



Introduction à la science des données

Instructeur: Patrick Boily



uOttawa

Institut de développement professionnel
Professional Development Institute



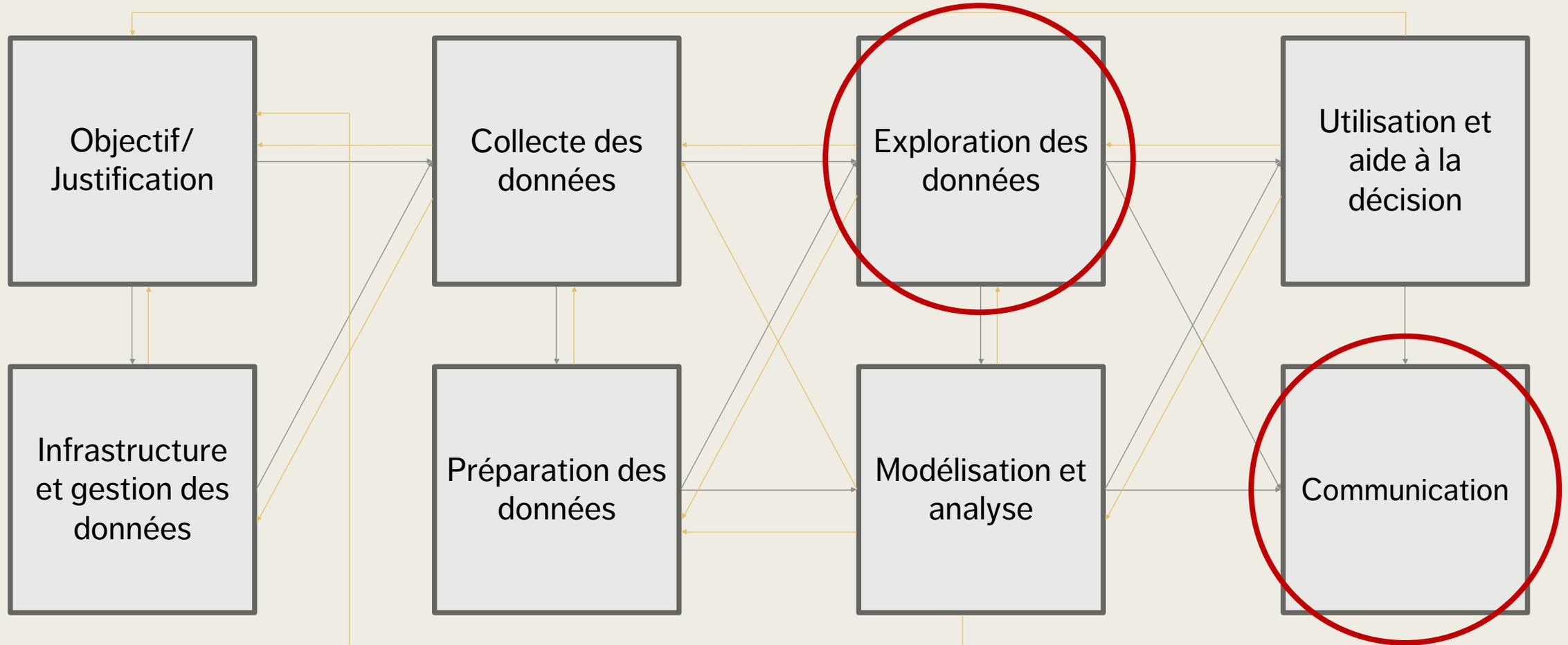
LES PRINCIPES DE LA VISUALISATION DES DONNÉES

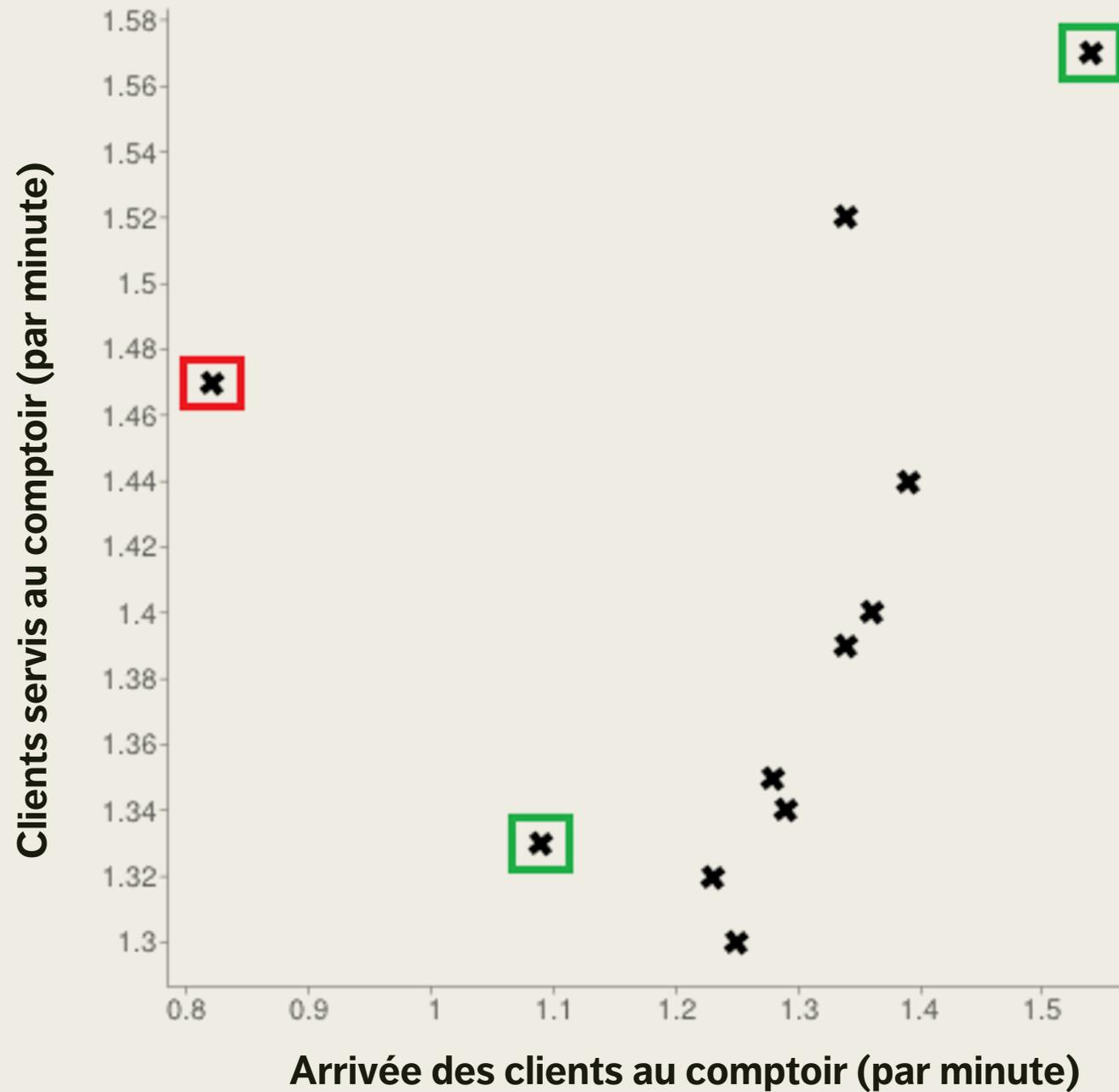
Patrick Boily

Data Action Lab | uOttawa | Idlewyld Analytics

pboily@uottawa.ca

LE « FLUX DE TRAVAIL » DE LA SCIENCE DES DONNÉES





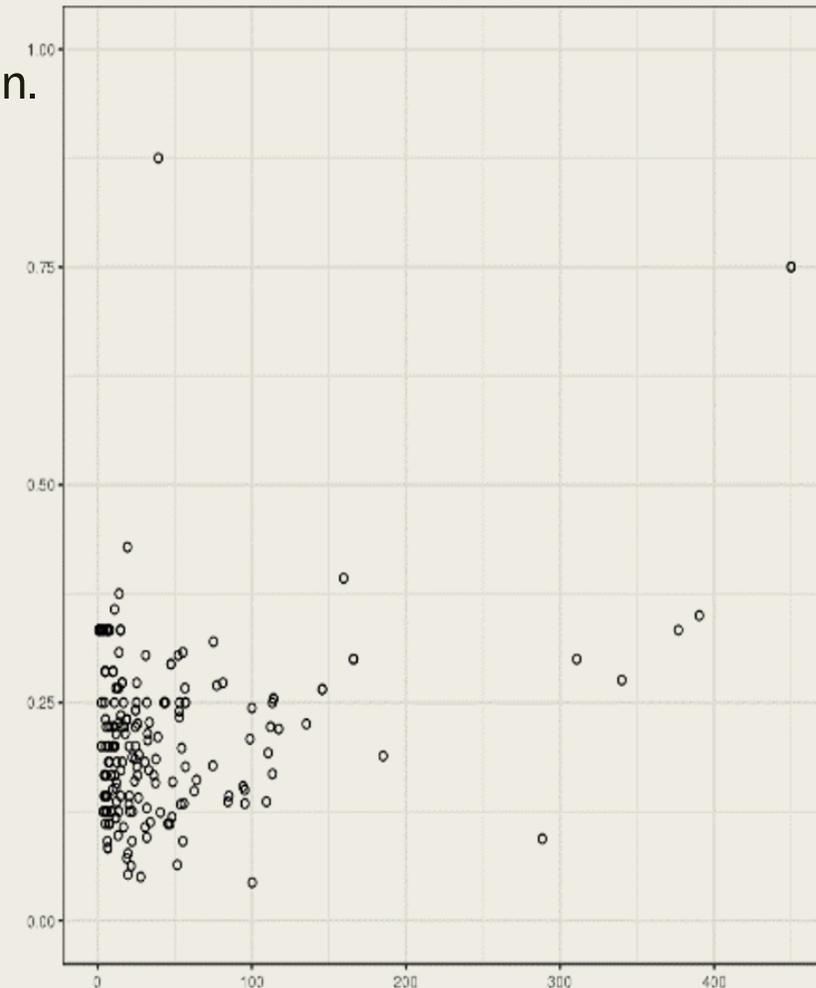
REPRÉSENTER LES OBSERVATIONS

2 variables peuvent être représentées par leur position dans le plan.

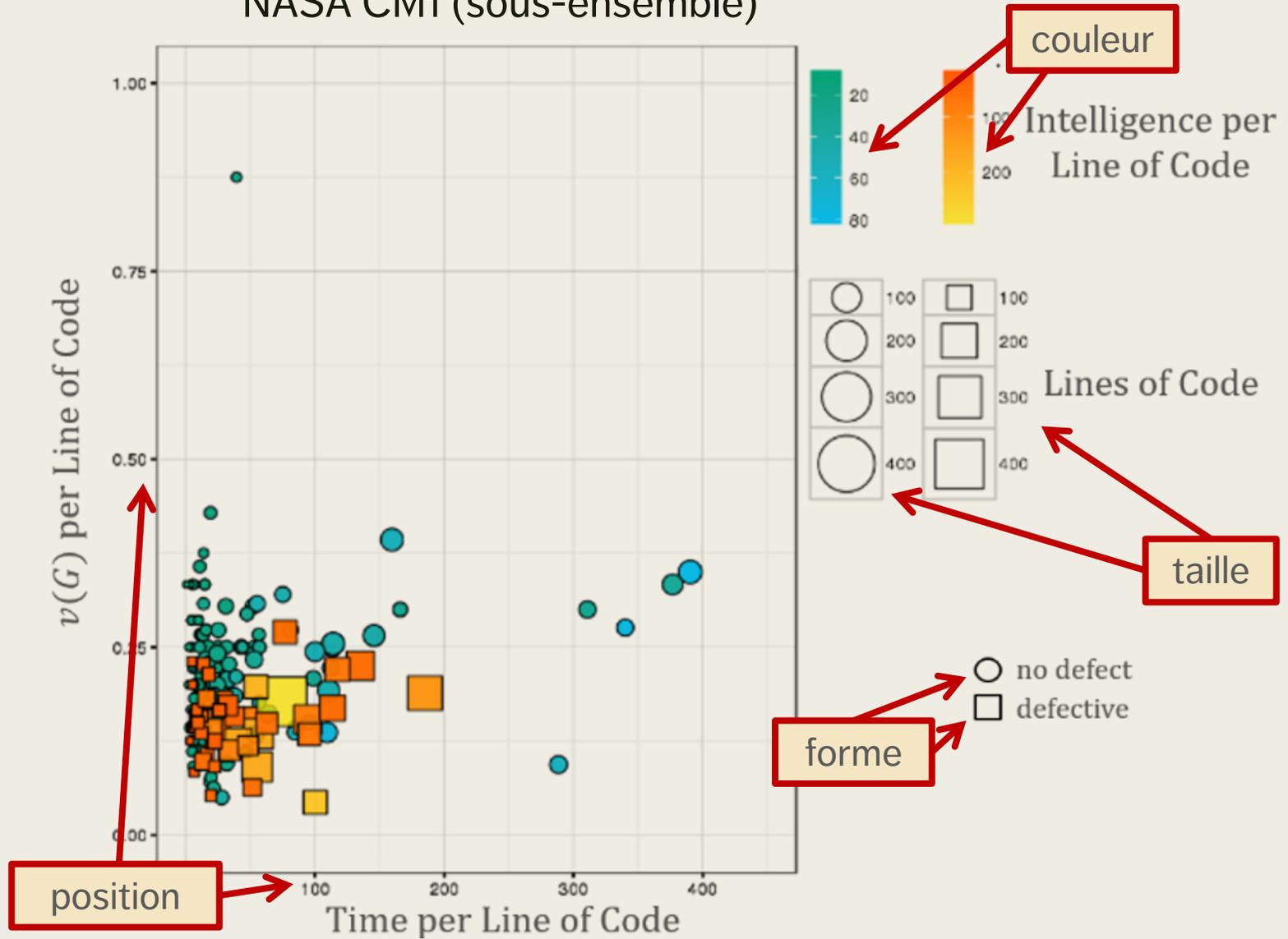
Des facteurs supplémentaires peuvent être représentés par :

- taille
- couleur
- valeur
- texture
- orientation de la ligne
- forme
- (mouvement ?)

NASA CM1 (sous-ensemble)



NASA CM1 (sous-ensemble)



VISUALISATIONS STANDARD

Graphique en ligne/graphique en creux/ligne de chiffres

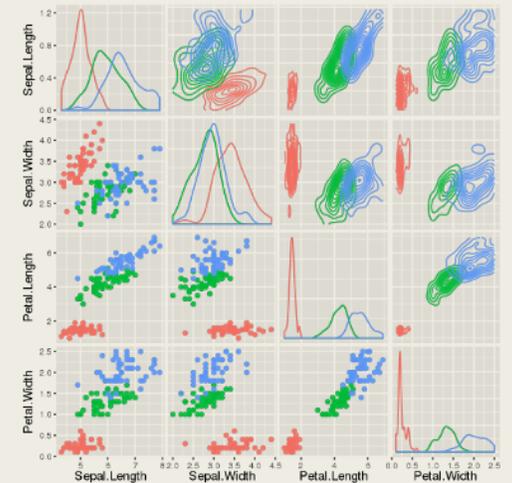
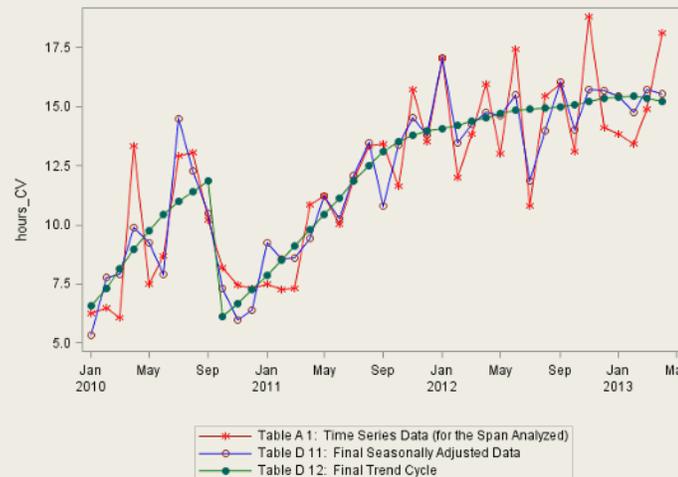
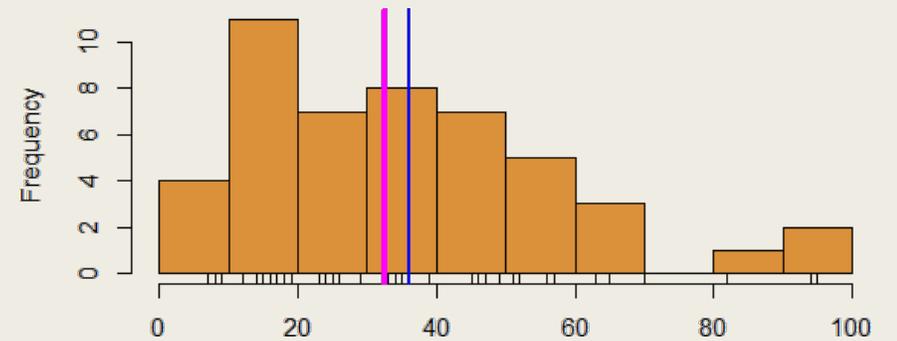
Histogramme

Graphique en ligne

Boîte à moustache

Graphique à barres

Nuage de points





LES PRINCIPES FONDAMENTAUX DE LA CONCEPTION ANALYTIQUE

Le **raisonnement** et la **communication** de nos pensées sont intimement liés à notre vie dans un univers multivariable, causal et dynamique.

Symétrie des présentations visuelles des preuves : les consommateurs devraient rechercher exactement ce que les producteurs devraient fournir, par exemple :

- comparaisons significatives
- réseaux de causalité et structure sous-jacente
- liens multivariés
- données intégrées et pertinentes
- documentation honnête
- accent mis sur le contenu

Le message est-il bien compris ?
Le message est-il transmis ?

UN MOT SUR L'ACCESSIBILITÉ

Les graphiques ne peuvent généralement pas être traduits en braille. Décrire les caractéristiques et les structures émergentes d'une visualisation est une solution possible... **si on peut les repérer.**

Les analystes doivent produire des visualisations claires et significatives, mais ils doivent également les décrire, ainsi que leurs caractéristiques, de manière à ce que tout le monde puisse « voir » les informations. Pour cela, il faut que les analystes aient « vu » toutes les informations, ce qui n'est pas toujours possible.

Conditions : daltonisme, basse vision, handicap moteur, handicap cognitif, TDAH, etc.

Meilleures pratiques : texte/éléments à fort contraste, zoom/agrandissement, navigation au clavier, conception assistée, résumés courts, fonction défaire/refaire, etc. [F. Elavsky]

A CLASSIFICATION OF CHART TYPES

Data comparison charts

Data reduction charts

Comparison

Composition

Distribution

Evolution

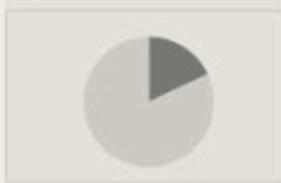
Relationship

Profiling

Bars



Pie



Histogram



Line



Scatterplot



Grouped bars



Dot plot



Bullet



Pareto



ID Scatterplot



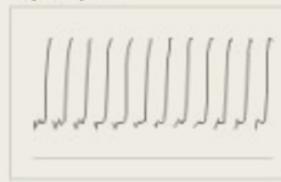
Horizon



Connected Scatterplot



Cycle plot



Scatterplot matrix



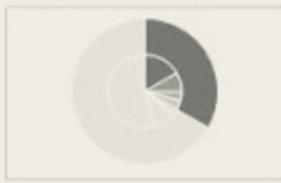
ID Scatterplot



Heat map



Multidimensional Pie



Boxplot



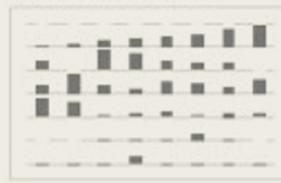
Step



Bubble



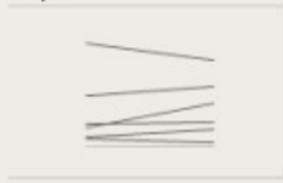
Reorderable matrix



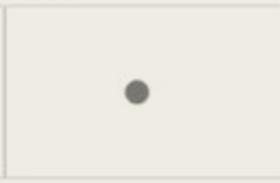
Horizon



Slope



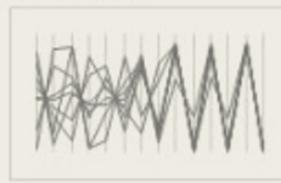
Alert



Connected Scatterplot



Parallel Plot



Trellis



TYPES DE GRAPHIQUES

Texte simple et tableaux

Diagrammes de dispersion

Graphiques en ligne

Diagrammes à barres

Diagrammes à barres empilées

Diagrammes à barres à 100%

Graphiques de zone

Treemaps

Graphiques à jauges

Cartes thermiques, choroplèthes

Cartes géographiques

Coordonnées parallèles

Visages de Chernoff

Nuages de mots

Diagrammes de réseau

Dendrogrammes et arbres

Sparklines

Graphiques interactifs

Petits multiples

etc.

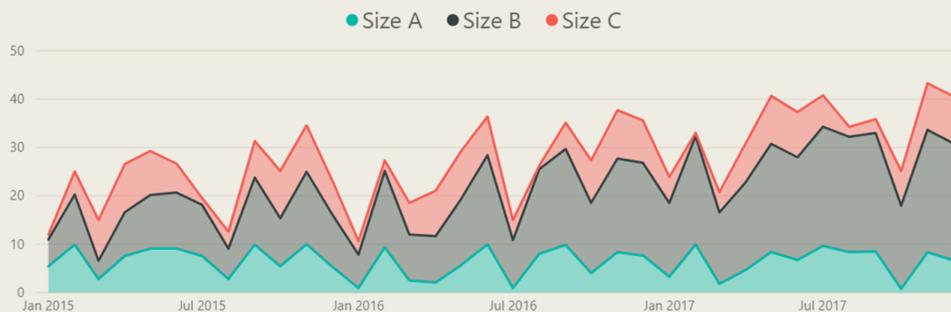
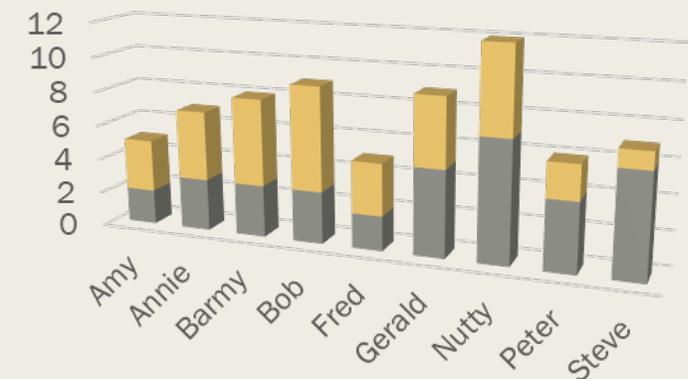
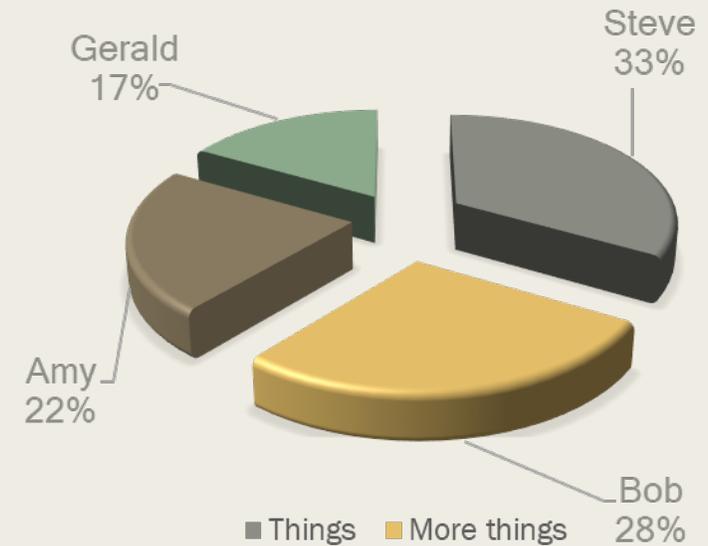
GRAPHIQUES À ÉVITER



ÉVITEZ (?) tout ce qui comporte un arc (à l'exception de graphiques à jauge) : tarte, beignet, etc. : le cerveau humain a du mal à **comparer les arcs** – sans étiquette, quelle est la différence entre Steve et Bob ?

ÉVITEZ les graphiques en 3D : il est difficile de les comparer visuellement (et ils sont trop **encombrés**).

ÉVITEZ les graphiques à zones empilées : ils sont bcp trop déroutants.



POINTS À RETENIR

Effective data visualizations **provide insights** and **facilitate understanding**.

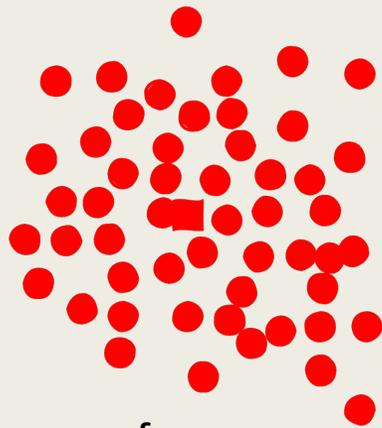
The basic principles can guide your visualization design and consumption.

Be **creative** but keep your data and your representations **honest**.

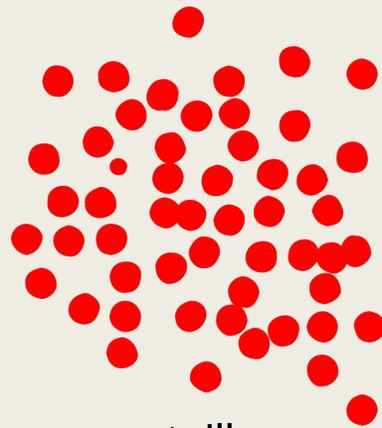
Be mindful of attempts to distort trends and conclusions with flashy visuals.

Data and code should be made available along with the displays.

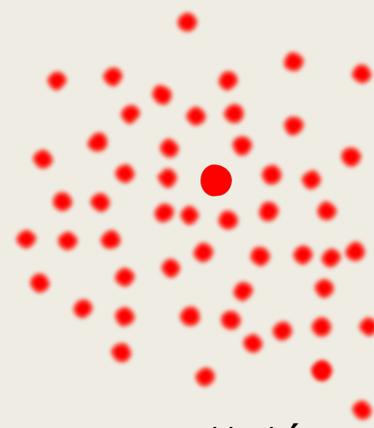
ATTRIBUTS PRÉ-ATTENTIFS



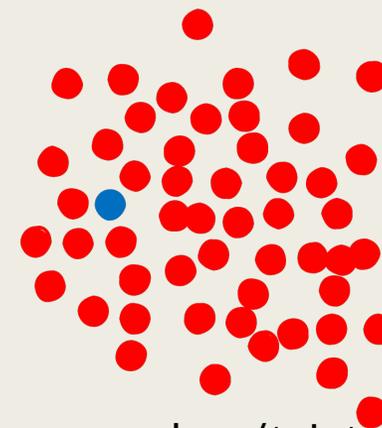
form



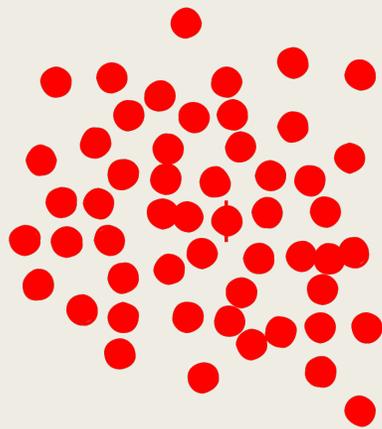
taille



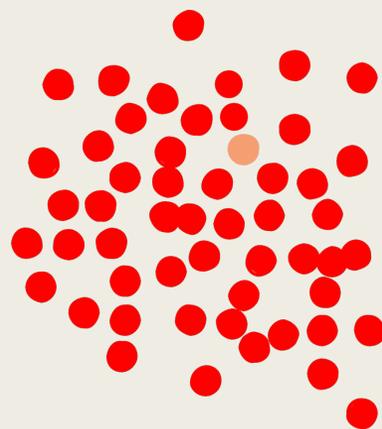
netteté



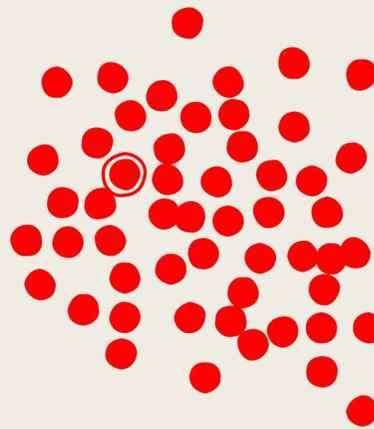
couleur/teinte



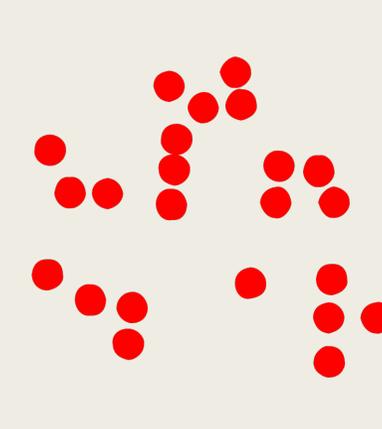
marquages



intensité/valeur



clôture



numérosité

DÉSENCOMBREMENT

LE DÉSORDRE, C'EST L'ENEMI !

- chaque élément d'une page ajoute une **charge cognitive**
- Identifiez tout ce qui n'apporte pas de valeur ajoutée et supprimez-le.
- Considérez la charge cognitive comme l'effort mental nécessaire pour traiter l'information (plus il est faible, mieux c'est).
- Tufte fait référence au **ratio données/encre** : « plus la part d'encre d'un graphique consacrée aux données est importante, mieux c'est ».
- Dans Resonate, Duarte parle de « maximiser le rapport signal/bruit », le signal étant l'information ou l'histoire que nous voulons communiquer.

DÉSENCOMBREMMENT

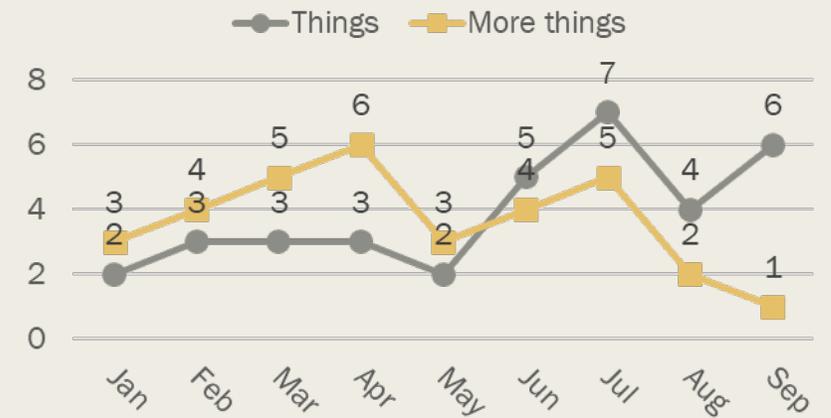
Utilisez les **principes de la Gestalt** pour organiser/souligner les données dans un graphique.

Alignez tous les éléments (graphiques, texte, lignes, titres, etc.).

- **Ne vous fiez pas** à l'œil, utilisez des cases de position et des valeurs.

Graphiques :

- supprimez les bordures, les lignes de la grille, les marqueurs de données
- clarifiez les étiquettes des axes
- étiqueter les données directement



DÉSENCOMBREMENT

Utilisez une police, une taille de police, une couleur et un alignement **uniformes**.

Ne faites pas pivoter le texte à un angle autre que 0 ou 90 degrés.

Utilisez des **espaces blancs** :

- les marges doivent rester libres de texte et de visuels
- n'étirez pas les visuels jusqu'au bord de la page ou trop près d'autres visuels
- considérez l'espace blanc comme une bordure

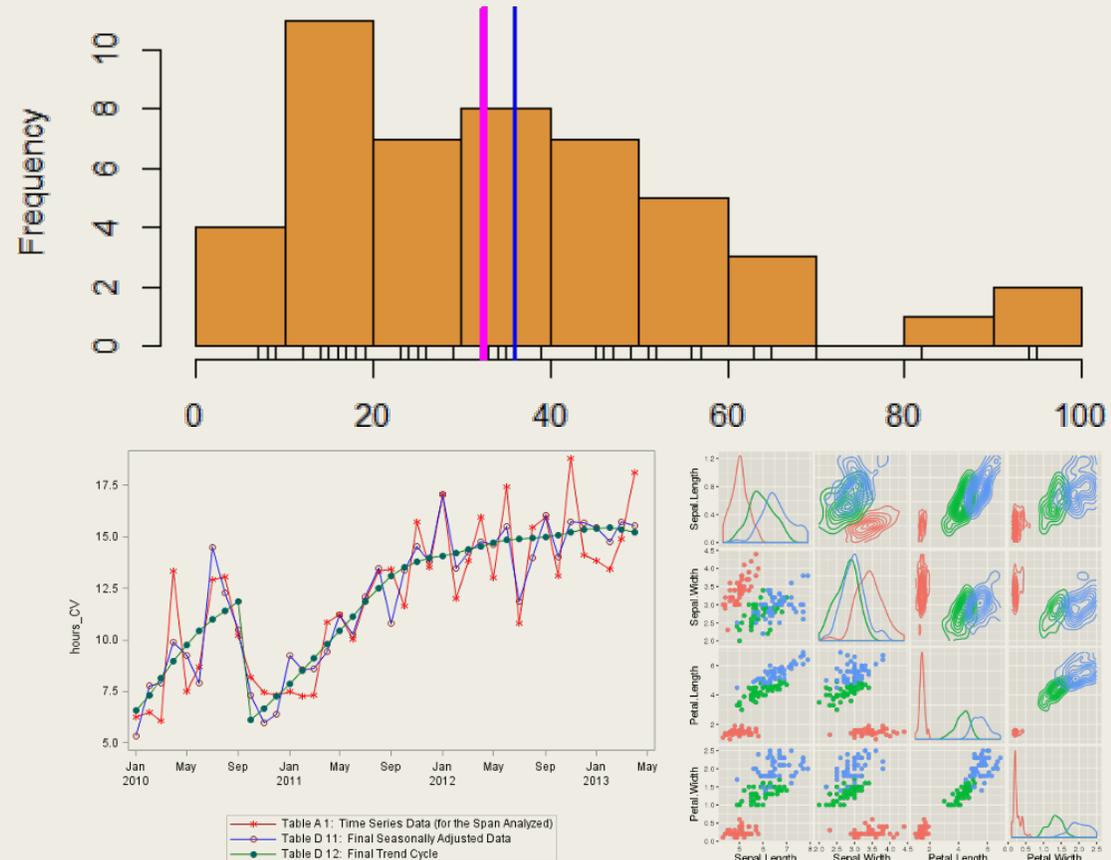
TAILLES DES GRAPHIQUES

En supposant que le tableau ait été désencombré :

- les choses d'importance égale ont la **même taille** ;
- la taille des autres éléments est proportionnelle à leur **importance**.

Comme il est rare que l'on mette plus de 3 ou 4 tableaux sur une page, les options de taille sont limitées.

Exception perpétuelle : les **cartes géographiques** peuvent nécessiter plus d'espace.

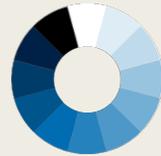


PALETTES DE COULEURS

Achromatique



Monochromatique



Complémentaire



Complémentaire divisé



Complémentaire divisé à gauche (ou à droite)



Analogique



Diade de couleurs



Triade de couleurs



Tétrade de couleurs



PALETTES DE COULEURS

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

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

Le cas échéant, choisissez un schéma basé sur l'identité de l'entreprise (cela maximise l'adhésion).

Créez un modèle (et respectez-le).

Téléchargez des images pour voir à quoi ressemblent les graphiques en fonction des différents degrés de daltonisme :

- <https://www.color-blindness.com/coblis-color-blindness-simulator> (*il existe d'autres outils*)