2014
- [Who Marries Whom](#), Bloomberg, 2016
- [Hipparcos Star 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

Les 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 \$\zeta\$ 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

Les visualisations interactives et animées

"Il y a toujours un danger que si certains types de techniques de visualisation prennent le dessus, les types de questions qui sont particulièrement bien adaptées à la collecte de données pour ces techniques en viennent à dominer le paysage, ce qui affectera ensuite les techniques de collecte de données, la disponibilité des données, l'intérêt futur, et ainsi de suite." (P. Boily)

Même lorsqu'elles sont bien faites, 85% des utilisateurs ne se soucient pas des visualisations interactives (NY Times).

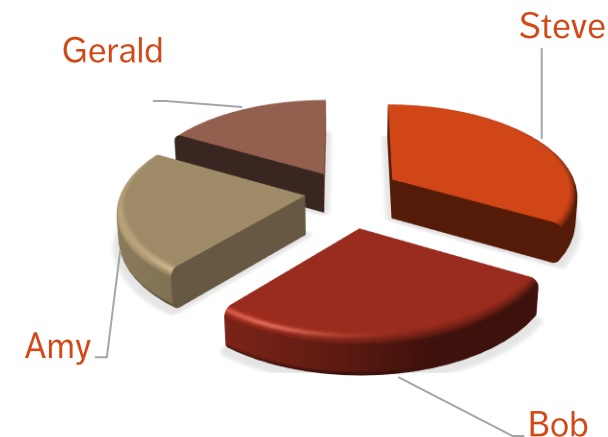
À garder en tête : explorez les données et essayez différentes méthodes

Des graphiques à éviter

TOUT ce qui a un arc (sauf la jauge)

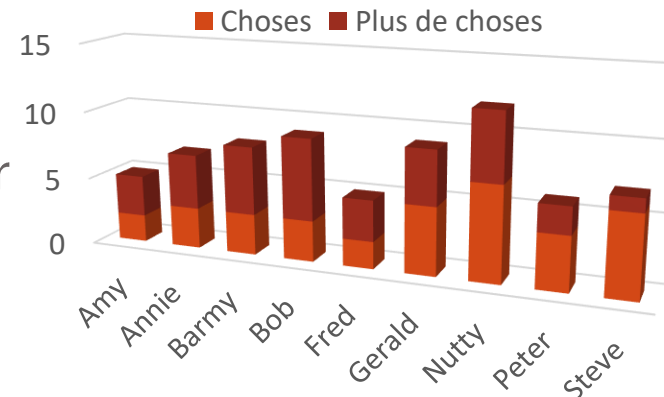
- tarte
- beignet

Le cerveau ne peut pas comparer les arcs et ils peuvent être trompeurs : quelle est la différence entre Steve et Bob dans le diagramme circulaire ?



TOUT 3D EST MAUVAIS !

- comme pour les arcs, nous ne pouvons pas facilement comparer visuellement les séries de données
- le 3d ajoute beaucoup trop d'encombrement



Lectures conseillées

Un catalogue de visualisations

The Practice of Data Visualization

Part IV: Some Practical Aspects of Data Visualization

9. Visual Toolbox



4. L'esthétique des graphiques

Les principes de la Gestalt

Les **principes de la Gestalt** forment les "lois" de la perception humaine.

Ils décrivent la manière dont les humains regroupent des éléments similaires, reconnaissent des modèles et simplifient des images complexes lorsqu'ils perçoivent des objets.

Les graphistes les utilisent pour organiser le contenu des graphiques, des tableaux de bord, des sites Web, etc. afin qu'ils soient **esthétiquement agréables** et **faciles à comprendre**.

Les principes de la Gestalt

"Gestalt" est un terme allemand qui signifie "**ensemble cohérent**".

Les premiers tels principes ont été élaborés lors des années 1920 par les psychologues allemands Wertheimer, Koffka et Kohler.

L'**objectif** : comprendre comment l'être humain donne du sens aux stimuli chaotiques qui l'entourent ("*le tout est plus grand que la somme des parties*").

Les principes de la Gestalt traitent de la compulsion humaine à trouver de l'ordre dans le désordre. Selon ces principes, l'esprit "informe" ce que l'œil voit en percevant une série d'éléments individuels **comme un tout**.

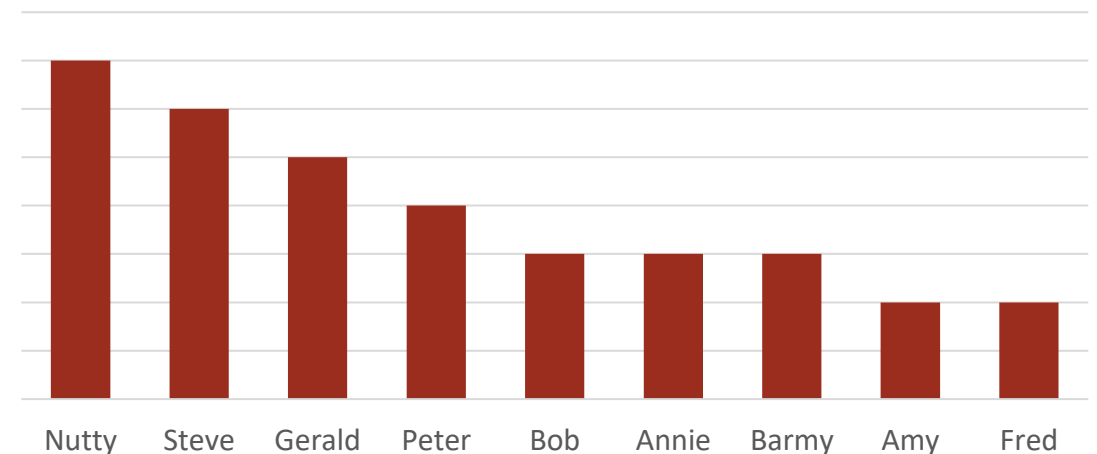
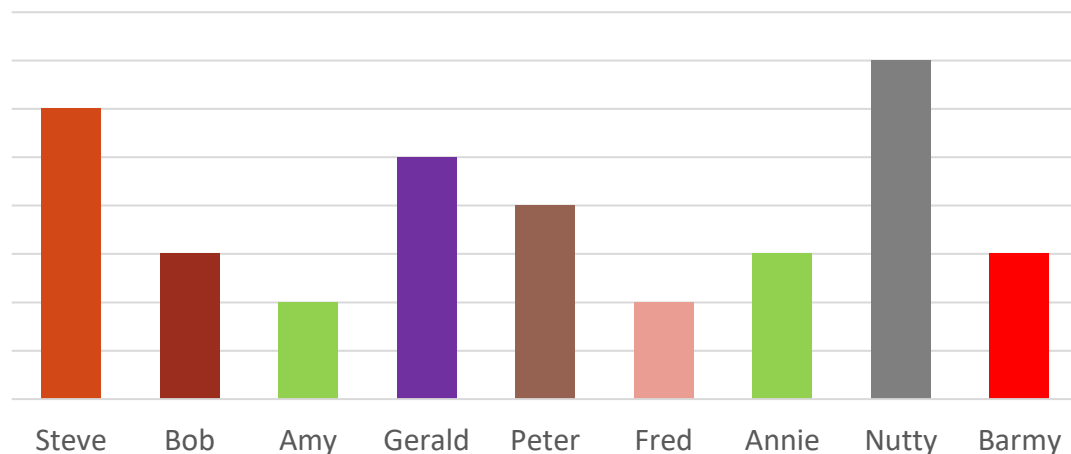
Les principes de la Gestalt

- simplicité
- continuation
- proximité
- similitude (invariance)
- point focal
- correspondance isomorphe
- dualité figure-fond
- sort commun*
- clôture*
- connexion uniforme*

Le principe de la simplicité

Le cerveau humain privilégie la **simplicité** : il tend à assimiler plus rapidement les motifs simples que ceux qui sont plus complexe

Leçon : disposez les données de manière simple et logique (dans la mesure du possible).

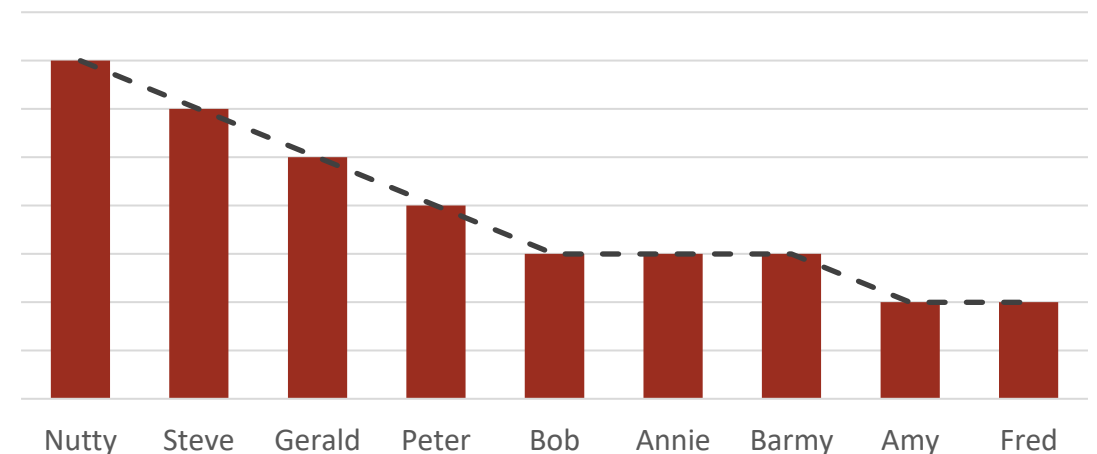
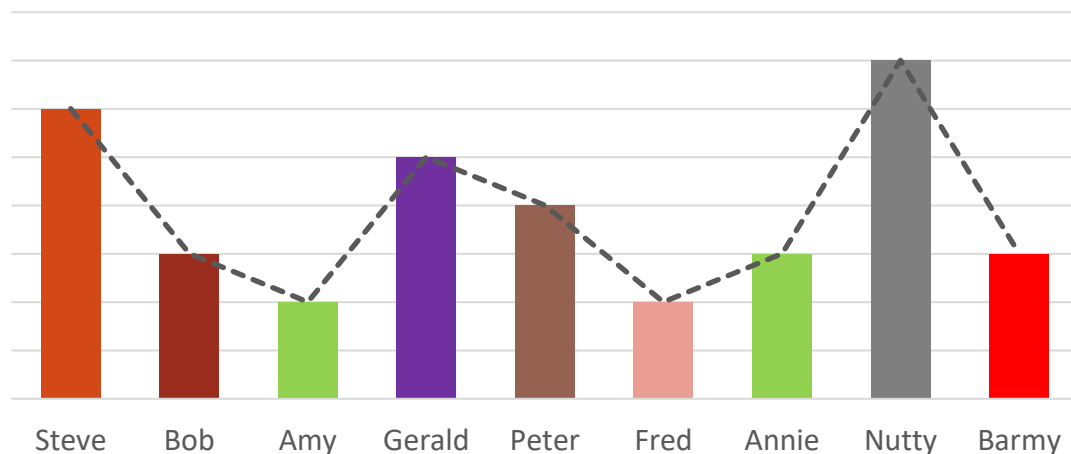


Le principe de la continuation

Nos yeux regroupent ce qui est **aligné** (par exemple, trié de haut en bas).

Dans le graphique de droite, les yeux suivent un **chemin continu** ; cela rend l'ensemble plus lisible en raison de la direction continue vers le bas.

Leçon : disposez les objets en ligne afin de faciliter le regroupement et la comparaison (lorsqu'il est logique de le faire).



Le principe de la proximité

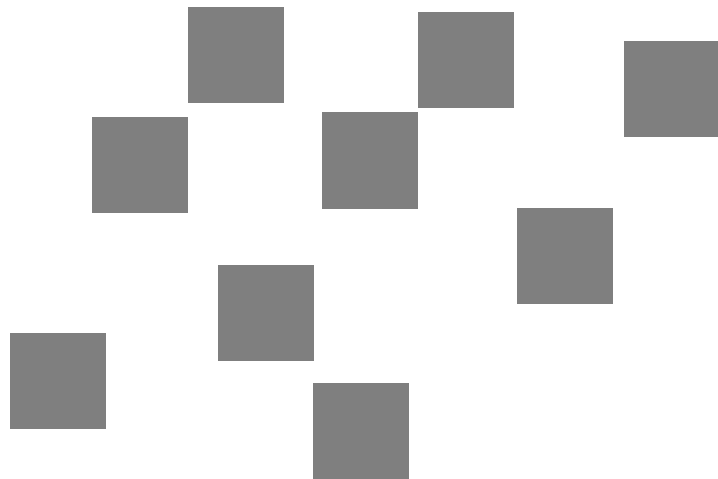
Les objets/formes qui sont à **proximité** semblent former des **groupes**.

L'effet généré par le groupe est plus "puissant" que celui généré par ces éléments constituants séparés.

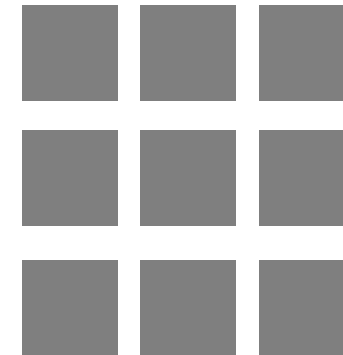
Les éléments regroupés créent l'**illusion** de formes/plans dans l'espace, même si les éléments ne se touchent pas.

Leçon : créez des regroupements par proximité qui soutiennent les priorités du graphique.

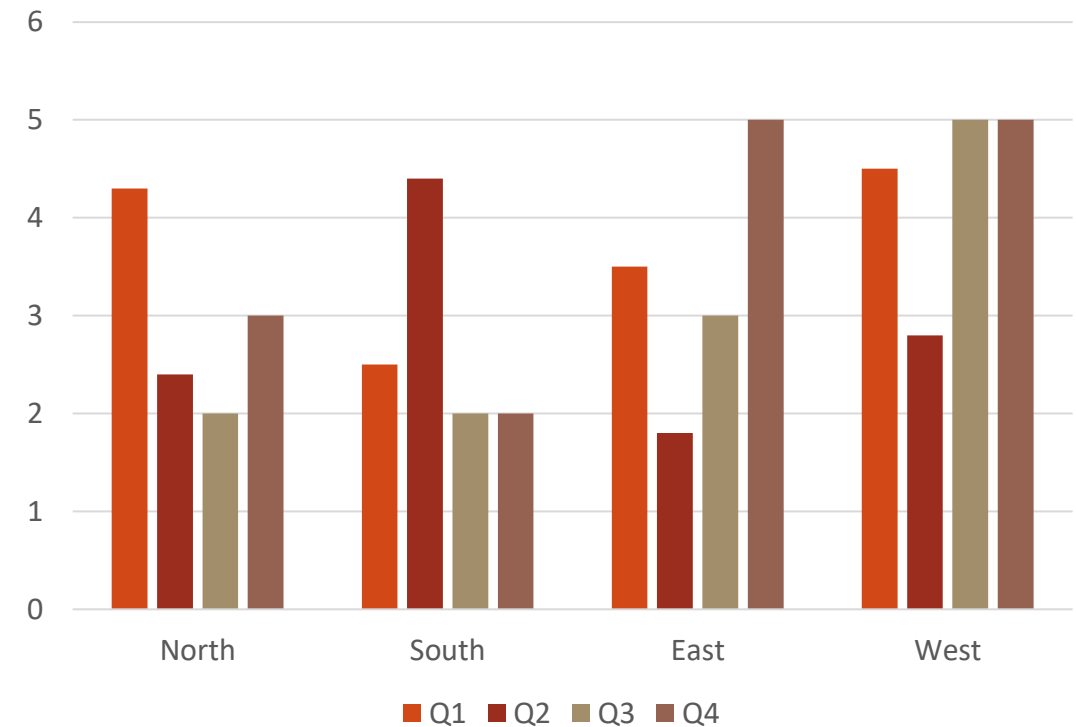
Le principe de la proximité



VS



Le principe de la proximité



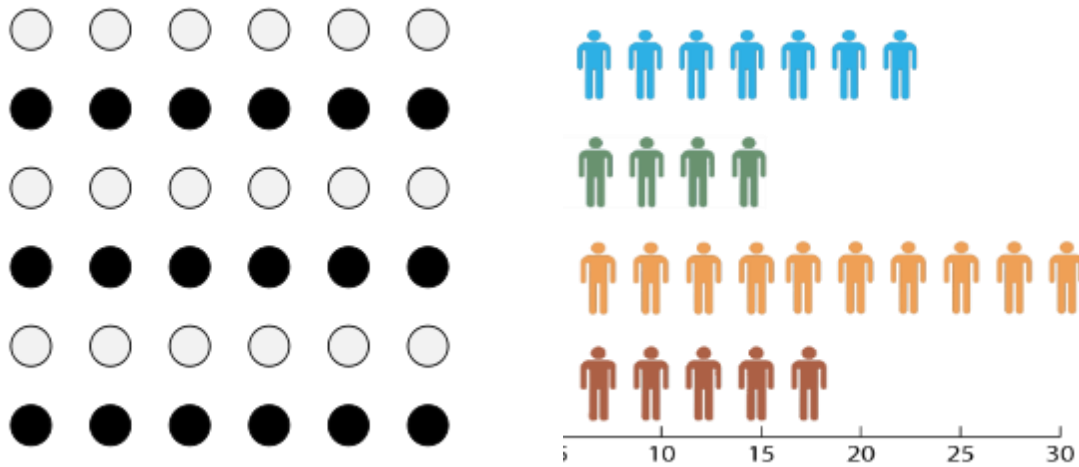
Le principe de la similitude

Les stimuli qui se ressemblent physiquement sont considérés comme **faisant partie du même objet** ; les stimuli qui ne se ressemblent pas sont considérés comme faisant partie d'un **objet différent**.

La similitude et la proximité s'associent souvent pour former une **hiérarchie visuelle**. L'un ou l'autre principe peut dominer, selon les applications.

Leçon : utilisez des caractéristiques similaires afin d'établir des relations et favoriser les regroupements d'objets.

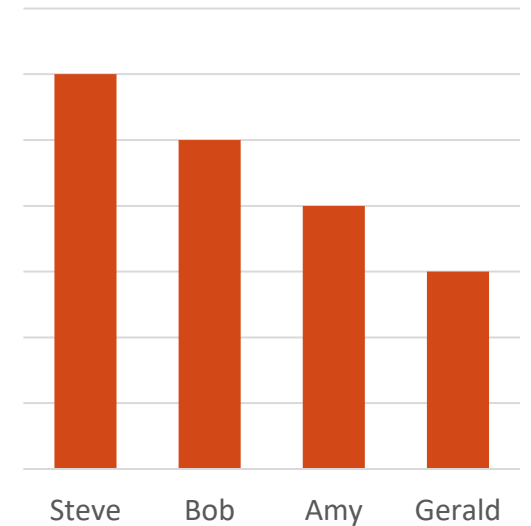
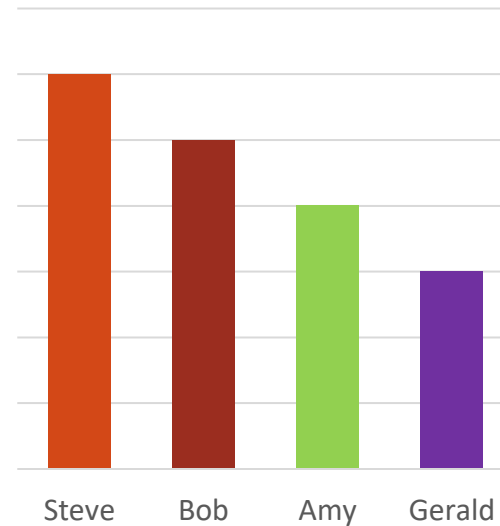
Le principe de la similitude



Dans ces exemples, la similitude domine sur la proximité : nous lisons les lignes avant de voir les colonnes.

Le principe de la similitude

On réduit la charge cognitive en rendant les choses similaires (couleur).

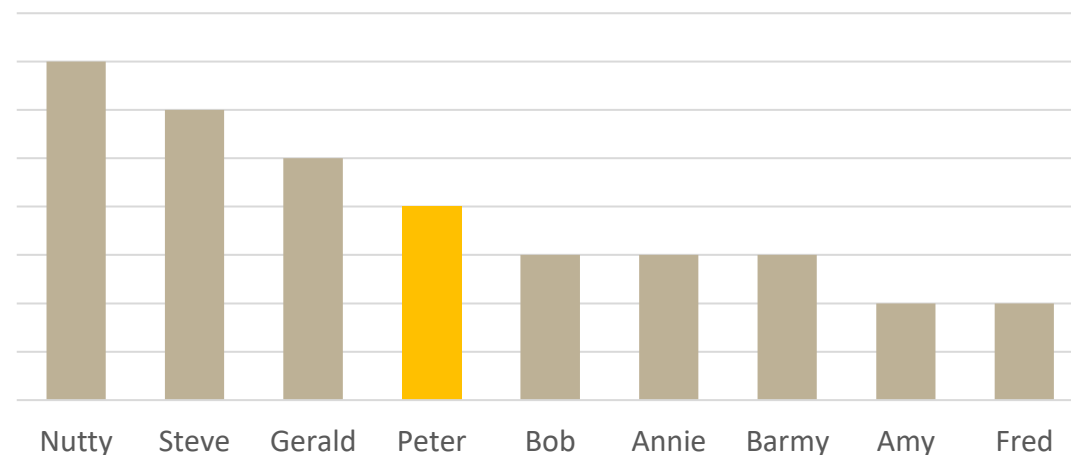
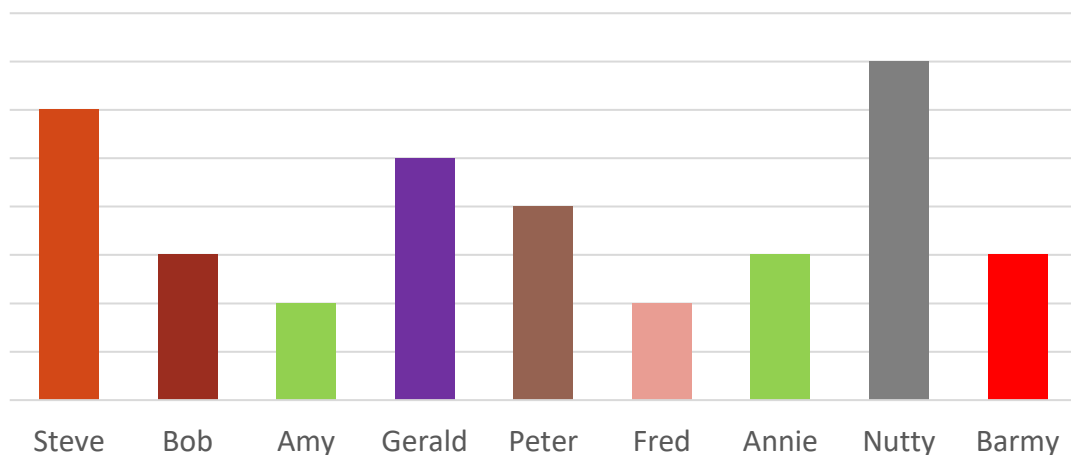


Le principe du point focal

En opposition à la similarité, le principe du **point focal** stipule que les objets d'apparence distinctive peuvent créer un point focal.

Pour mettre en évidence les performances d'un vendeur, faites en sorte que la couleur de son graphique à barres soit différente.

Leçon : utilisez les caractéristiques pour mettre en valeur des points focaux.

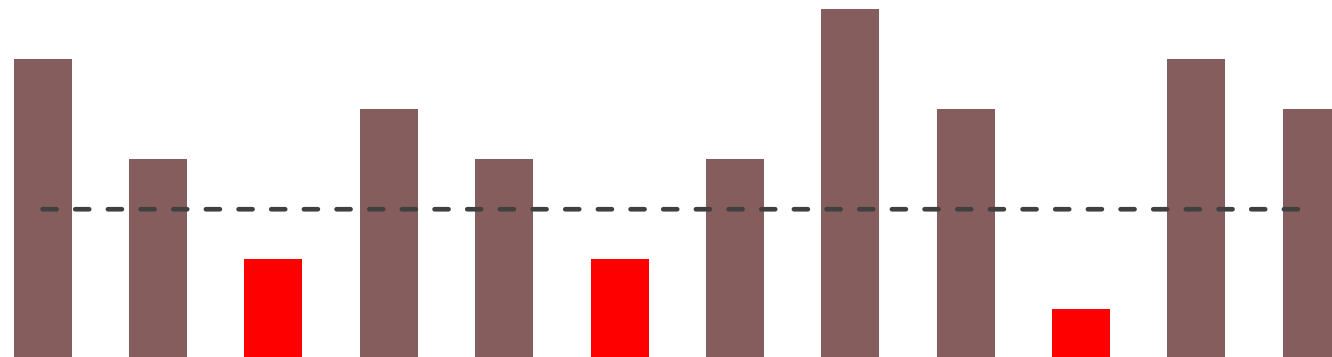


Le principe de la correspondance

Les gens interprètent et réagissent aux images en se basant sur des expériences passées/partagées (notamment pour le choix des couleurs).

Le **rouge** est souvent associé au "**mauvais**" et le **vert** au "**bon**" (daltonisme ?). Nous pouvons coder les couleurs des graphiques en conséquence.

Leçon : tenez-vous en aux conventions/pratiques bien établies (en partie ?).



Le principe de la dualité

Les éléments du graphique sont perçus soit comme des **figures** (des objets de premier plan), soit comme faisant partie de l'**arrière-plan** (“background”).

Les objets de premier plan sont **promus** par le cerveau, les objets d'arrière-plan sont **rétrogradés**.

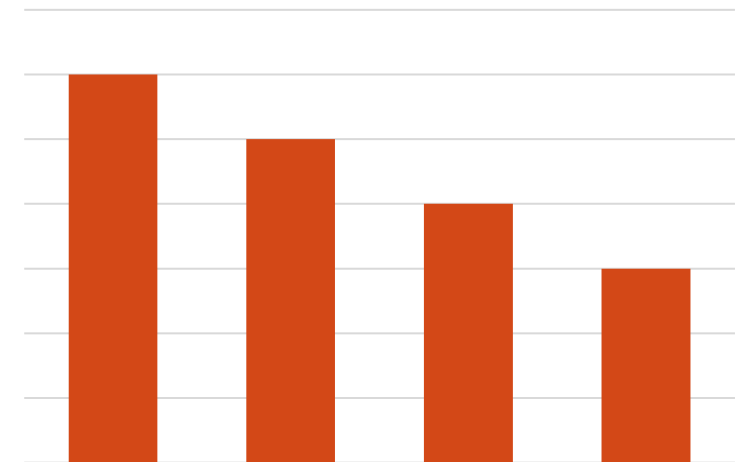
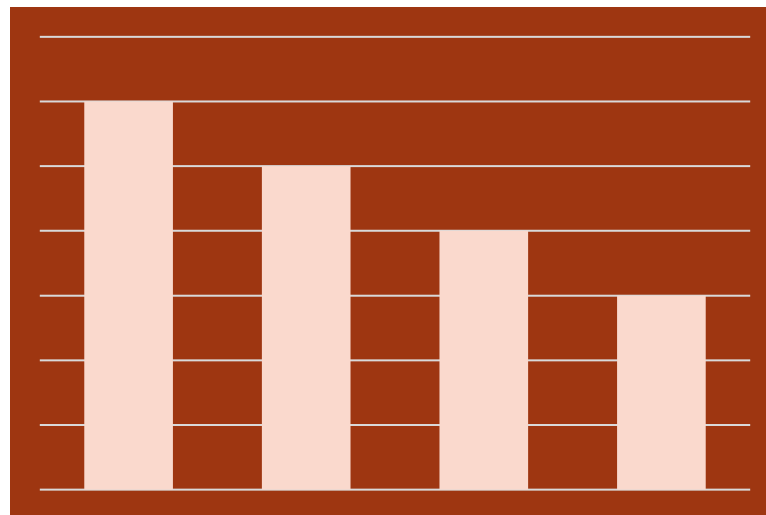
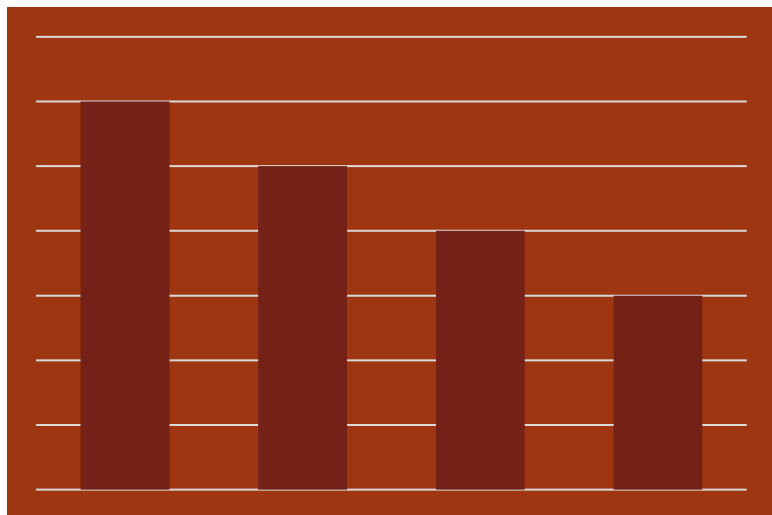
Un **contraste élevé** permet de distinguer facilement les deux types d'objets.

Leçon : veillez à ce qu'il y ait suffisamment de contraste entre le premier plan du graphique et son arrière-plan.

Le principe de la dualité

En raison du **faible contraste** entre la figure et le fond dans le graphique de gauche, il y a une **charge cognitive supplémentaire**.

L'augmentation du contraste à droite améliore la lisibilité.



Le désencombrement

Le désordre, c'est l'ennemi !

Chaque élément d'une page ajoute une charge cognitive

- identifiez et supprimez tout ce qui n'apporte pas de valeur ajoutée
- la charge cognitive est l'effort mental requis pour traiter l'information (faible = préférable)

Tufte fait référence au **rapport données/encre** – “plus la part d'encre d'un graphique consacrée aux données est importante, mieux c'est”.

Dans *Resonate*, Duarte suggère de "**maximiser le rapport signal/bruit**", le signal étant l'information ou l'histoire que nous cherchons à communiquer.

Le désencombrement

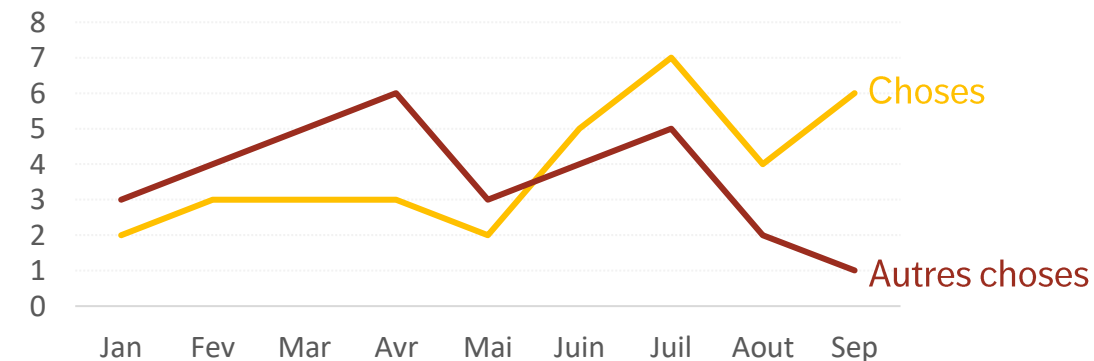
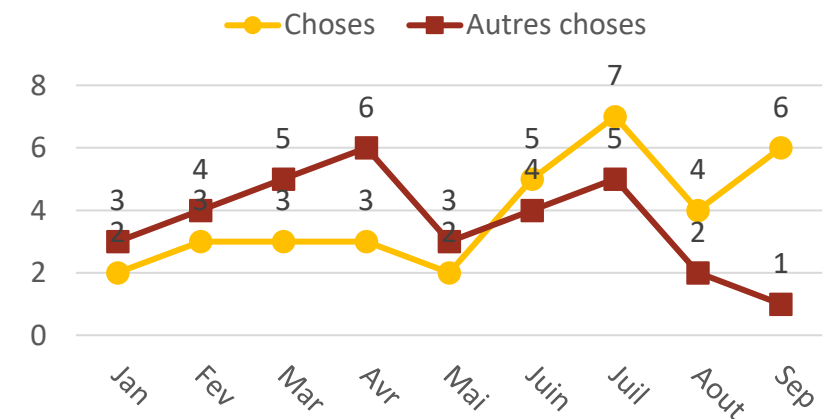
Utilisez les principes de la Gestalt pour **organiser et mettre en évidence** les données dans le graphique.

Alignez les éléments (graphiques, texte, lignes, etc.) :

- n'y allez pas au pif !!

Graphiques :

- supprimez la bordure, les grilles, les marqueurs
- nettoyez les étiquettes d'axe
- étiquetez les données directement



Le désencombrement

Utilisez des polices, une taille de police, des couleurs, et un alignement **cohérents/consistents**.

Ne faites pas pivoter le texte à un angle autre que 0 ou 90 degrés (notez l'incompatibilité entre l'anglais et le français pour la lecture du texte vertical).

Utilisez les **espaces blancs** :

- les marges doivent rester exemptes de texte et de visuels
- n'étirez pas les visuels jusqu'au bord de la page ou trop près des autres graphiques
- utilisez les espaces blancs comme frontière

Un peu de théorie chromatique

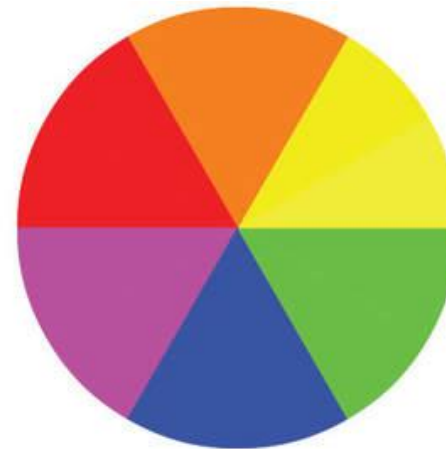
C'est un sujet compliqué – vous pouvez en lire plus sur le sujet :

- <http://www.deanenettles.com/webexamples/colorexamples/>
- <https://www.sessions.edu/color-calculator/>

Roues chromatiques :



couleurs primaires



couleurs secondaires



couleurs tertiaires

Les palettes chromatiques

Achromatique (incolore, utilisant uniquement les noirs, les blancs et les gris)



Monochromatique (schémas à une couleur)



Complémentaire (couleurs directement en face l'une de l'autre sur la roue chromatique)



Split complémentaire (2 des 3 couleurs sont adjacentes ; 1 des couleurs est opposée)



Quelques conseils chromatiques

En ce qui concerne la couleur, utilisez-la avec **parcimonie** (on apprend aux graphistes à "bien faire les choses, en noir et blanc").

Sur la base de la Gestalt, les schémas **monochromes** peuvent être efficaces.

Au besoin, primez le schéma d'identité de l'entreprise.

Créez un **gabarit** (et utilisez-le !).

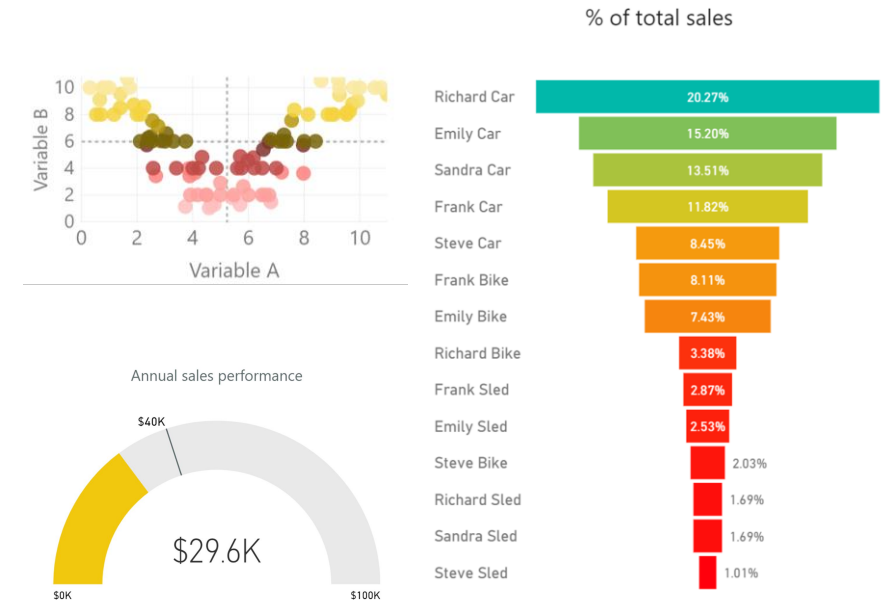
Téléchargez les images afin de voir à quoi elles ressemblent pour les daltoniens :

- <https://www.color-blindness.com/coblis-color-blindness-simulator> (ce n'est pas le seul outil)

Quelques considérations sur la taille

En supposant que les tableaux aient été **désencombrés** :

- les choses de même importance devraient avoir la même taille
- les autres choses s'échelonnent en fonction de leur importance



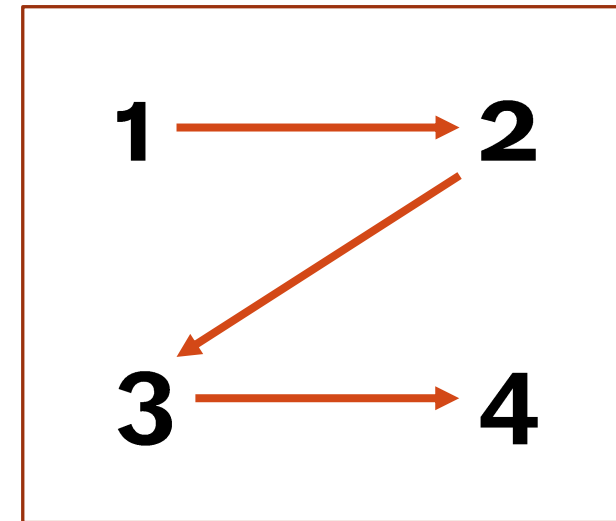
Quelques considérations sur la position

Comment doit-on placer les éléments sur une page ?

En Occident, la plupart des gens commencent en **haut à gauche** et font des zigzags jusqu'en **bas à droite**.

Une **règle simple** : ne faites pas travailler les gens trop dur

- message principal : en haut à gauche/en haut à droite
- info par ordre de préférence
- les gens se concentrent moins au fur et à mesure qu'ils scannent, donc la complexité devrait diminuer au fur et à mesure que l'on se déplace vers le coin inférieur droit



Session 2

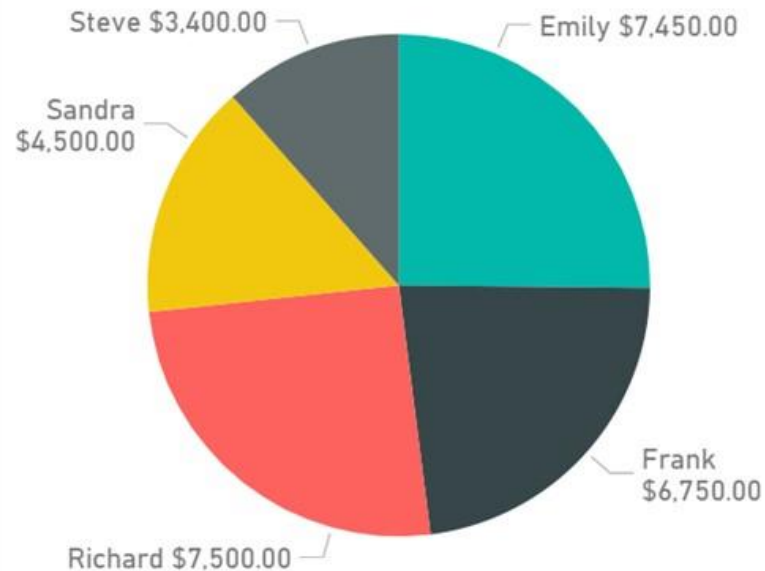
Sales Dashboard

\$ sales



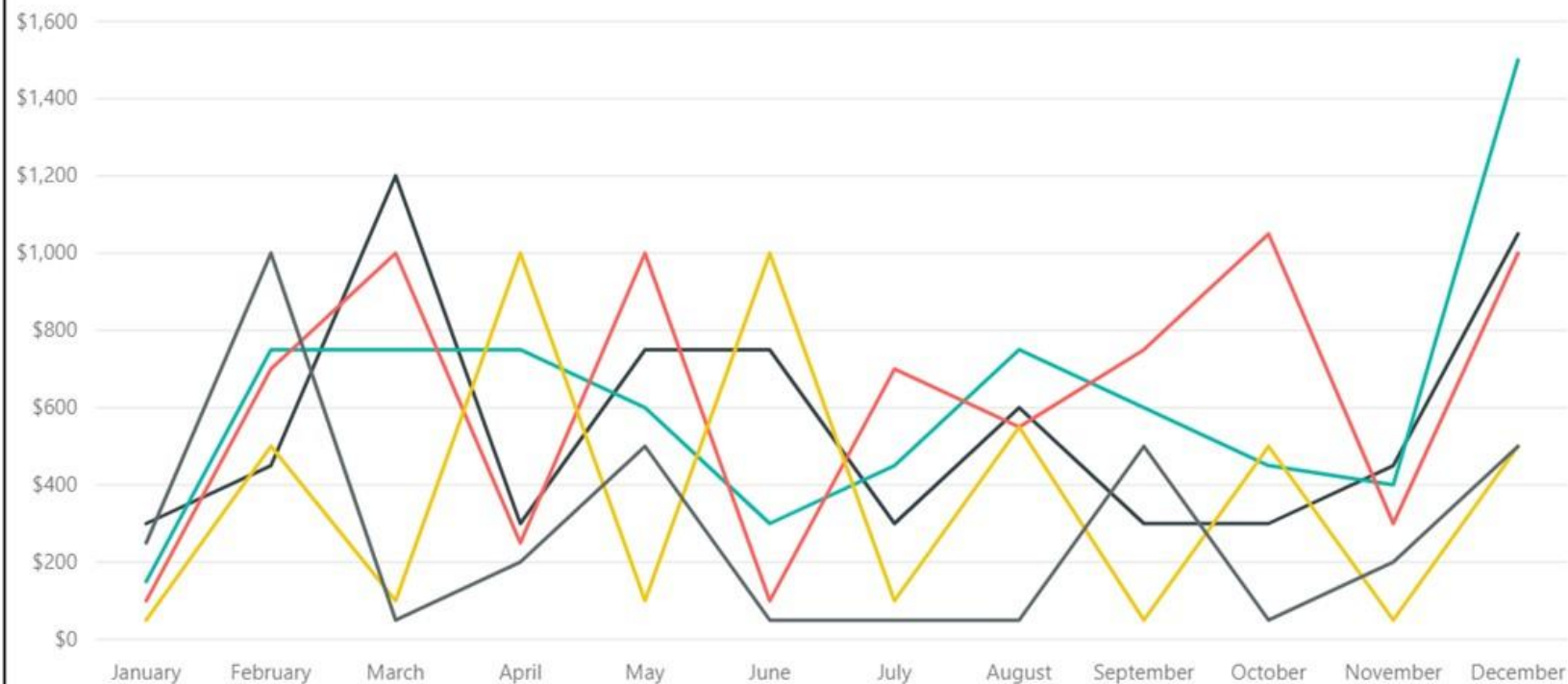
\$ sales by Salesperson

Salesperson Emily Frank Richard Sandra Steve



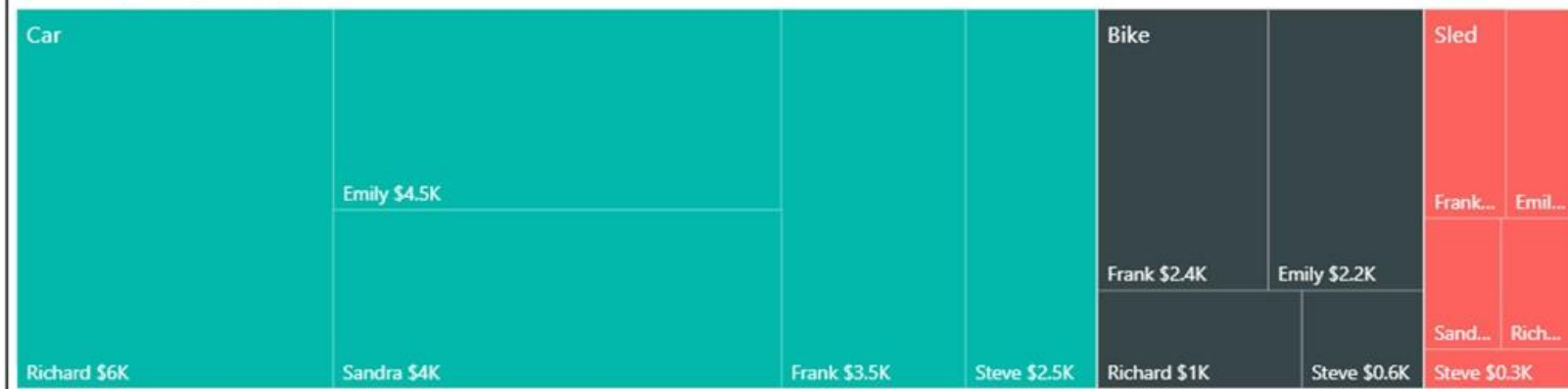
\$ sales by Month and Salesperson

Salesperson Emily Frank Richard Sandra Steve



\$ sales by Product and Salesperson

Product Car Bike Sled



Sales Dashboard

Annual Sales for 2017

Total Sales

\$29.6K



Lectures conseillées

L'esthétique des graphiques

The Practice of Data Visualization **Part II: The Foundations of Visual Design**

4. The Mechanics of Visual Perception

5. Visual Design and Data Charts

- 5.1 Visualization and Memory

- 5.2 Aesthetic Considerations

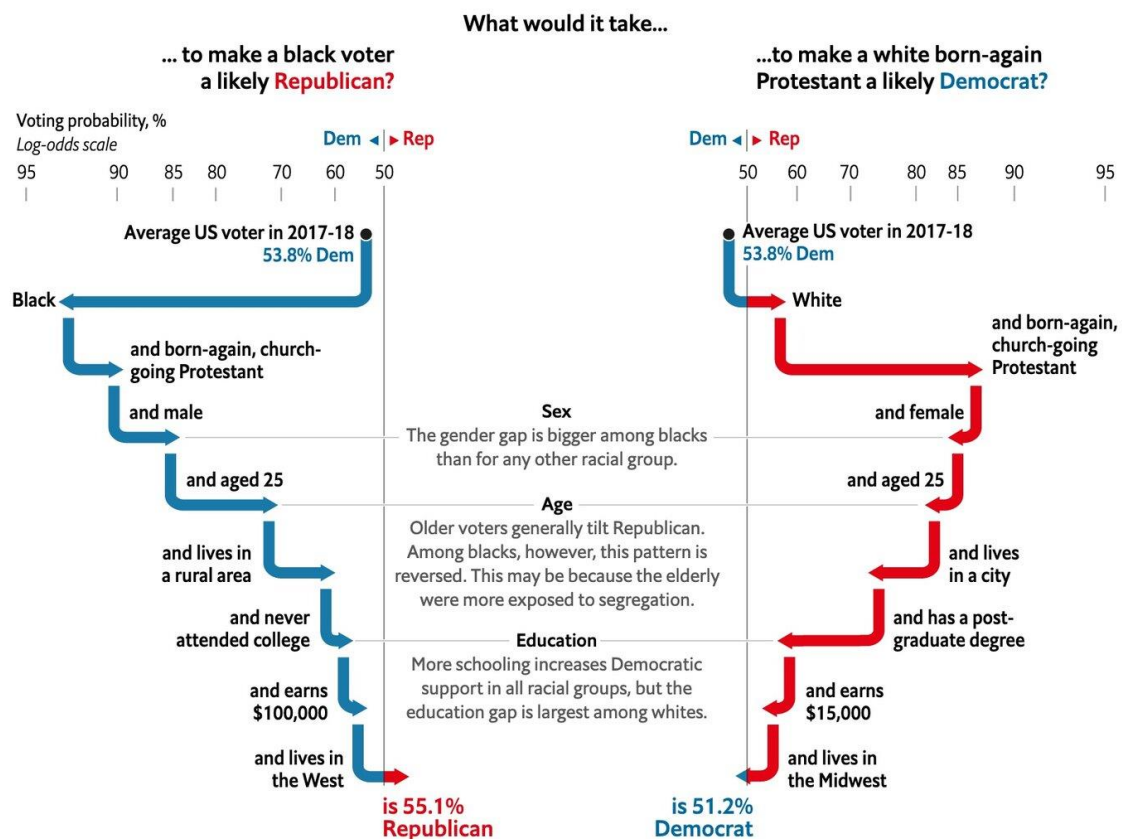
- 5.3 Decluttering

Exercices

L'esthétique des graphiques

Commentez l'esthétique des graphiques suivants, en fonction :

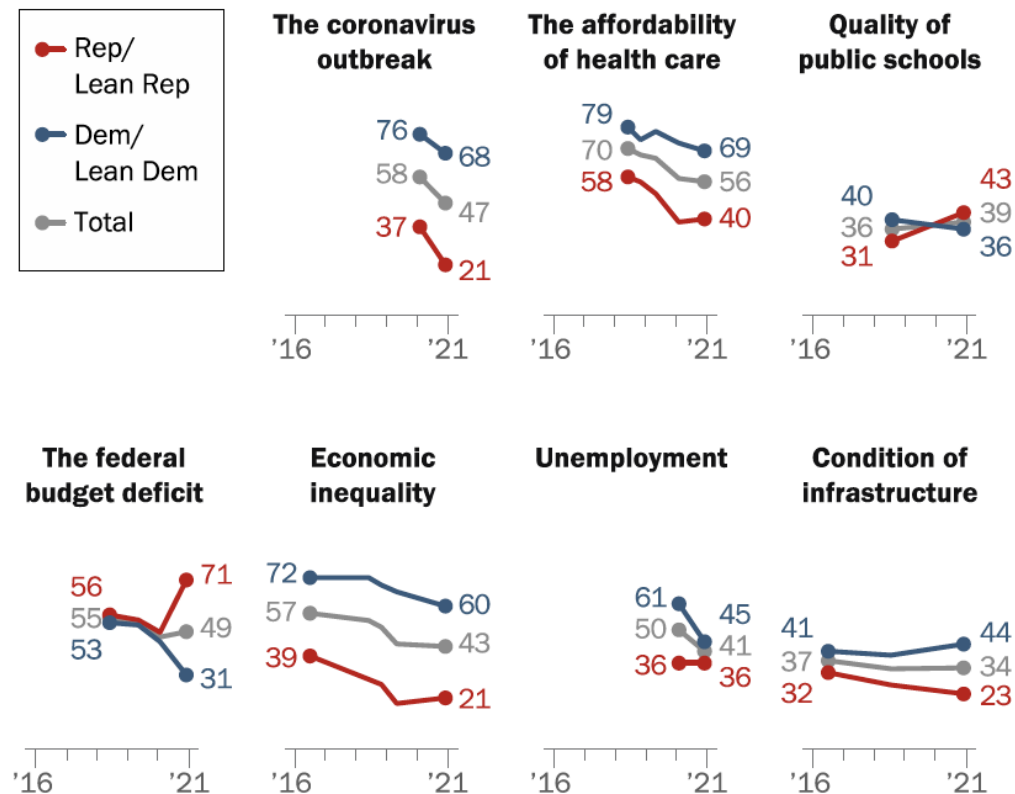
- des principes de la Gestalt
- de l'utilisation des couleurs
- de l'encombrement
- de la taille et de la position
- etc.



Sources: YouGov; *The Economist*

Republican concern about the budget deficit increases sharply; Democratic concern declines

% who say ____ is a **very big problem** in the country today

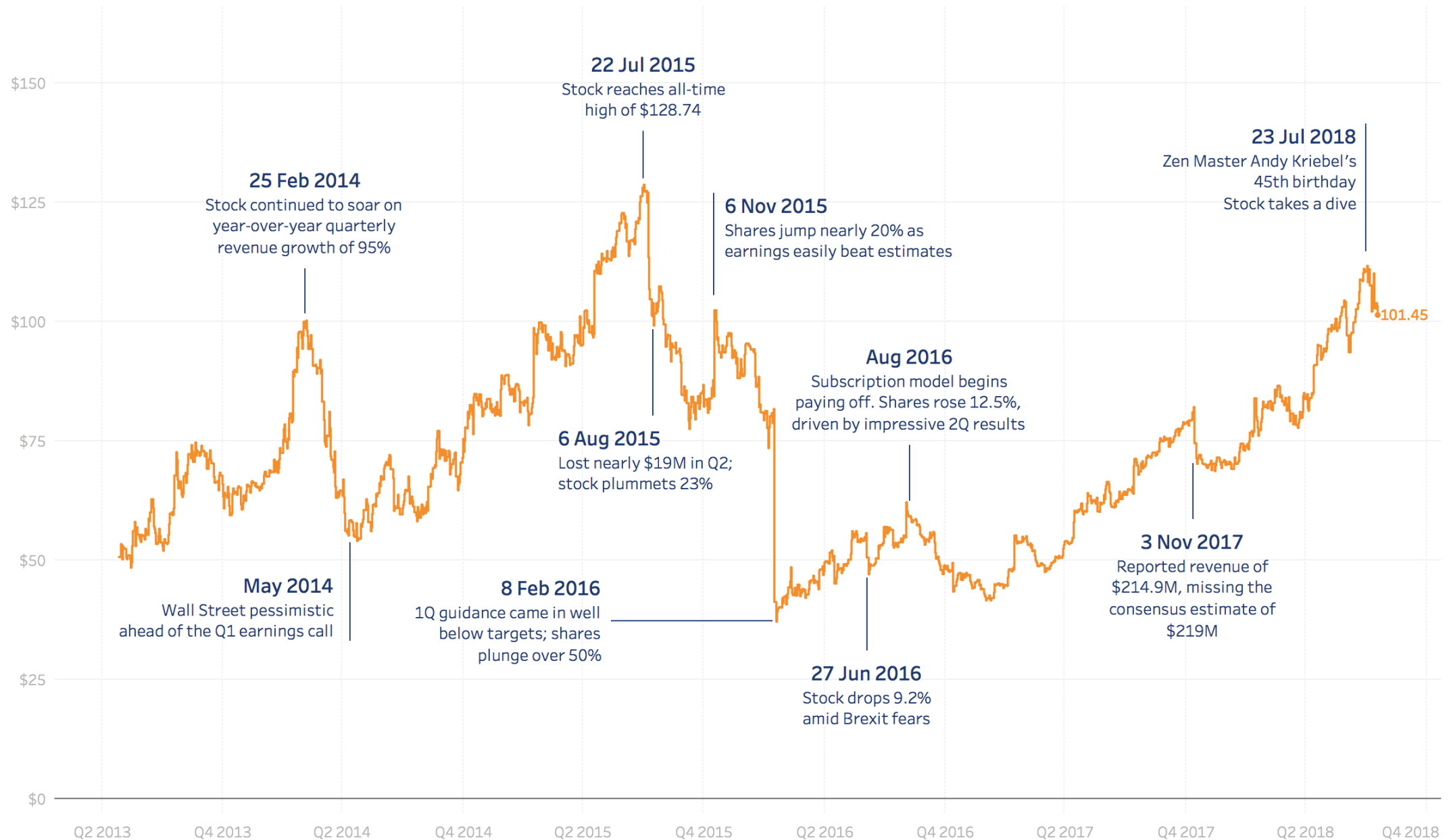


Note: March 2019 and earlier wording for economic inequality was "The gap between the rich and poor." See topline for details.

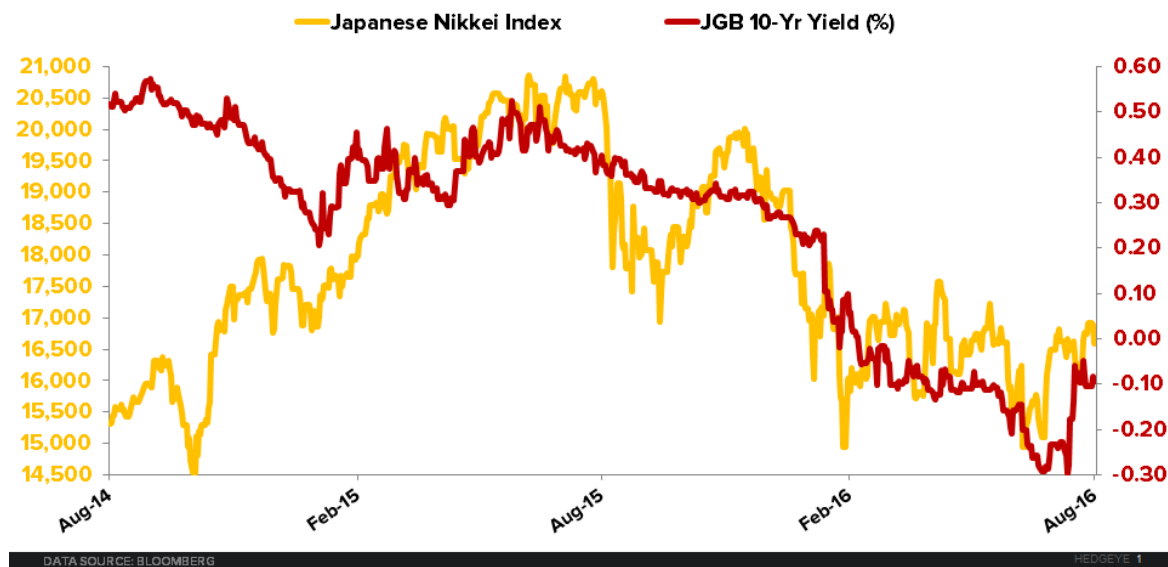
Source: Survey of U.S. adults conducted April 5-11, 2021.

PEW RESEARCH CENTER

The Roller Coaster Ride of Tableau's Stock

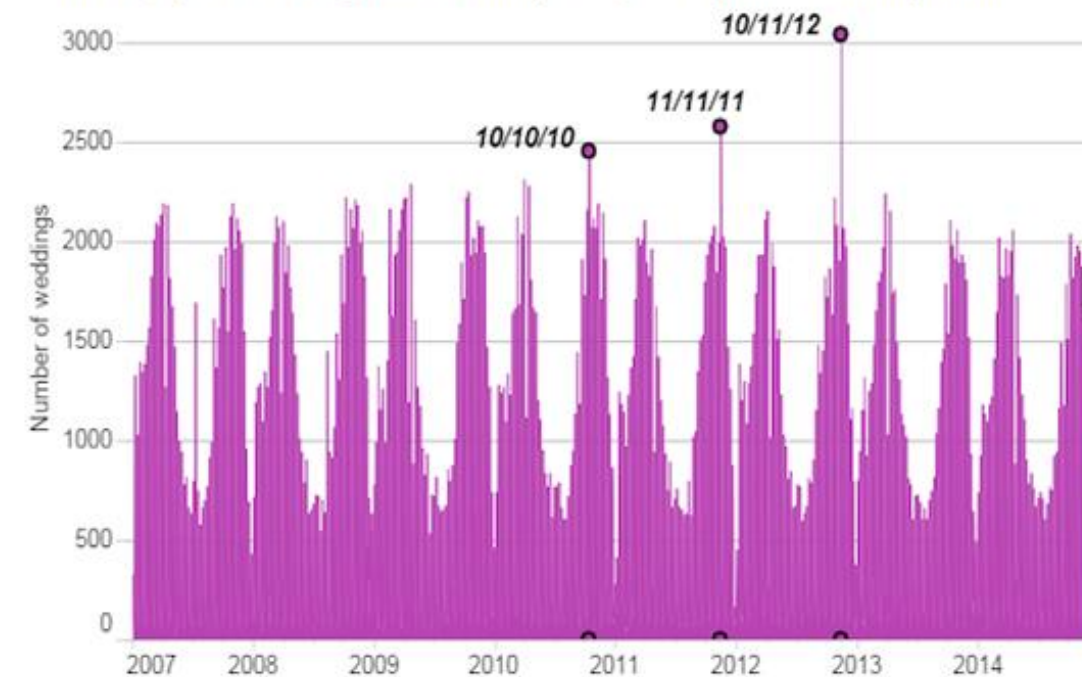


BAD ISN'T GOOD FOREVER: JAPAN

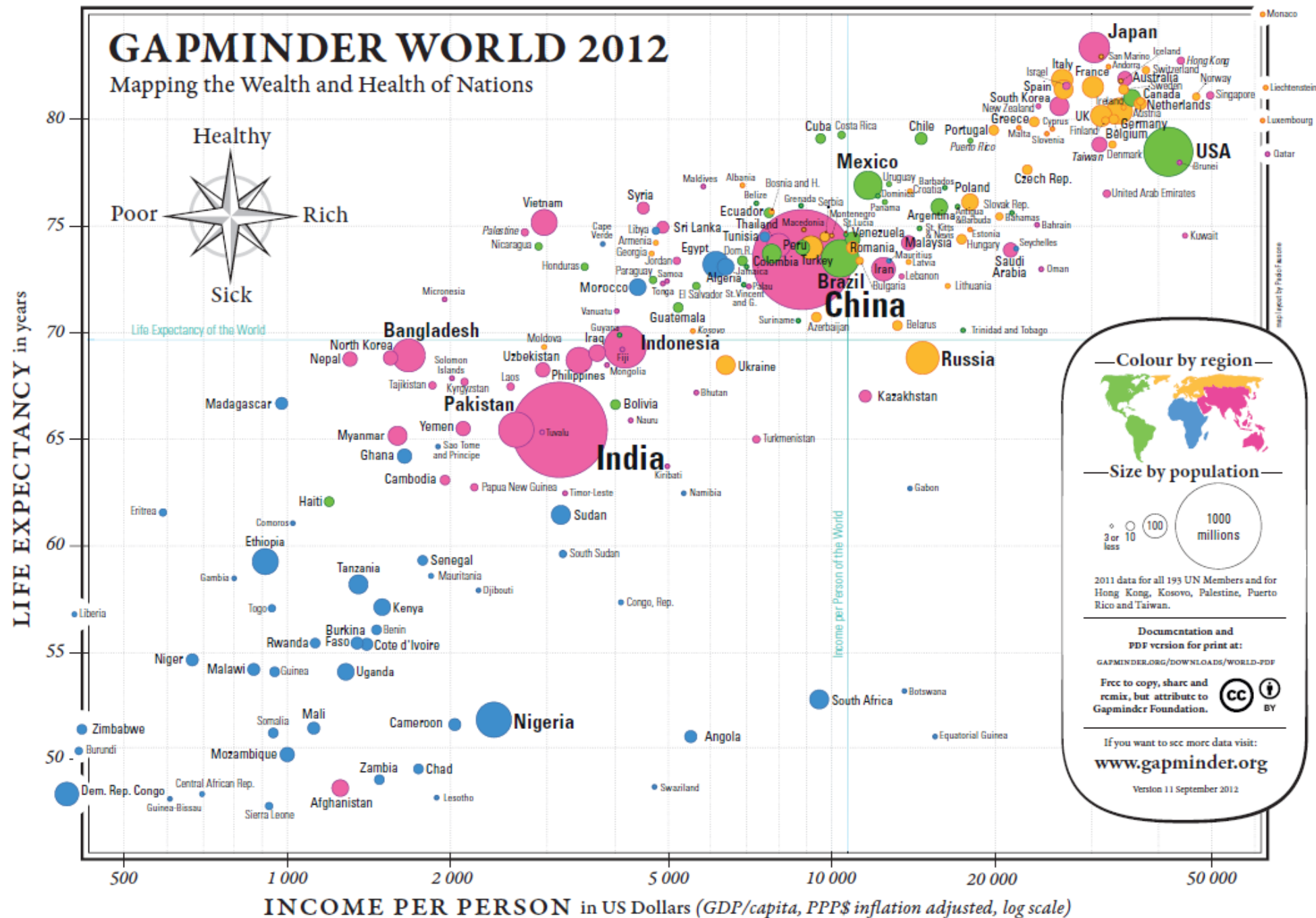


Weddings in Australia

The most popular wedding dates form repeating or sequential number patterns.



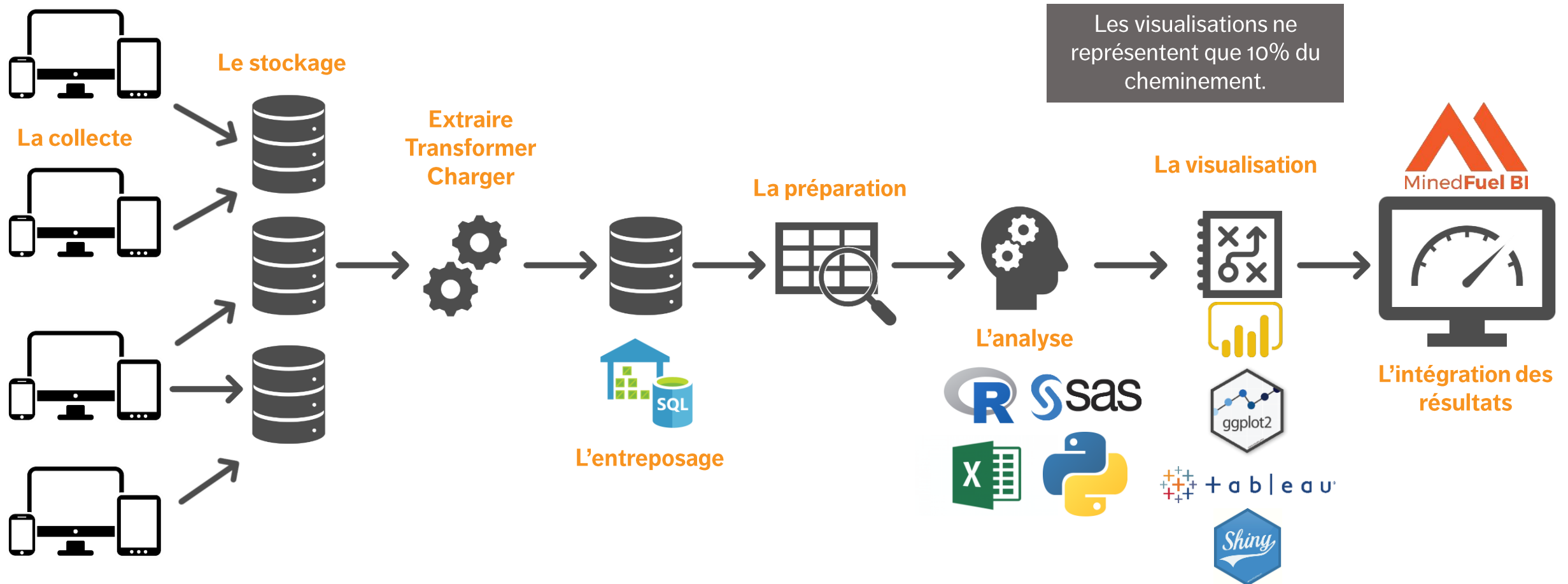
Graphic: Inga Ting | Source: ABS 2015





5. Les tableaux de bord

L'environnement des données

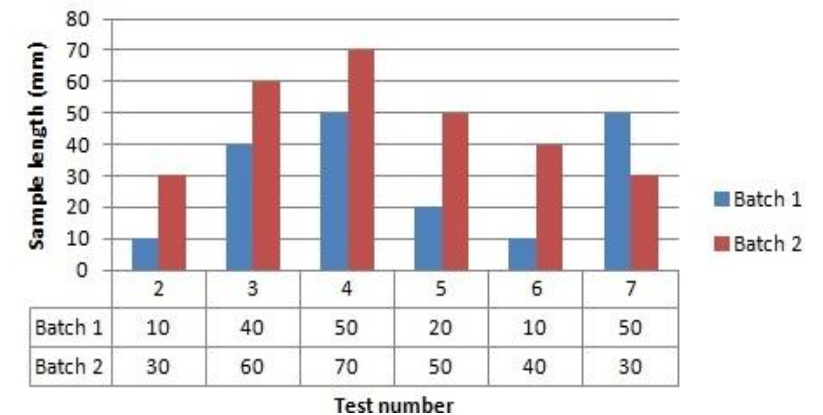


La vue d'ensemble

Le passé est **axé sur les données** :

- principalement Excel (ou des outils de “reporting” tel que Cognos)
- principalement des chiffres, des tableaux et des graphiques non interactifs
- distribué sur les ordinateurs de bureau, par e-mail, dans une présentation PowerPoint
- statiques, principalement tournés vers l'arrière (indicateurs retardés)
- Les KPI et les tableaux de bord sont quelque peu artificiels

Region	Jan-06	Feb-06	Mar-06	Apr-06	May-06	Jun-06	Jul-06
Actuals							
Seattle	111	653	1,598	3,411	3,972	5,092	5,298
Boise	26,779	27,867	29,153	30,557	33,402	35,400	35,450
Portland	33,078	34,401	37,535	39,916	41,357	45,306	46,670
Spokane	25,417	26,669	28,092	29,020	29,674	30,501	30,830
North Region	199,841	211,053	226,789	242,957	256,605	273,640	277,770
Plan							
Seattle	693	468	790	1,383	2,205	3,180	4,210
Boise	29,525	26,062	27,088	28,269	29,536	30,821	32,160
Portland	32,276	34,708	36,737	38,857	41,066	43,364	45,750
Spokane	30,500	26,644	27,987	29,430	30,994	32,594	34,230
North Region	191,783	203,916	216,524	230,474	246,390	263,378	281,220
Variance							
Seattle	-582	185	808	2,029	1,767	1,912	1,078
Boise	-2,746	1,805	2,064	2,288	3,866	4,578	3,288
Portland	802	-307	798	1,059	291	1,942	920
Spokane	-5,082	25	105	-410	-1,320	-2,093	-3,398
North Region	-8,057	7,437	10,265	12,483	10,245	10,264	-3,410



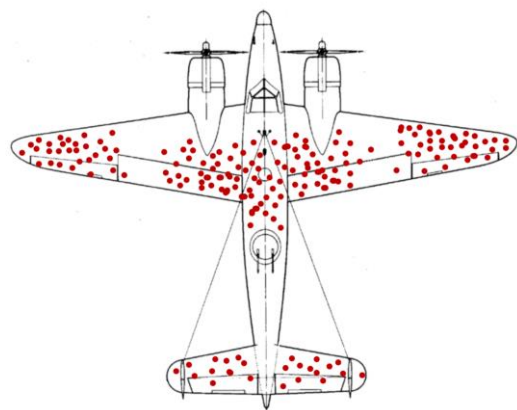
La vue d'ensemble

L'avenir est **axé sur les récits/histoires** :

- de nouveaux outils : Power BI, Tableau, Qlickview, Shiny, etc.
- principalement des visualisations, occasionnellement des chiffres et des tableaux
- distribués sur le web (à l'interne et à l'externe)
- dynamique et tournés à la fois vers l'arrière et vers le futur (indicateurs avancés et retardés)
- les données pour tous

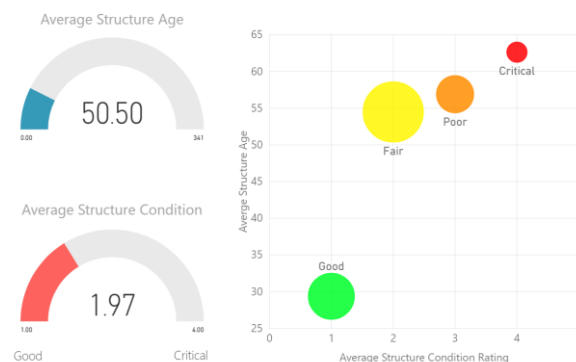


Définir le contexte



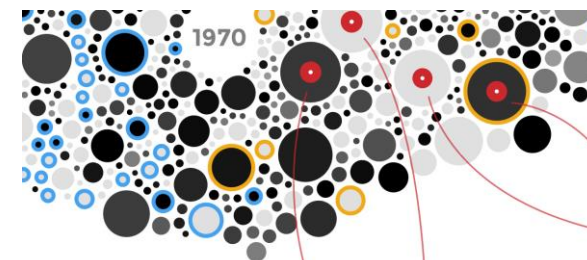
Directory of Federal Real Property (DFRP) Dashboard

You have selected 20,186 properties that contain 35,148 structures



Access to Information and Privacy (ATIP) search

You have currently selected 28,711 requests totaling 6,597,612 pages of information



The Beatles

No other artist or band has more songs in the Top 2000 as the Beatles. With 38 songs they are responsible for 14% of all titles before 1970. Nonetheless, only 5 years ago they still had 50 songs in the list.

- 4 Piano Man
Billy Joel: 1974
- 5 Child in Time
Deep Purple: 1972

Secondes

Minutes

Fractions d'une heure

Heures

← Infographies/Data Viz →

← Tableaux de bord →

← Rapports et exploration →

← Art de données →

Les tableaux de bord

Un **tableau de bord** est un affichage visuel de données utilisé afin surveiller des conditions et/ou faciliter la compréhension d'une situation.

Exemples :

- un écran interactif qui permet aux gens d'explorer les demandes d'assurance automobile par ville, province, âge du conducteur, etc.
- un PDF montrant les principales mesures d'audit qui est envoyé par e-mail à la DG d'un département sur une base hebdomadaire
- un écran mural qui affiche les statistiques du centre d'appels en temps réel
- un app mobile qui permet aux administrateurs d'hôpitaux d'examiner les temps d'attente sur une base horaire et quotidienne pour l'année en cours et l'année précédente

Quelques questions à considérer

Dans le tableau de bord d'une voiture, un petit nombre d'**indicateurs clés** (vitesse, niveau d'essence, etc.) doivent être compris en un **coup d'œil**. Les tableaux de bord n'en tenant pas compte ont des conséquences catastrophiques.

Il est nécessaire de répondre aux questions suivantes avant de concevoir le tableau de bord :

- qui est le consommateur du tableau de bord ?
- quelle histoire le tableau de bord raconte-t-il ?
- quelles données (catégories) seront utilisées ?
- qu'est-ce qui apparaîtra sur le tableau de bord ?
- comment le tableau de bord sera-t-il utilisé ?



Quelques directives de conception

Nick Smith suggère les 6 règles d'or suivantes :

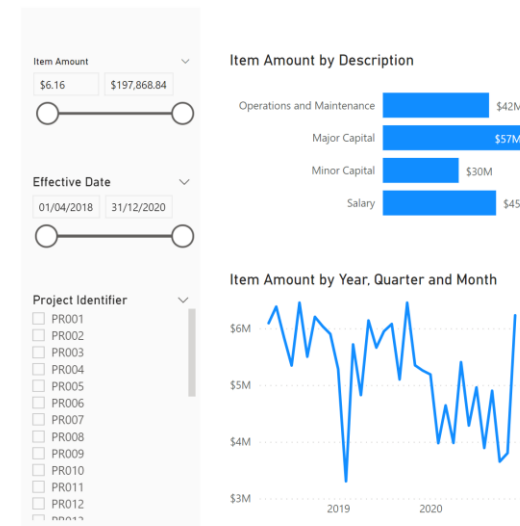
- **considérez le public** (qui essayez-vous d'informer ? la DG a-t-elle vraiment besoin de savoir que les serveurs fonctionnent à 88% de leur capacité ?)
- **choisissez le bon type de tableau de bord** (opérationnel, stratégique, analytique)
- **regroupez les données de manière logique, utilisez l'espace de manière judicieuse** (divisez les domaines fonctionnels : produit, ventes/marketing, finances, personnel, etc.)
- **faîtes en sorte que les données soient pertinentes pour le public** (étendue et portée des données, différents tableaux de bord pour différents départements, etc.)
- **évitez d'encombrer le tableau de bord** (présentez uniquement les métriques les plus importantes)
- **rafraîchissez vos données à la bonne fréquence** (en temps réel, quotidiennement, hebdomadairement, mensuellement, etc.)

Les types de tableaux de bord

Exploration : utilisation de visualisations comme outils pour explorer/comprendre les données

- haut niveau d'interactivité
- haut niveau de détail
- **tous les** aspects des données doivent être représentés (tableaux, colonnes, calculs, etc.)
- pas d'annotations/d'explications nécessaires

Financial Data Exploration



\$59.78K
Average of Item Amount

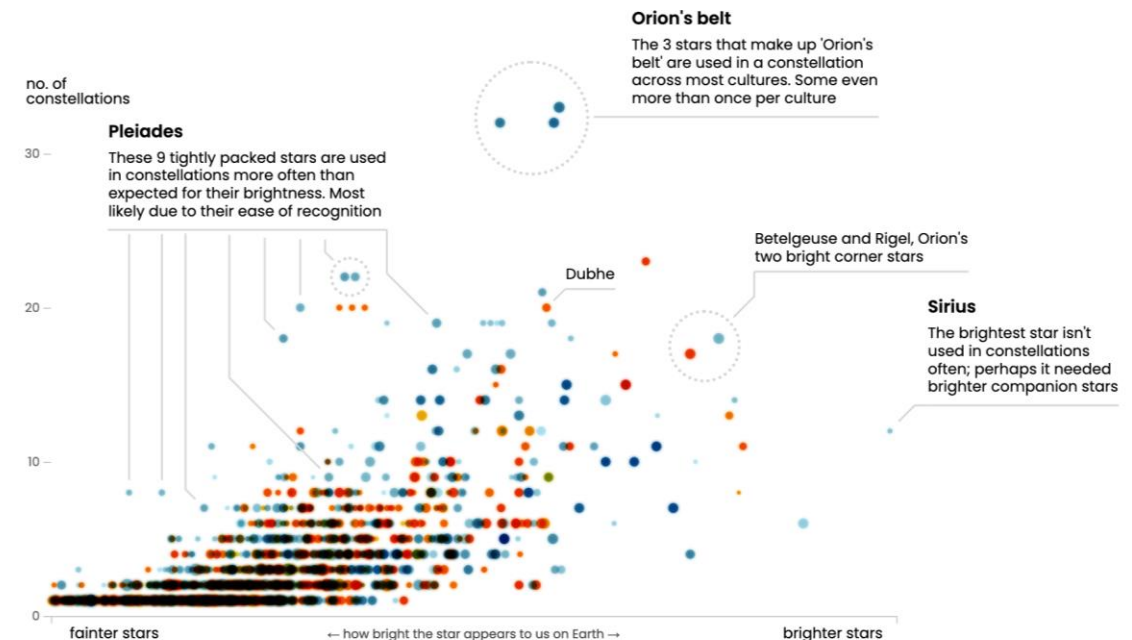
\$173.78M
Sum Item Amount

Journal Voucher Type Code	2018	2019	2020	Total
MC	\$18,110,444.61	\$21,810,187.90	\$17,448,129.75	\$57,368,762.27
PR001	\$687,677.21	\$1,792,024.46	\$907,127.55	\$3,386,829.23
PR002	\$788,825.39	\$565,031.07	\$813,175.69	\$2,167,032.15
PR003	\$1,517,664.95	\$612,091.00	\$1,093,131.35	\$3,222,887.30
PR004	\$800,174.27	\$719,551.46	\$1,155,498.57	\$2,675,224.30
PR005	\$611,844.01	\$1,559,623.99	\$505,962.54	\$2,677,430.55
PR006	\$869,847.19	\$1,142,078.50	\$567,309.21	\$2,579,234.90
PR007	\$1,254,247.56	\$1,202,463.46	\$1,121,613.47	\$3,578,324.48
PR009	\$536,301.11	\$1,466,714.57	\$654,848.18	\$2,657,863.87
PR010	\$1,025,185.44	\$1,124,411.66	\$810,384.12	\$2,959,981.22
PR011	\$1,323,665.62	\$947,916.20	\$951,129.63	\$3,222,711.45
PR012	\$894,949.35	\$1,321,602.78	\$1,142,398.09	\$3,358,950.22
PR013	\$810,720.06	\$1,397,946.44	\$943,871.63	\$3,152,538.13
PR015	\$1,115,244.24	\$1,238,919.57	\$1,211,122.76	\$3,565,286.57
PR017	\$1,163,245.06	\$1,346,151.02	\$595,533.30	\$3,104,929.39
PR018	\$888,426.84	\$1,297,179.23	\$1,177,356.88	\$3,362,962.95
PR019	\$942,777.50	\$1,028,710.89	\$748,386.14	\$2,719,874.53
PR022	\$842,076.88	\$697,992.57	\$1,105,900.34	\$2,645,969.79
PR023	\$1,219,843.67	\$1,143,895.90	\$1,115,052.77	\$3,478,792.34
PR027	\$817,728.27	\$1,205,883.13	\$828,327.52	\$2,851,938.92
MIC	\$8,733,325.92	\$11,316,310.76	\$9,855,321.54	\$29,904,958.22
PR001	\$488,147.03	\$447,373.91	\$493,012.00	\$1,428,532.94
PR002	\$288,526.70	\$794,250.21	\$275,485.45	\$1,358,262.36
PR003	\$249,707.20	\$301,928.04	\$339,914.44	\$891,549.68
Total	\$53,750,707.93	\$65,112,680.21	\$54,913,391.39	\$173,776,979.54

Les types de tableaux de bord

Mise en récit: utilisation de visualisations comme outils qui expliquent les données et qui communique l'histoire

- faible niveau d'interactivité
- faible niveau de détail
- seulement les **principaux aspects** des données doivent être représentés
- les annotations et les explications font avancer le récit

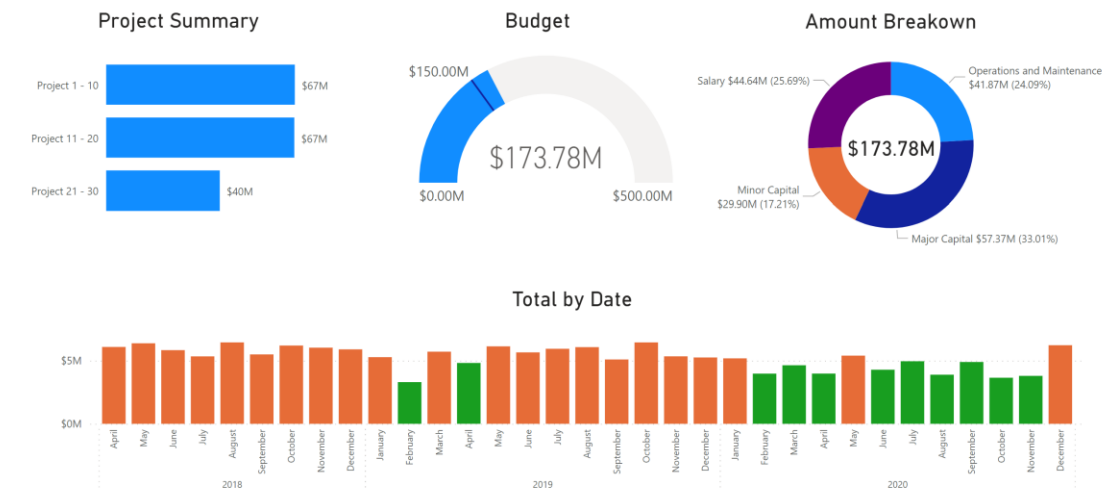


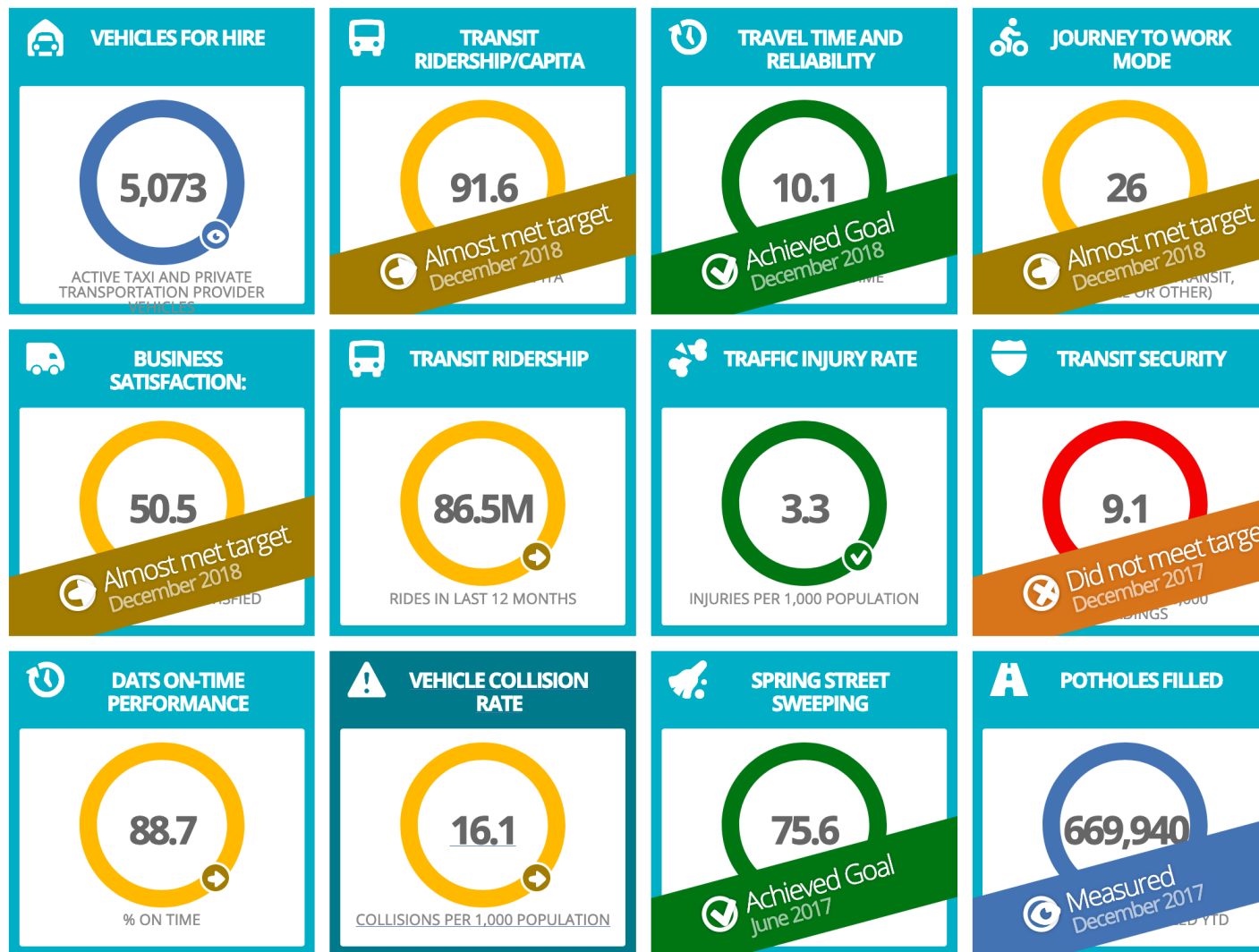
Les types de tableaux de bord

Prise de conscience de la situation :
utilisation de visualisations comme outils qui fournissent un “instantané” des données

- niveau moyen d'interactivité
- pas "scénarisé" mais bien organisé
- les données récapitulatives doivent être représentées
- les anomalies sont mises en évidence
- souvent utilisé pour les présentations internes

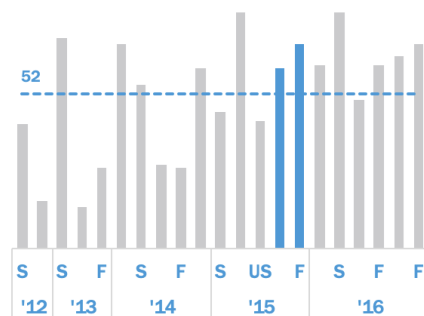
Financial Snapshot





✓ Meets or Exceeds Target
 ➕ Near Target
 ✖ Needs Improvement
 🕒 Measuring
 📊 Collecting Data

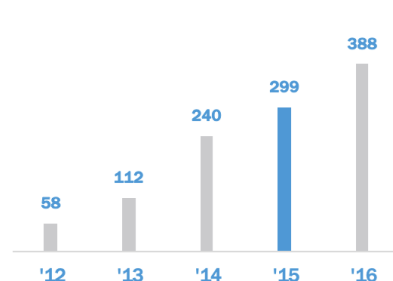
Students



1097

Total Students in five years

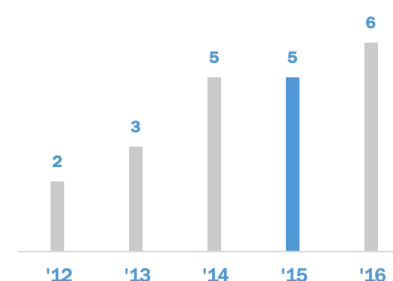
Enrollments



687

Total Students in 2015-2016

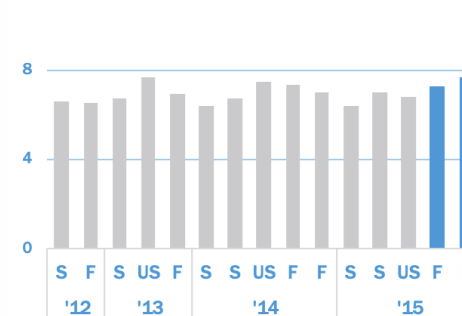
Classes



21

Total Classes in five years

Ratings



7.7 of 8

Most recent instructor rating (out of 8.0)

Semesters

2015 Fall Semester 001

Questions

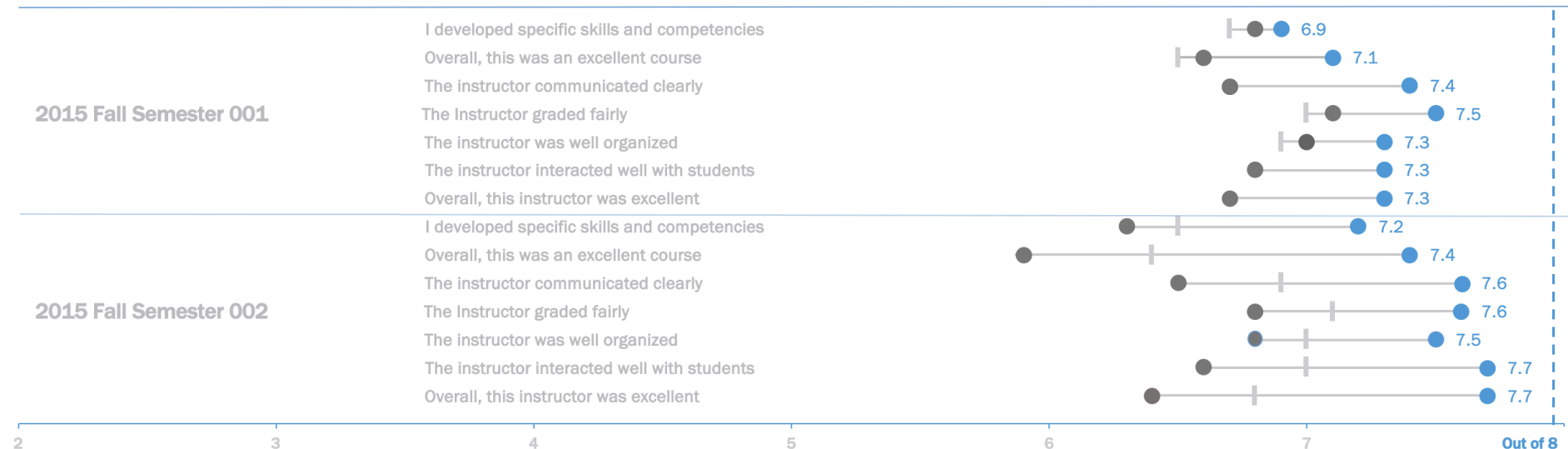
- I developed specific skills and competencies
- Overall, this was an excellent course
- The instructor communicated clearly
- The Instructor graded fairly
- The instructor was well organized
- The instructor interacted well with students
- Overall, this instructor was excellent

2015 Fall Semester 002

- I developed specific skills and competencies
- Overall, this was an excellent course
- The instructor communicated clearly
- The Instructor graded fairly
- The instructor was well organized
- The instructor interacted well with students
- Overall, this instructor was excellent

● BANA ■ College ● Shaffer

Ratings



Lectures conseillées

Les tableaux de bord

Data Understanding, Data Analysis, Data Science
Volume 2: Fundamentals of Data Insight

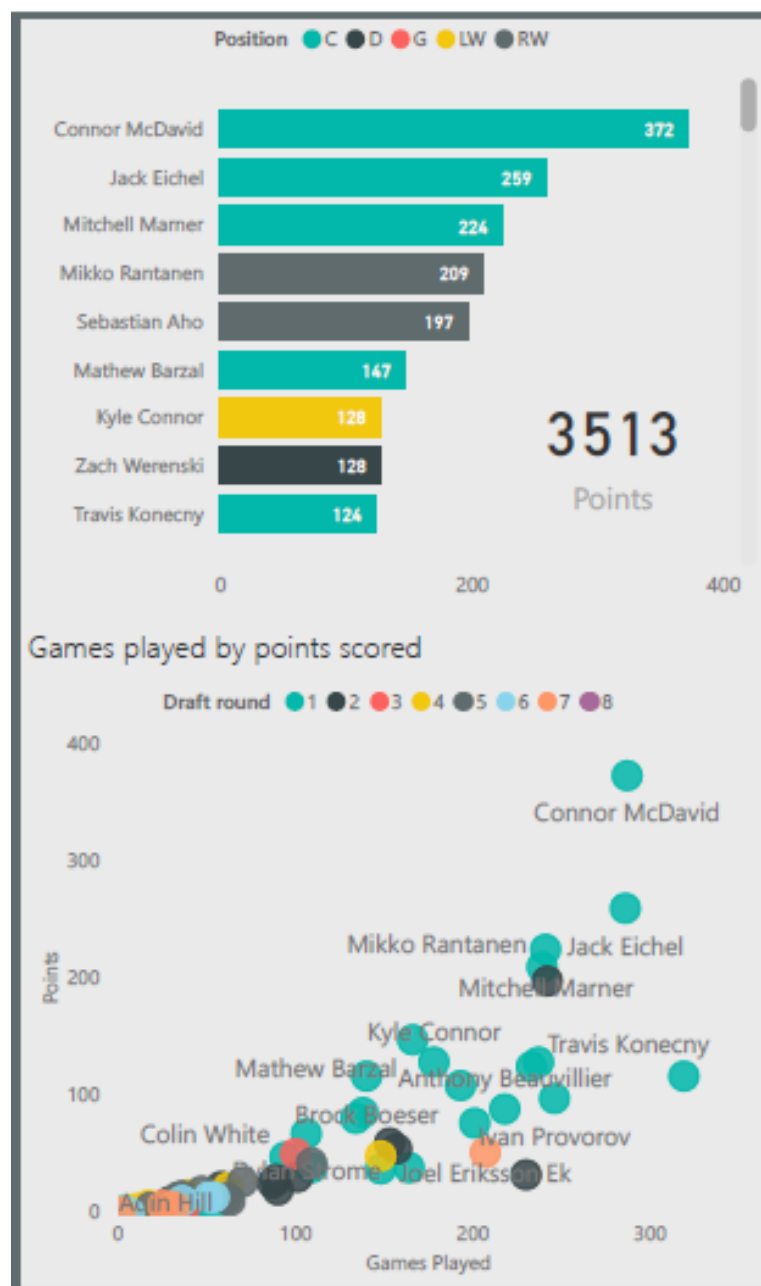
18. Data Exploration and Data Visualization
18.3 Dashboards

Exercices

Les tableaux de bord

1. En équipe ou individuellement, identifiez quelques visualisations de données qui vous attirent. Quelles histoires sont racontées par la visualisation ? Quel type de données nécessitent-elles ?
2. En équipe ou individuellement, identifiez des scénarios de travail pour lesquels la visualisation de données pourrait s'avérer utile. Quel aperçu pourrait être tiré de telles visualisations ? Ces visualisations obtiendraient-elles l'adhésion de vos superviseurs/employeurs ? Quelle quantité de travail serait nécessaire pour passer de la conception à la réalisation ? Les obstacles sont-ils principalement de nature technique ? Liés à l'acquisition de données ?
3. Considérez les tableaux de bord suivants. Pouvez-vous déterminer, en un coup d'œil, qui est leur public ? Quels sont leurs types ? Leurs forces et leurs faiblesses ? Comment pourriez-vous les améliorer ?

Top Scorers

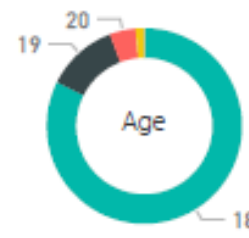
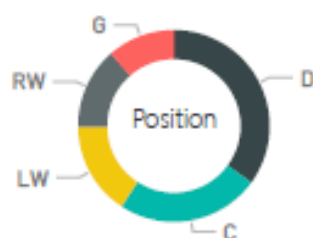
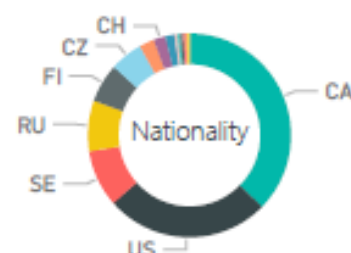
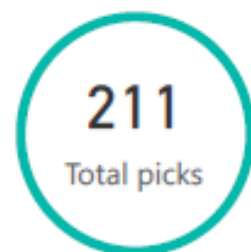


2015 NHL Draft Class

Select NHL team

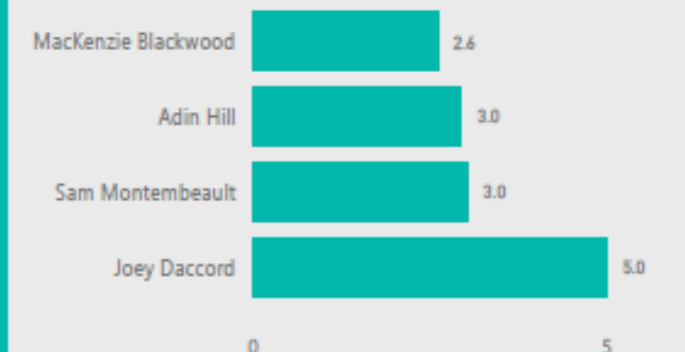
All

Overall	Player	Team	Position	Games Played
1	Connor McDavid	Edmonton Oilers	C	287
2	Jack Eichel	Buffalo Sabres	C	286
3	Dylan Strome	Arizona Coyotes	C	106
4	Mitchell Marner	Toronto Maple Leafs	C	241
5	Noah Hanifin	Carolina Hurricanes	D	319
6	Pavel Zacha	New Jersey Devils	C	201
7	Ivan Provorov	Philadelphia Flyers	D	246
8	Zach Werenski	Columbus Blue Jackets	D	237
9	Timo Meier	San Jose Sharks	RW	193
10	Mikko Rantanen	Colorado Avalanche	RW	239

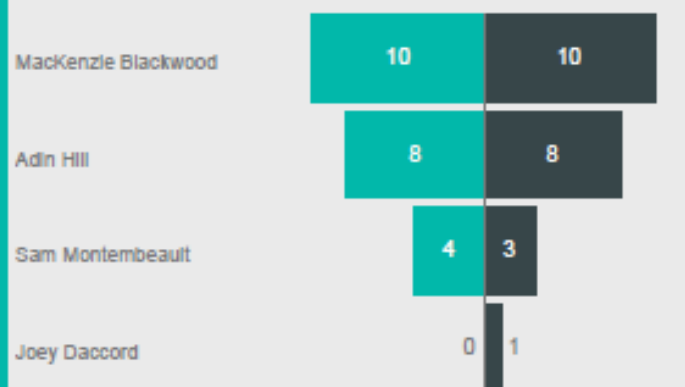


Goalies

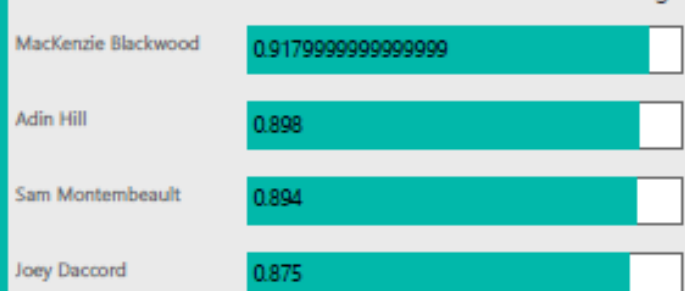
Goals Against Average



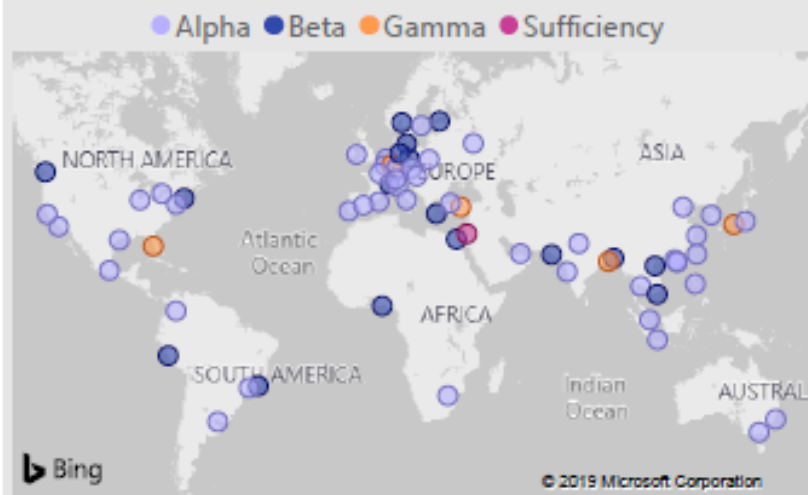
Wins VS Losses



Save Percentage



Global Cities Overview



68

Cities

44

Countries

City	Country	Rating
Bangkok	Thailand	Alpha
Brussels	Belgium	Alpha
Buenos Aires	Argentina	Alpha
Chicago	USA	Alpha
Frankfurt	Germany	Alpha
Guangzhou	China	Alpha
Istanbul	Turkey	Alpha
Jakarta	Indonesia	Alpha
Los Angeles	USA	Alpha
Madrid	Spain	Alpha
Melbourne	Australia	Alpha
Mexico City	Mexico	Alpha
Milan	Italy	Alpha
Moscow	Russia	Alpha

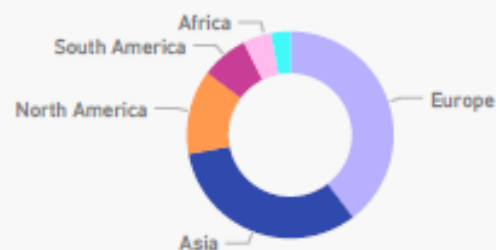
Alpha

Beta

Gamma

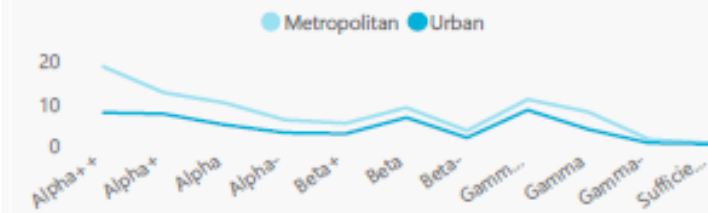
Sufficiency

Summary

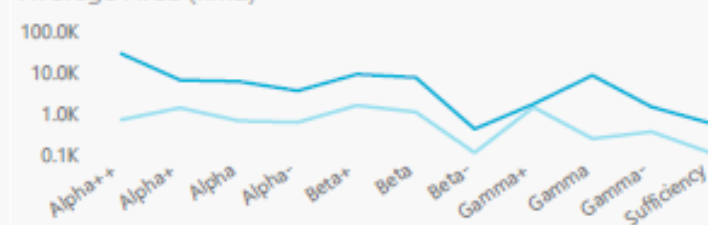


Metro vs Urban

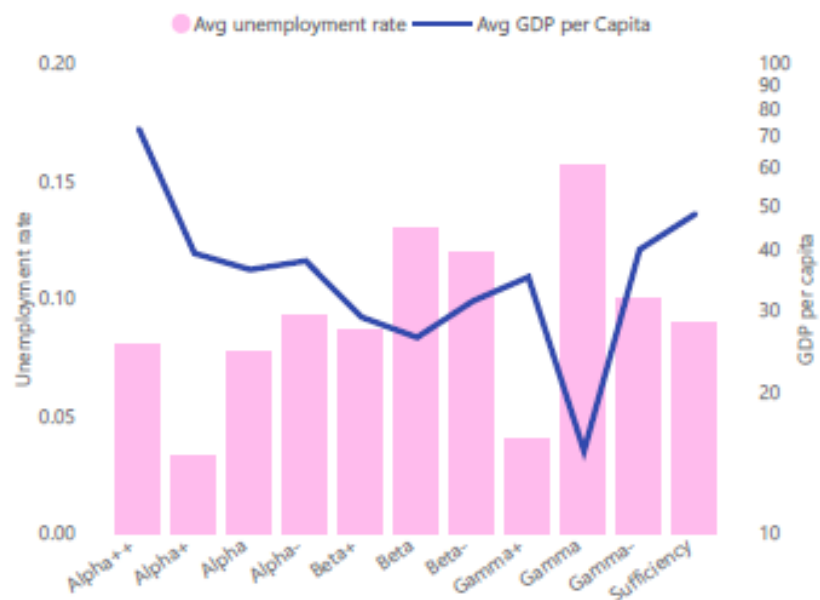
Average Population (millions)



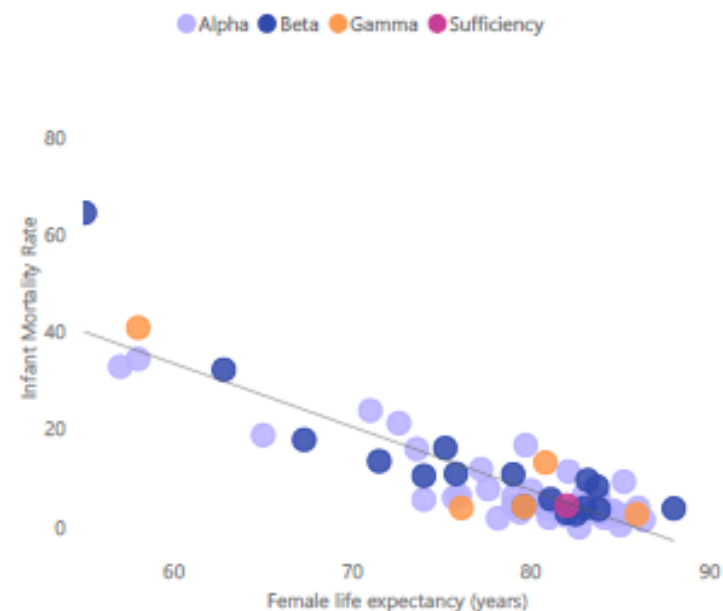
Average Area (km2)

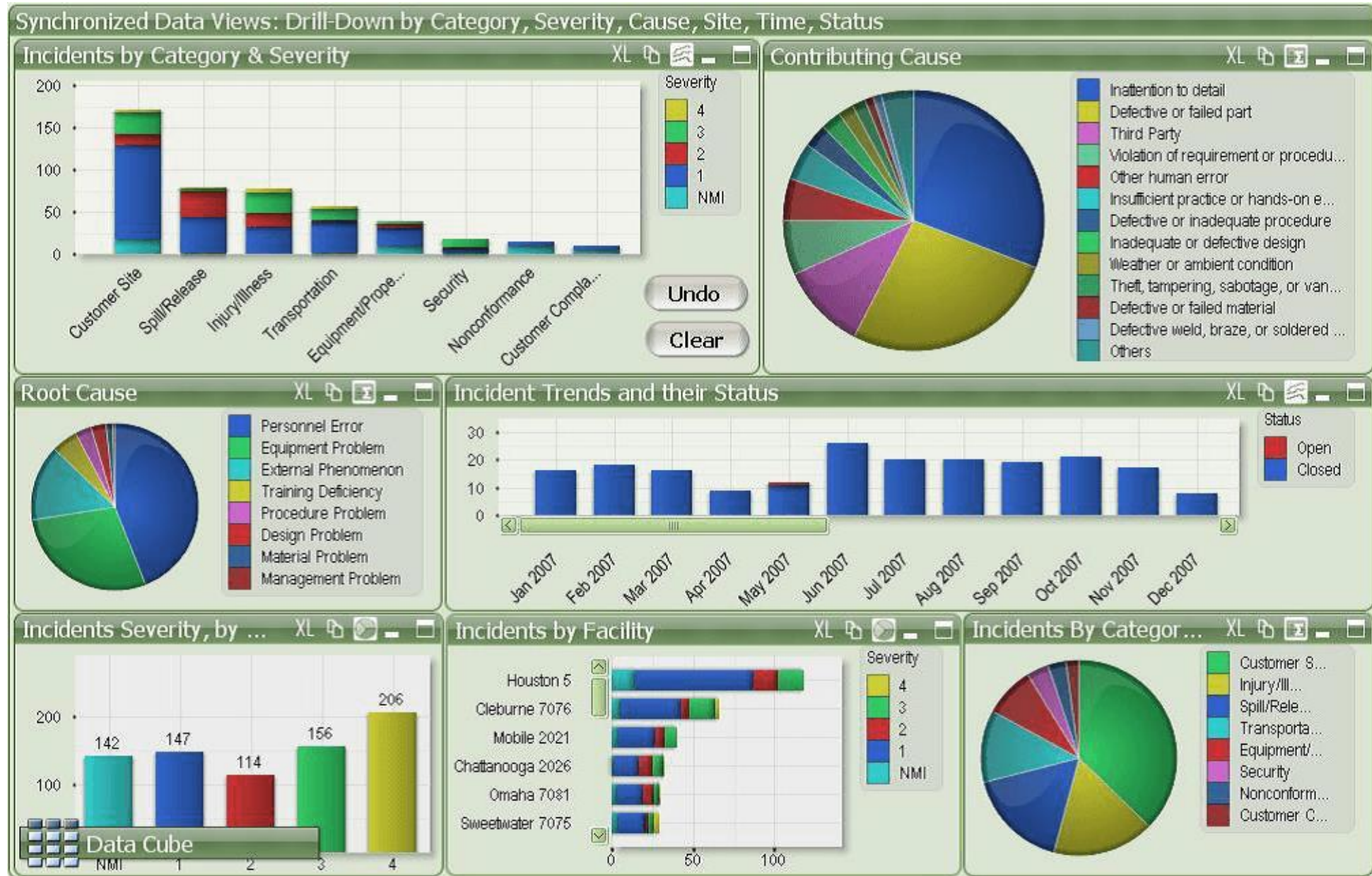


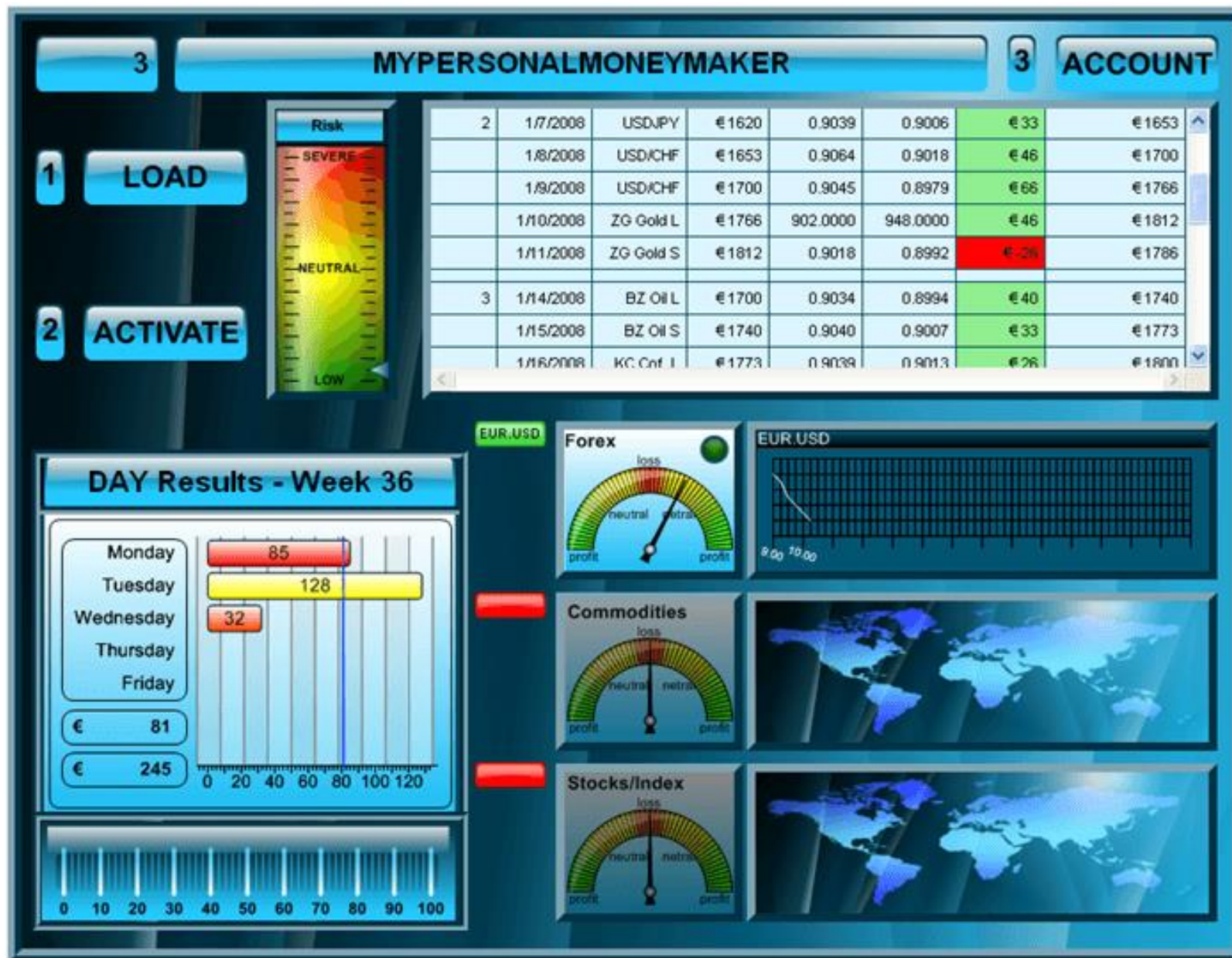
Average unemployment rate vs GDP per capita



Female life expectancy vs Infant Mortality



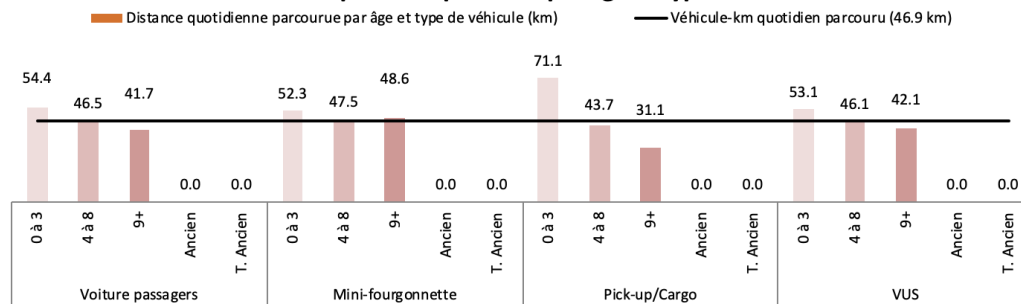




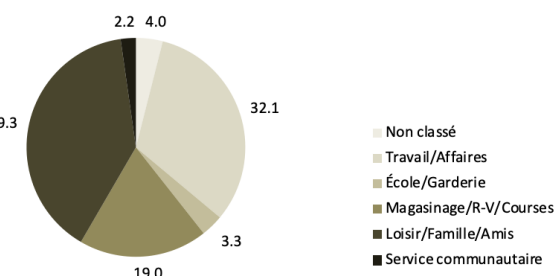
Ontario – 1er trimestre 2012

Caractéristiques des déplacements

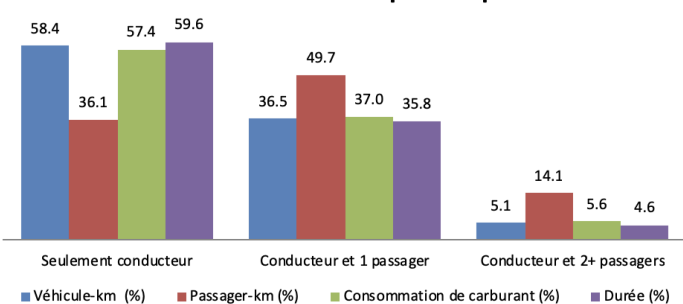
Véhicule-km quotidien parcouru par âge et type de véhicule



Passager-km quotidien parcouru par but des déplacements (%)



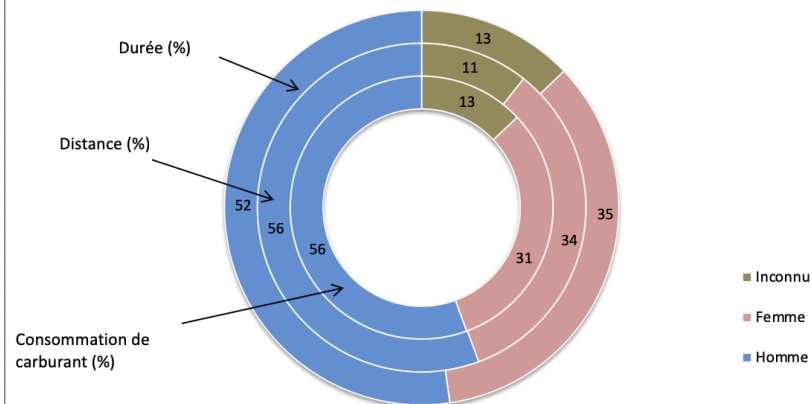
Distance, passager-km parcouru, durée et consommation de carburant par occupation



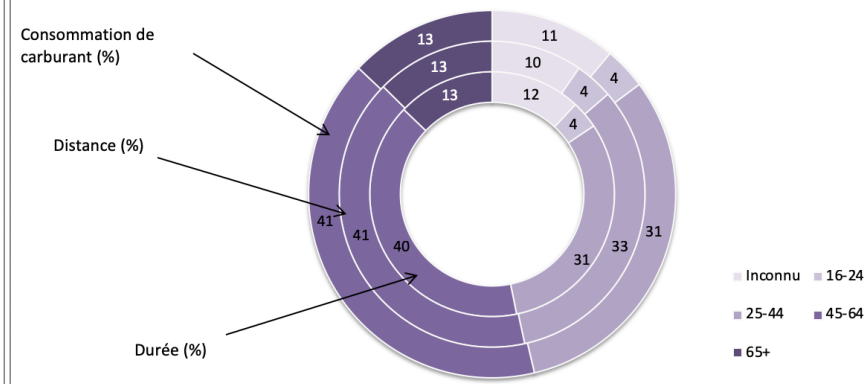
Proportion de déplacements par segments de distance



Durée, distance et consommation de carburant par sexe

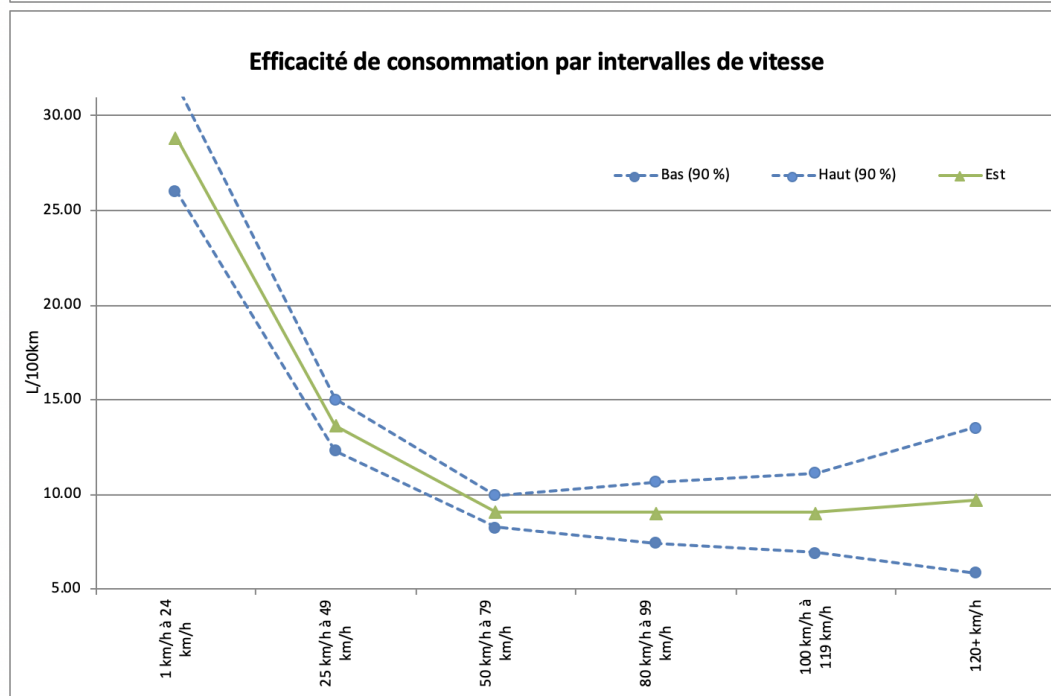
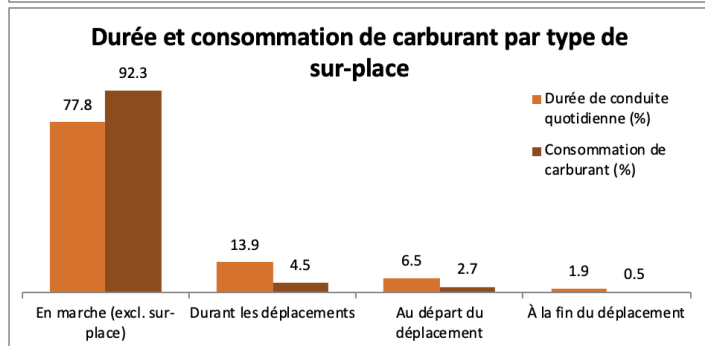
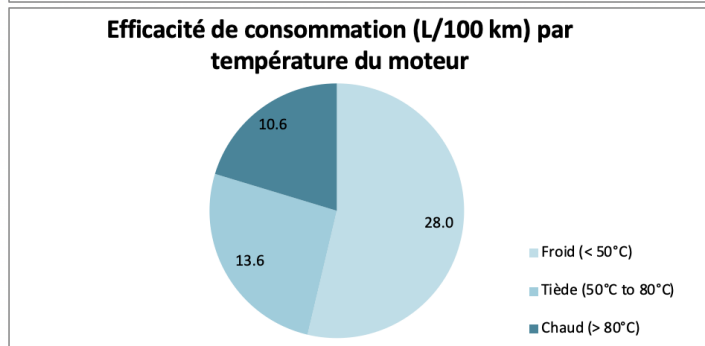
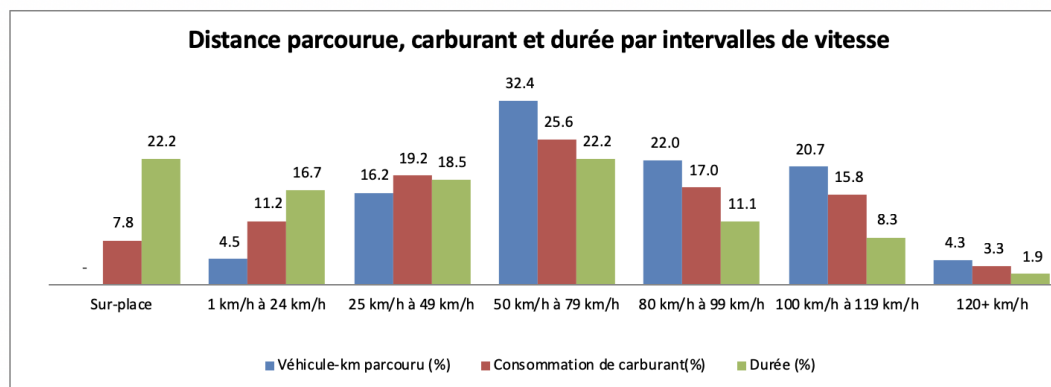
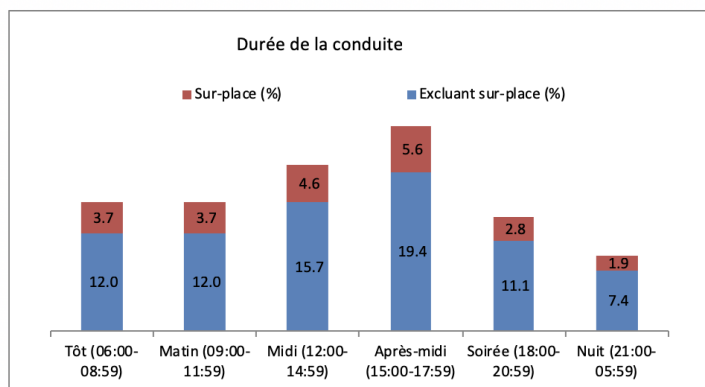


Consommation de carburant, distance et durée par âge des conducteurs



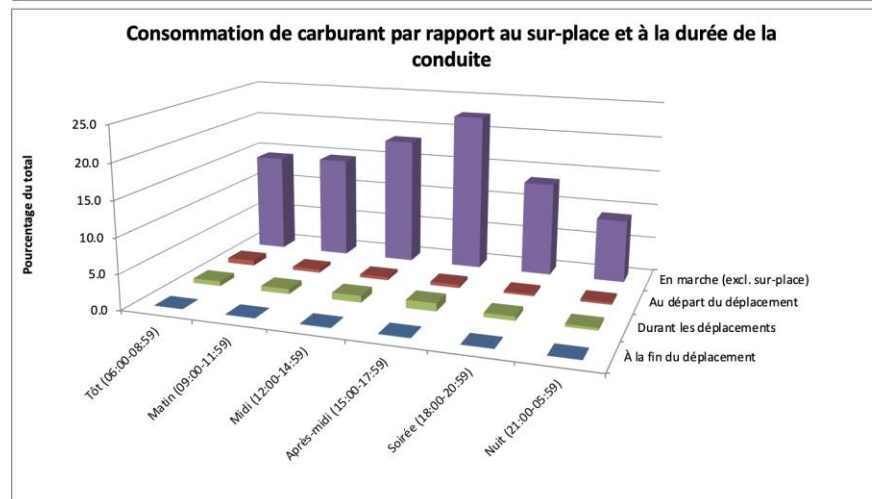
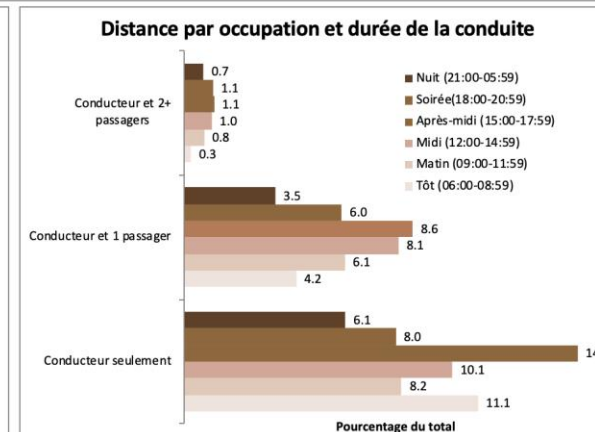
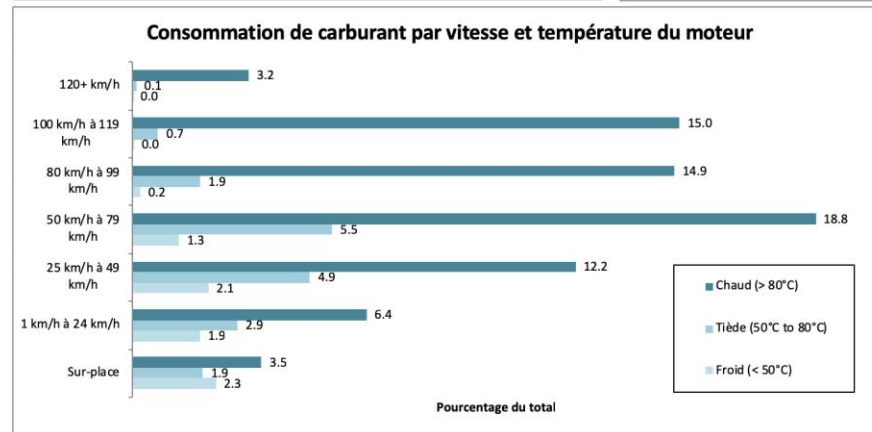
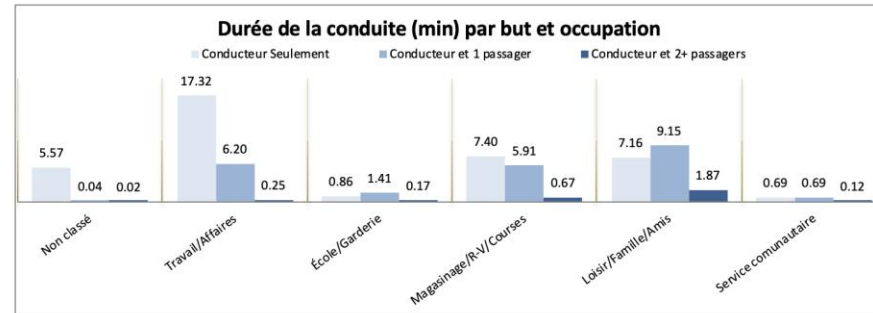
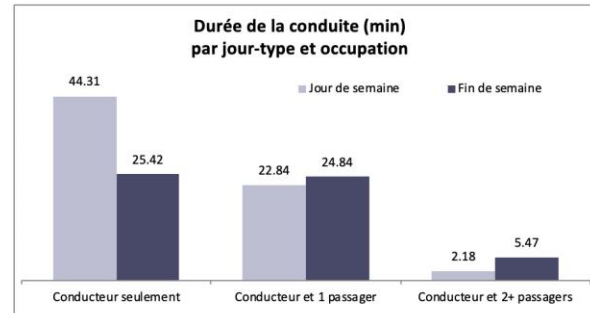
Ontario – 1er trimestre 2012

Sous-caractéristiques des déplacements



Ontario – 1er trimestre 2012

Caractéristiques mixtes sur les déplacements



What-If Analysis: Impact of Minimum Wage [https://bigbookofdashboards.com/dashboards.html]



Proposed Minimum Wage

Required Service

\$16.00

3

Developed by Matt Chambers
<http://sirvialot.blogspot.com/>

