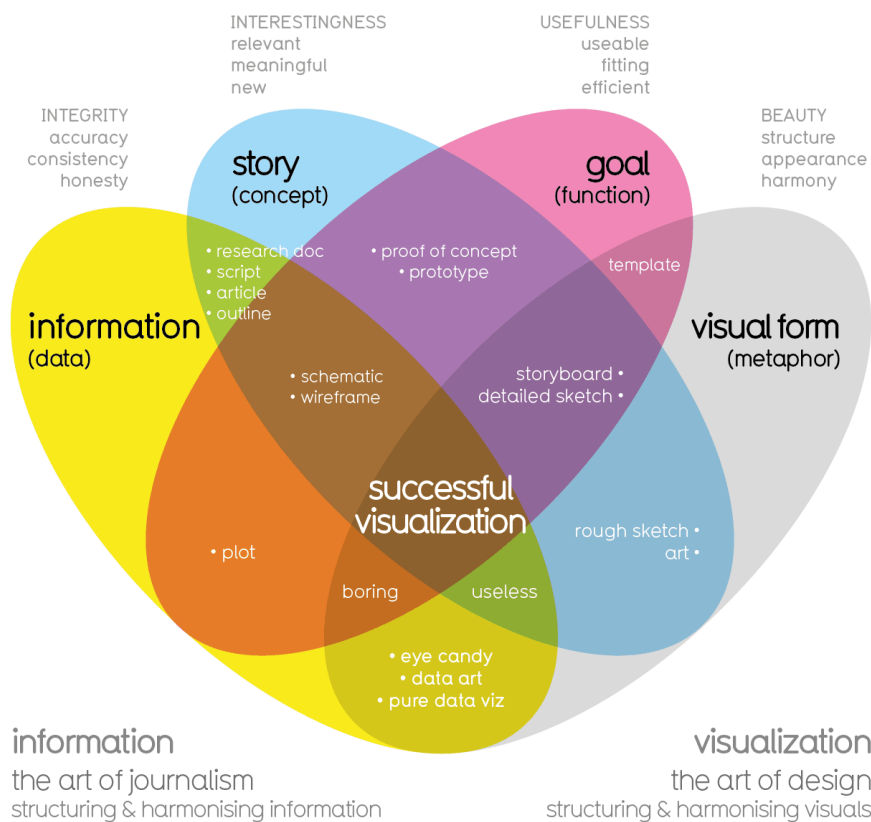




## DataViz - Itération 2

*Quel graphique pour quelle information ?*

### What Makes a Good Visualization?



David McCandless  
InformationisBeautiful.net

find out more  
[bit.ly/KIB\\_Books](https://bit.ly/KIB_Books)

## Objectifs de la journée

L'objectif de cette journée est de vous familiariser avec les graphiques dits "classiques": quand et comment les utiliser, mais aussi quand ne *pas* les utiliser. Vous allez explorer les pièges (tout aussi classiques) à éviter à tout prix et comment améliorer chacun de ces graphiques pour en sortir son plein potentiel de storytelling!

## Modalités

- Durée du projet : 1 jour
- Ce projet sera réalisé en équipe pour la plus grande partie
- Des slides serviront de support aux discussions tout au long de la journée (ces slides seront mis à disposition sur le google drive à la fin de la journée).

## Démarche Pédagogique

Cette itération est découpée en 4 étapes :

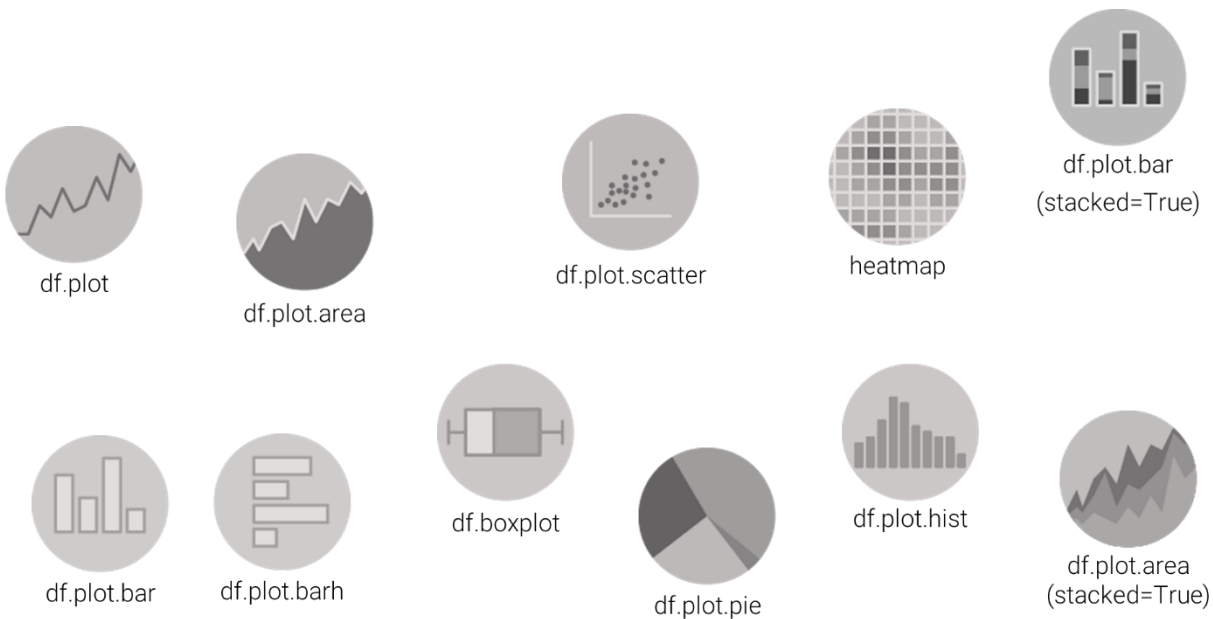
1. Comprendre les intérêts de chacun des graphes les plus classiques
2. Un peu de lecture
3. Revue de pièges courants
4. Étude de cas

## Compétences développées

Cette journée permet de valider la compétence suivante sur Campus Skills:

- **Connaître les types de graphiques "classiques" (line, hist, scatter, bar, pie, boxplot, area, heatmap) et savoir quand les utiliser:** *Je suis capable de choisir le type de graphique le plus approprié pour transmettre un message en fonction du type de données disponible.*

## Étape 1 : Les graph classiques



### Objectifs de l'activité :

Se familiariser avec les graphiques dits "classiques" et savoir quand les utiliser :

- Selon le type de données disponibles : *Variables qualitatives ou quantitatives*
- Selon l'information que l'on souhaite faire passer : *distribution, corrélation, classement, proportion, évolution ...*

### Consignes

Travail en équipes de 3 ou 4 personnes.

1. Pour chacun des deux jeux de données, proposer des visualisations (parmi les graphs classiques listés ci-dessus) qui permettent de répondre aux questions suivantes:

- **Dataset 1 : Titanic (dataset très classique fourni par Kaggle)**
  - Quelle est la répartition des passagers par classe ?
  - Quelles différences selon le sexe ?
  - Distribution des passagers selon âge et sexe
  - Quel est le taux de survie selon le sexe ? l'âge ? la classe ?
  - Y a-t-il un lien entre prix payé pour le billet et l'âge des passagers ?
- **Dataset 2 : Consommation et productions nationales**
  - Comment varient les différentes grandeurs présentées dans ce jeu de données ?
  - Quel est l'évolution de la consommation ?
  - Quand a-t-on un excédant / un déficit de production ?
  - Y a-t-il une corrélation entre les différents types de production ?

⚠ seul le **type** de graphique importe ici (pas son style) ⚠

2. Créez un tableau avec les différentes figures produites (*une capture d'écran de la figure suffit*). Spécifiez le type de graph (scatter, line plot, ...) ainsi que le type de variables représentées (quantitatives ou qualitatives). Entamez une réflexion sur la meilleure manière de faire passer le message (c'est-à-dire de répondre à la question posée): que manque t'il à la figure, quelles sont les simplifications possibles, l'usage de couleurs etc...
3. Téléchargez le tableau sur le drive afin de pouvoir comparer les différents tableaux créés par chaque groupe. On en discute ensemble en fin de matinée!

### Compétence développée :

Je sais choisir le type de graph le plus pertinent permettant de répondre à une question posée.

## Étape 2 : Un peu de lecture

Prenez le temps de lire les deux ressources suivantes :

- Comment créer un graphique : [What Questions to Ask When Creating Charts](#)
- Le choix des couleurs : [What to consider when choosing colors for data visualization](#)

## Étape 3 : Pièges courants

**Objectifs de l'activité :**

Bar plot, line plot, pie chart, et boxplot... quels sont les problèmes les plus courants ?

**Data-to-viz, une mine d'informations:**

Concernant les pièges à éviter : [A collection of dataviz caveats by data-to-viz.com](#)

Pour savoir quel graphique choisir : [From data to Viz](#)

**Consignes :**

- Réfléchir ensemble aux « pièges » présentés
- Passer un peu de temps à explorer l'arbre proposé par [From data to Viz](#)  
→ est-ce que vous auriez répondu différemment aux questions posées dans [l'Étape 1](#)?

## Étape 4 : Études de cas

**Objectifs de l'activité :**

Pour plusieurs cas d'étude: analyse et correction de figures.

**Ces études de cas proviennent du livre :**

[Storytelling with data](#) de Cole Nussbaumer Knafic

**Consignes :**

Pour chacune des figures projetées :

- Par équipe de 4, lister les défauts des éléments composant la figure (type de graphique, couleurs, titre, légende, ...) et proposer des améliorations
- Comparaison des figures créés par chaque équipe
- Analyse de la figure proposée par l'auteur

**Note:** les propositions de figures améliorées seront réalisées avec du papier/crayon, afin de se concentrer sur l'aspect visualisation, oubliez le code! Par contre n'oubliez pas les couleurs! 😊