

Introduction de la formation **DataViz** Unité **Démographie économique** (Ined)

Bénédicte Garnier (Service Méthodes
Statistiques)


Datavisualization : contraintes

- de *lisibilité* : le résultat doit être compréhensible directement ;
- de *pertinence* : la visualisation doit avoir du sens ;
- de *validité* : les choix de représentation doivent être corrects statistiquement ;
- d'*interactivité* : Dans l'idéal, l'utilisateur doit pouvoir naviguer dans le résultat pour mieux comprendre et analyser.

Transformer pour faire passer un message

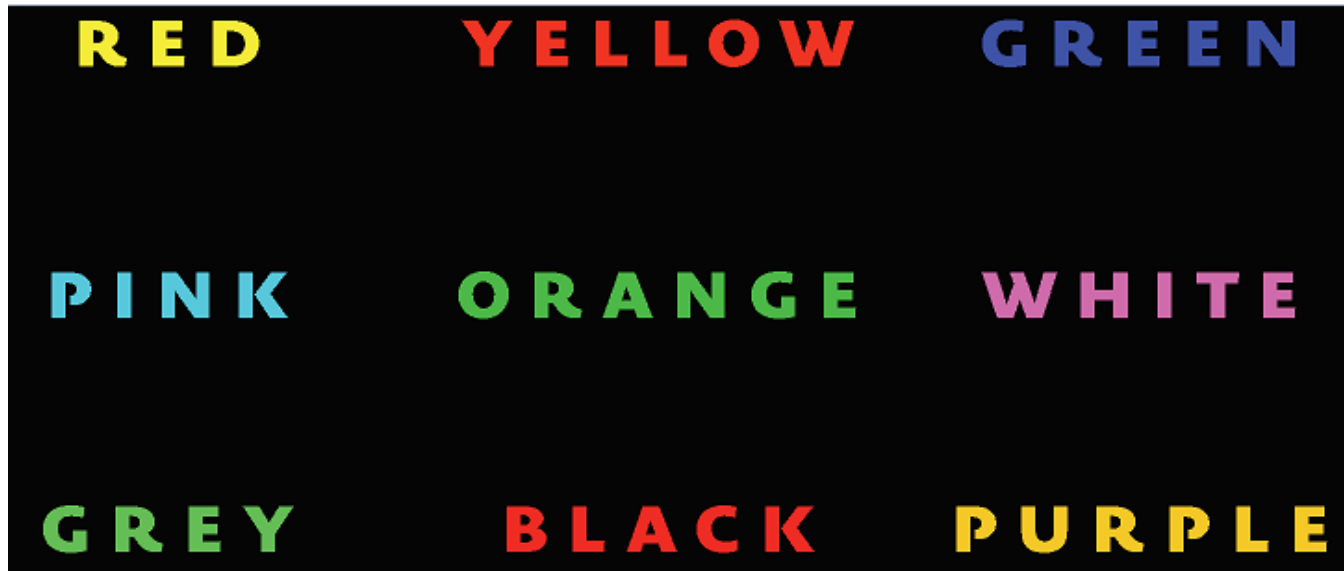
Exemple : usage de la couleur

3	.	1	4	1	5	9	2	6	5	3	5	8	9	7	9		3	.	1	4	1	5	9	2	6	5	3	5	8	9	7	9														
3	2	3	8	4	6	2	6	4	3	3	8	3	2	7		3	2	3	8	4	6	2	6	4	3	3	8	3	2	7		3	2	3	8	4	6	2	6	4	3	3	8	3	2	7
9	5	0	2	8	8	4	1	9	7	1	6	9	3	9		9	5	0	2	8	8	4	1	9	7	1	6	9	3	9		9	5	0	2	8	8	4	1	9	7	1	6	9	3	9
9	3	7	5	1	0	5	8	2	0	9	7	4	9	4		9	3	7	5	1	0	5	8	2	0	9	7	4	9	4		9	3	7	5	1	0	5	8	2	0	9	7	4	9	4
4	5	9	2	3	0	7	8	1	6	4	0	6	2	8		4	5	9	2	3	0	7	8	1	6	4	0	6	2	8		4	5	9	2	3	0	7	8	1	6	4	0	6	2	8
6	2	0	8	9	9	8	6	2	8	0	3	4	8	2		6	2	0	8	9	9	8	6	2	8	0	3	4	8	2		6	2	0	8	9	9	8	6	2	8	0	3	4	8	2
5	3	4	2	1	1	7	0	6	7	9	8	2	1	4		5	3	4	2	1	1	7	0	6	7	9	8	2	1	4		5	3	4	2	1	1	7	0	6	7	9	8	2	1	4

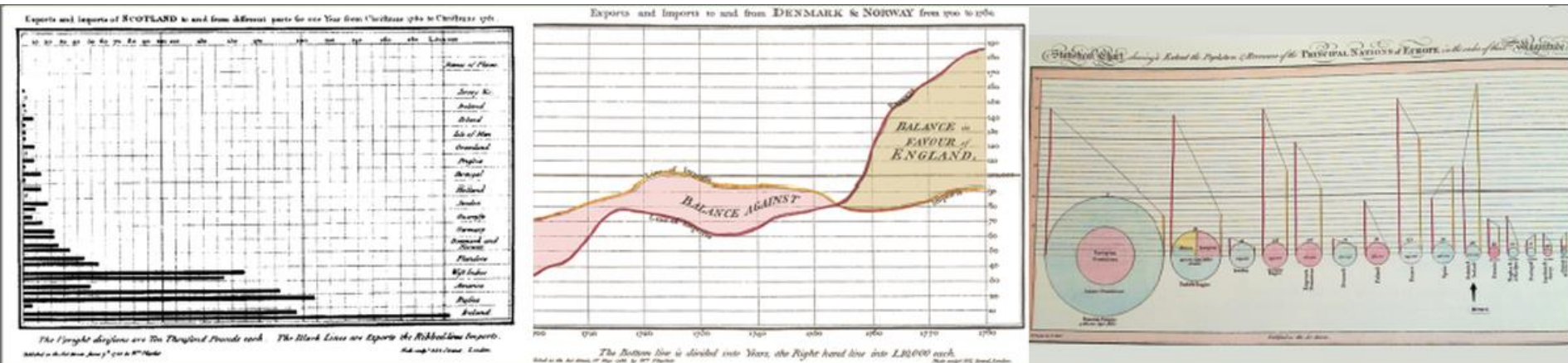


Conventions et usages

Le producteur encode les informations avec des *signes* et le lecteur décode les informations avec *ses codes* et *ses conventions* et *ses a priori*

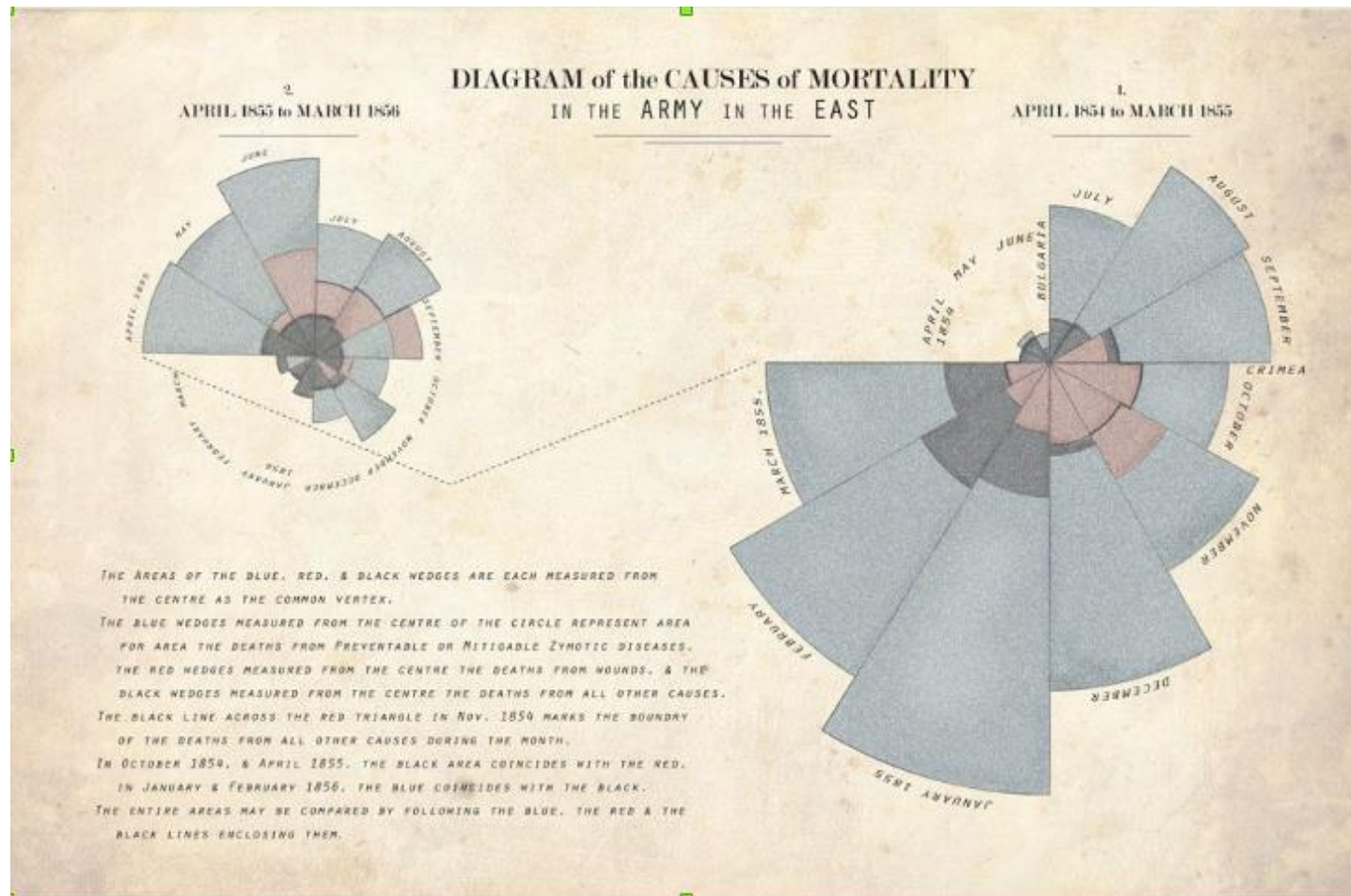


William Playfair (1786)

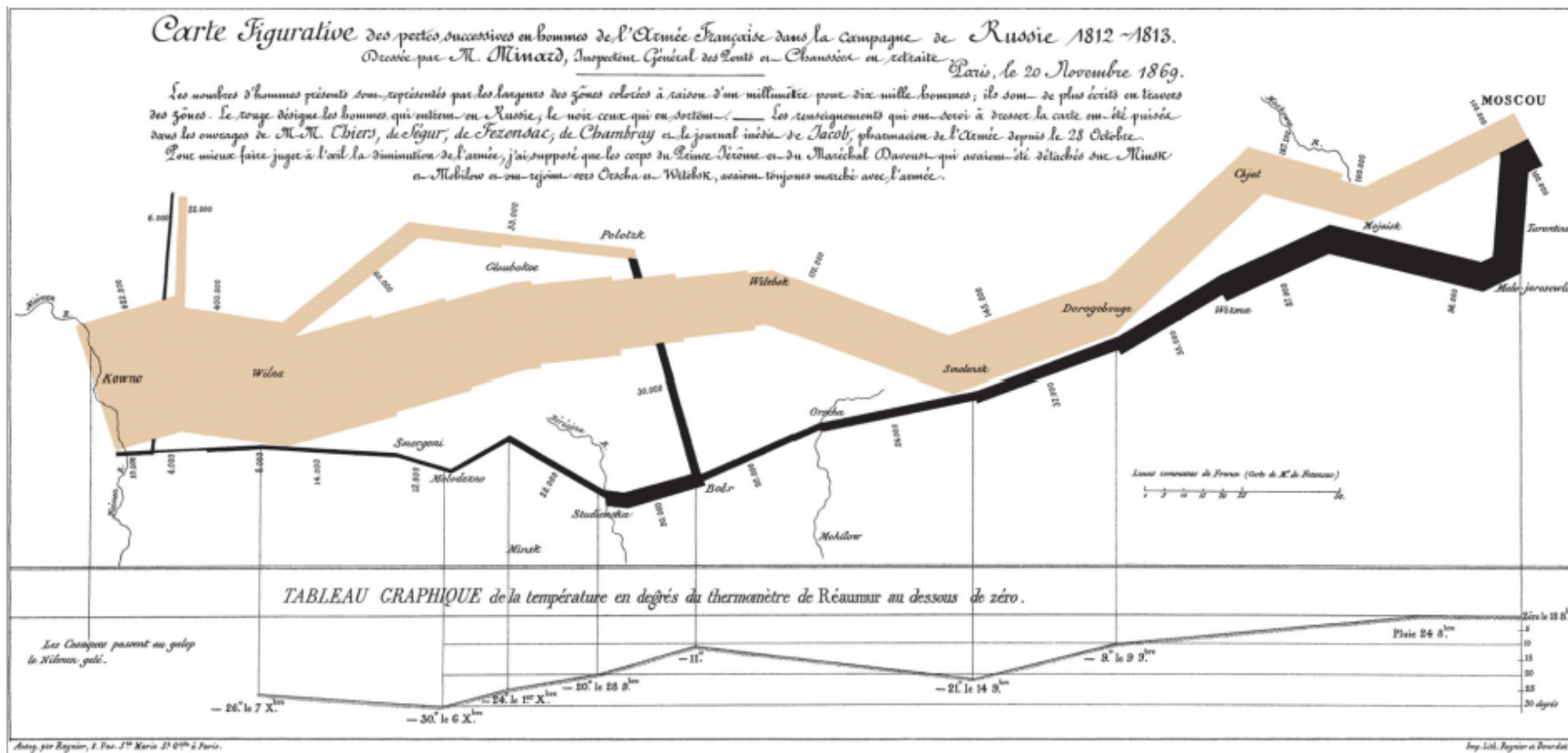


invente trois types de conception graphique : la série statistique sous forme de courbes, le graphique à barres et le graphique à secteurs

Florence Nightingale (1858)

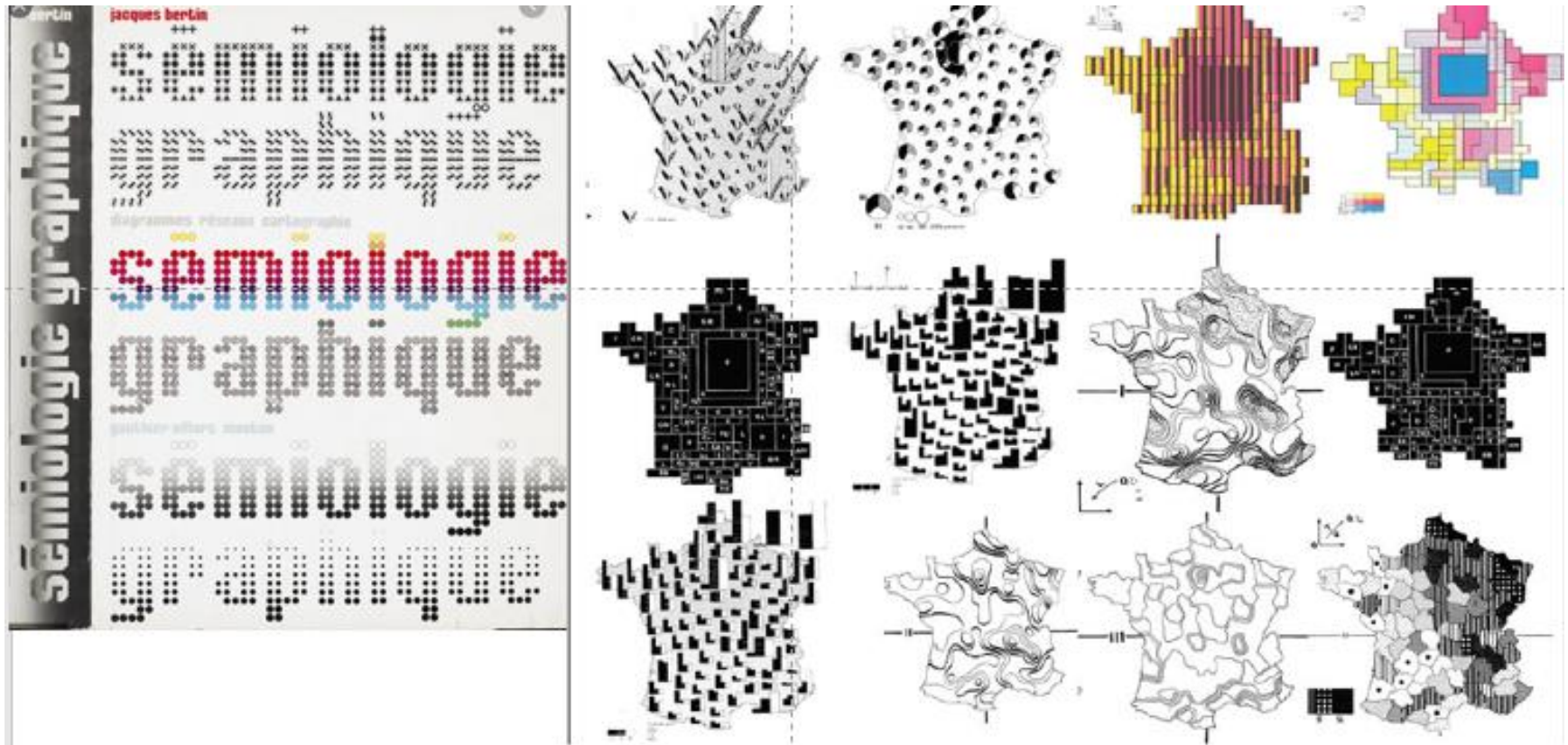


Charles Joseph Minard (1869)



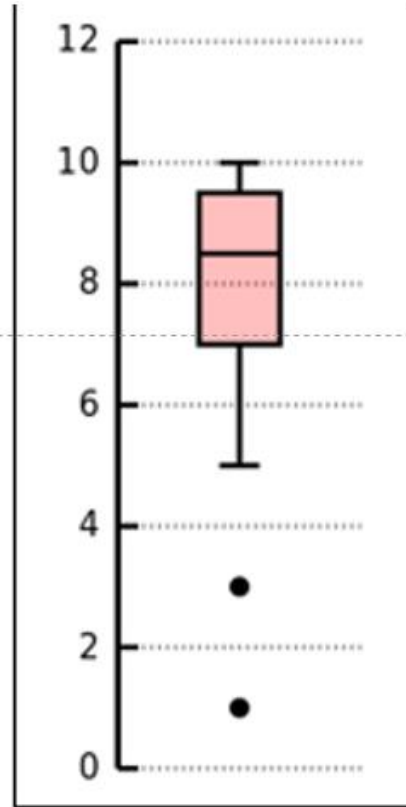
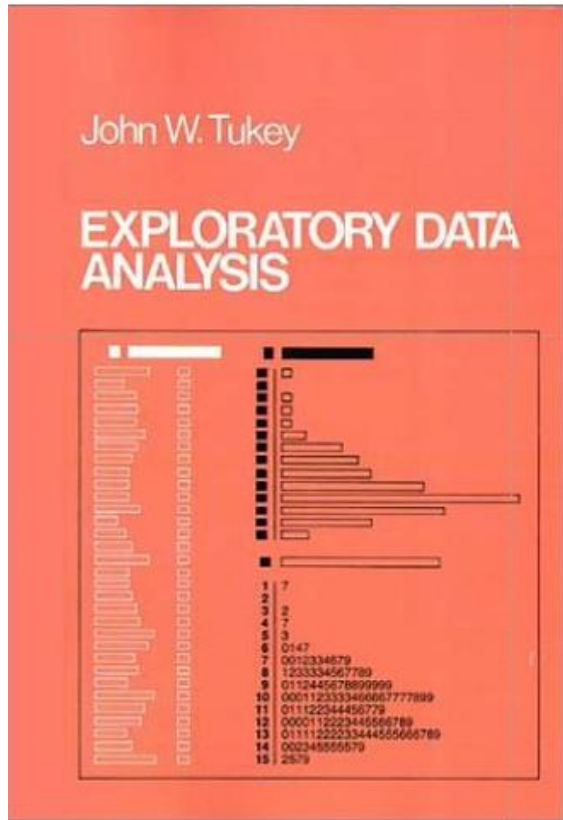
Armée de Napoléonienne, qui arrive à Moscou avec moins d'un quart de son effectif de départ, avant d'être décimée sur le voyage du retour

Jacques Bertin (1967)



Semiologie graphique : **ensemble de règles** permettant de transmettre une information correcte et d'aboutir à une image cartographique facilement **accessible** au lecteur

John Wilder Tukey (1970)



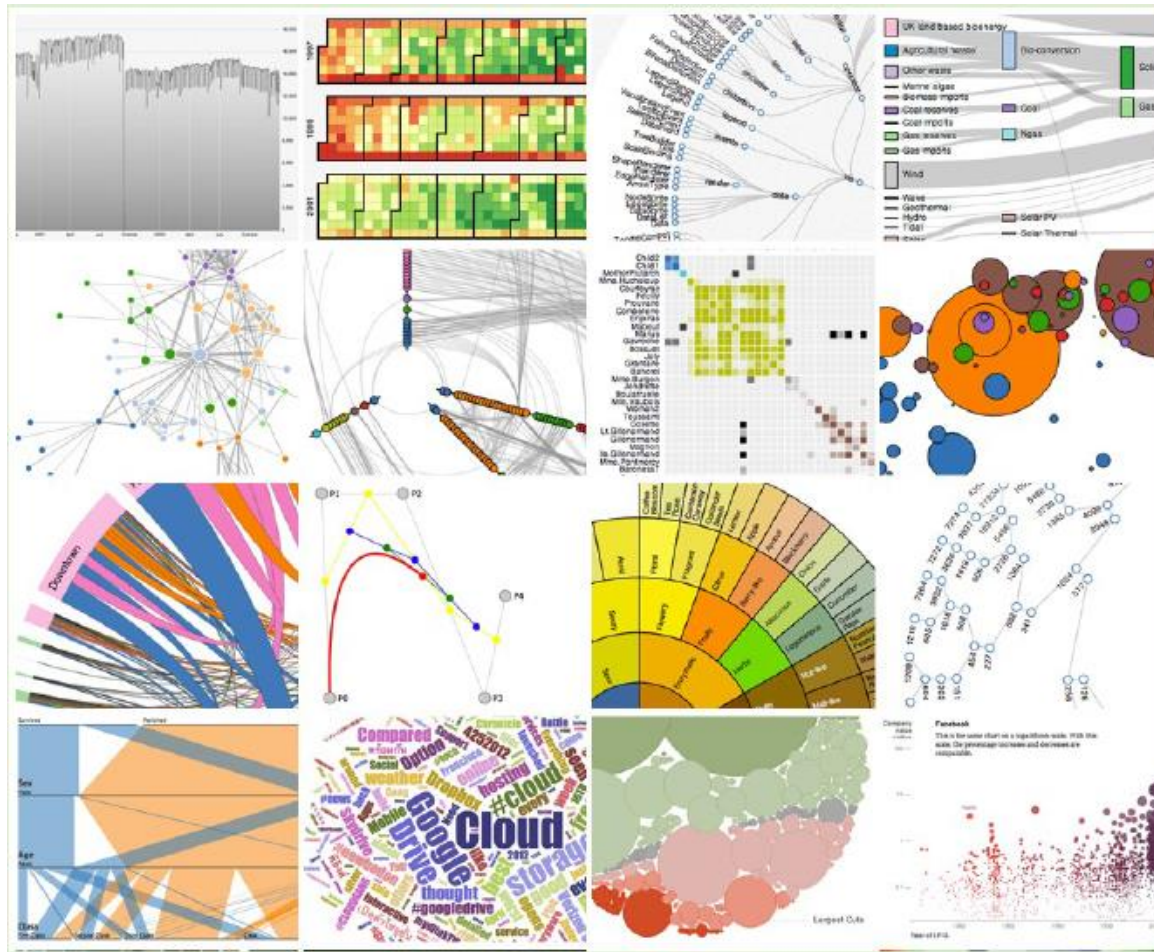
Invente les graphiques *Stem and Leaf, Box & Whiskers Plot* pour représenter schématiquement une distribution

We need both Exporatory & confirmatory statistics - J. Tukey (1980)

Edward. Tufte (1983) - The Visual Display of Quantitative Information

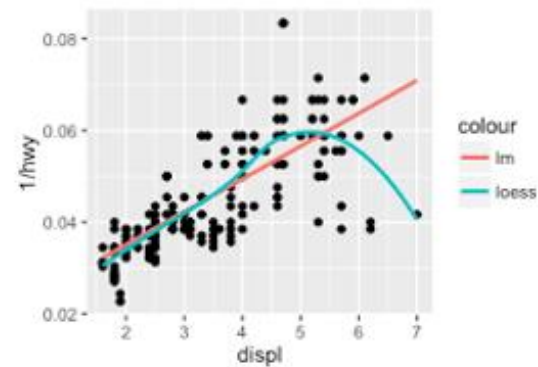
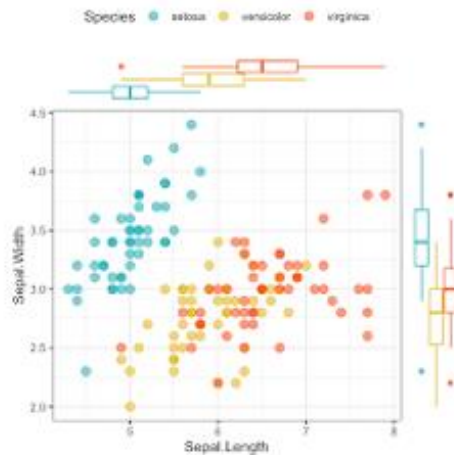
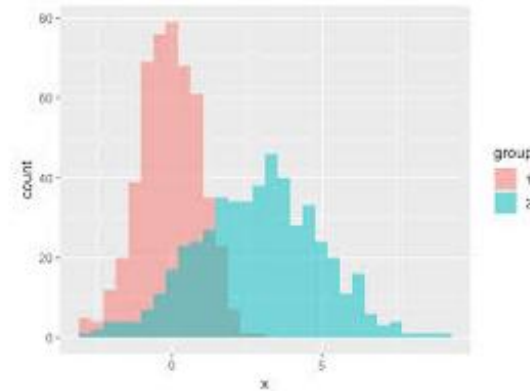
- Présenter les données ; Induire le lecteur à penser au contenu ;
- Éviter les distorsions ; Montrer beaucoup de choses dans peu d'espace ;
- Rendre des grand jeux de données cohérents ;
- Encourager la comparaison des différentes données ;
- Révéler les données à plusieurs niveaux (du plus large au plus fin) ;
- ...

Mike Bostock (2009) - librairie javascript *D3.js*



permet de visualiser des données sur le web de façon interactive et animée

Hadley Wickham - ggplot2



librairie open source sur R de visualisation de données

Les 3 Fonctions de la dataviz

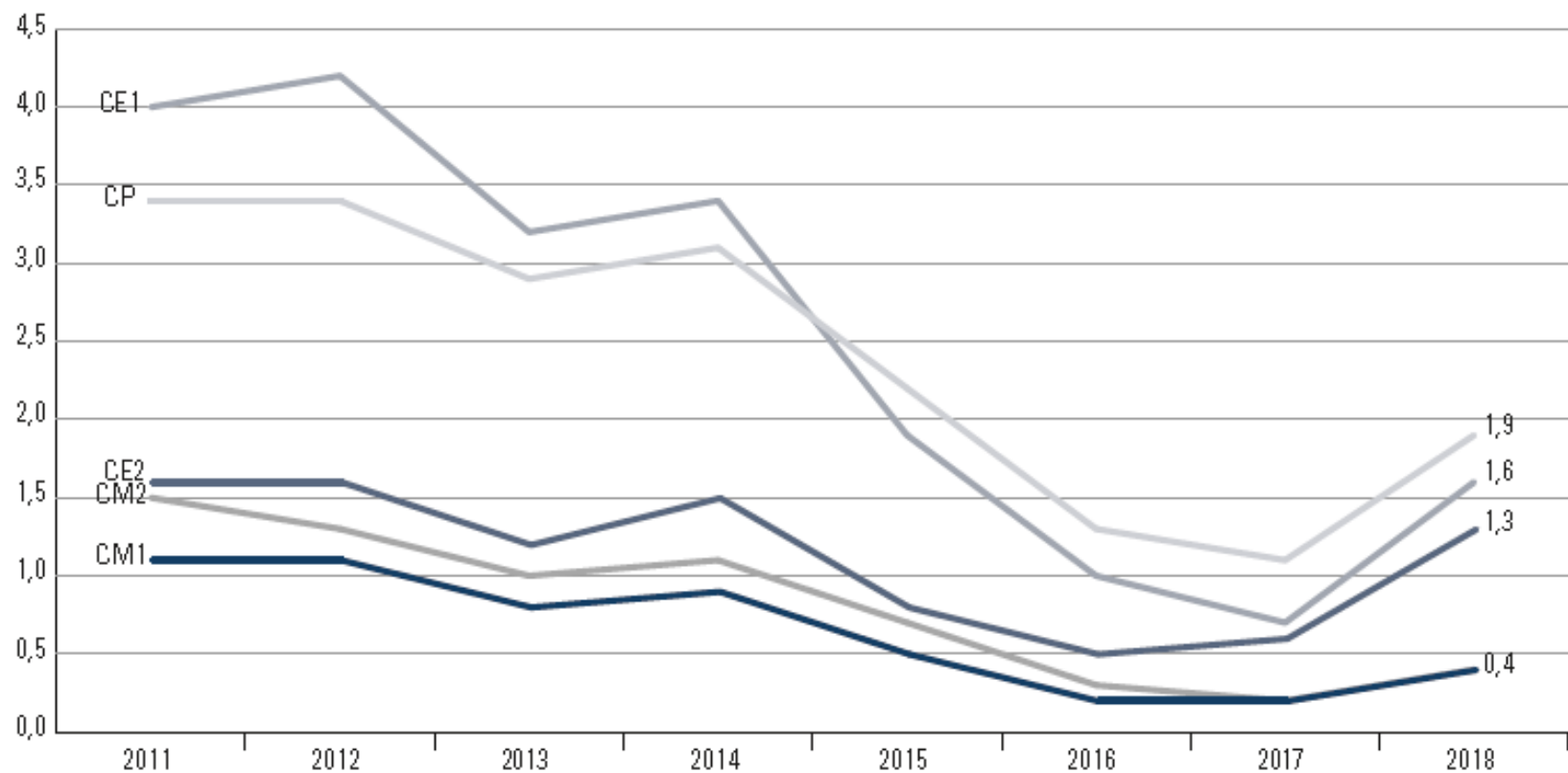
- **Explorer** : faire émerger des hypothèses à éprouver par d'autres moyens.

The greatest value of a picture is when it forces us to notice what we never expected to see (J. Tukey).

- **Confirmer** : utiliser le visuel comme un moyen de prouver - plus convaincant que les mots ?
- **Présenter** : faire usage des techniques de communication visuelle.

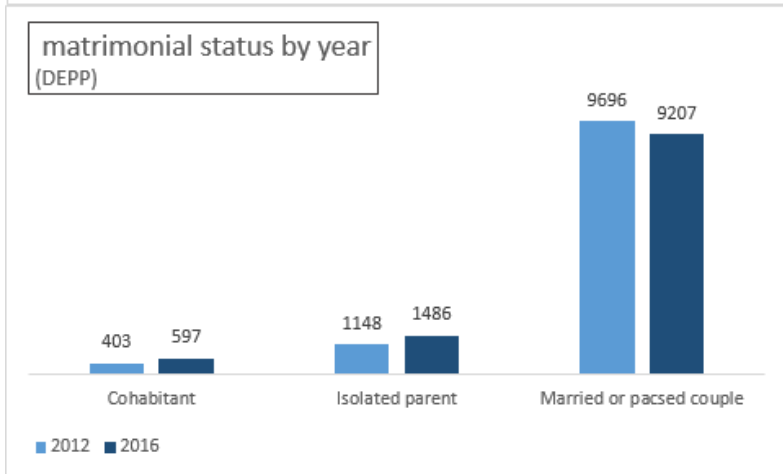
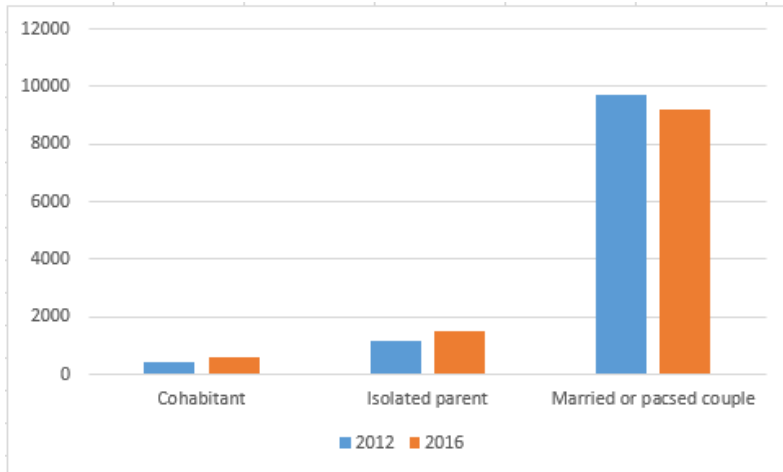
Usage de courbes pour confirmer

1 Évolution des taux de redoublement en élémentaire dans le secteur public

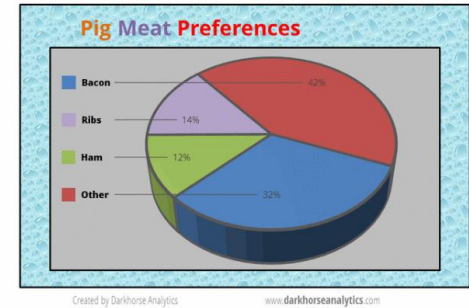


► Champ : France métropolitaine avant 1999; France métropolitaine + DOM (Mayotte à partir de 2014), Public + Privé.

Clarifier le message

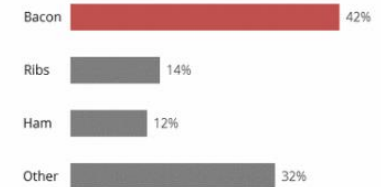


Before



Emphasize the bacon

Pig Meat Preferences



<https://www.darkhorseanalytics.com/blog/salvaging-the-pie>

Sémiologie graphique

Choisir le *graphique adapté*, ses *couleurs* et travailler son *habillage*. Nécesite au préalable d'identifier sa *finalité* (but/public/support).

La carte est un document graphique qui utilise le **langage visuel** composé de signes graphiques élémentaires : les **points**, les **traits** et les **taches**.

La manière dont varient ces signes par changement de forme de taille, de valeur et/ou de couleur est appelée **variable visuelle** (Bertin).

Les variations possibles des variables visuelles

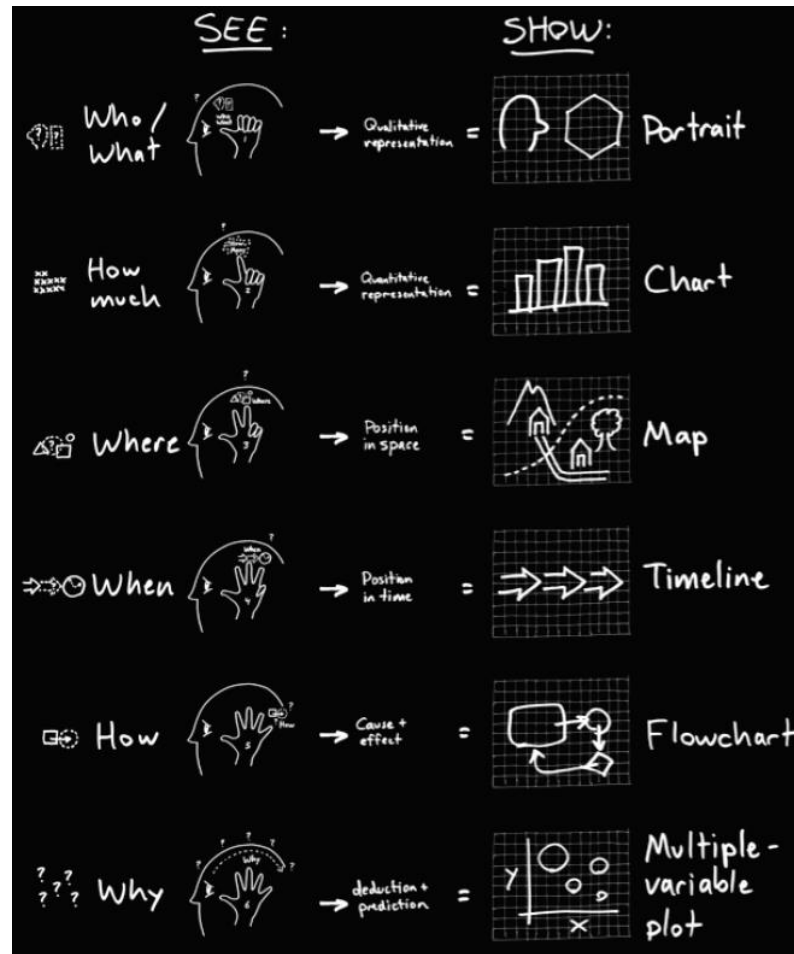
LES VARIABLES DE L'IMAGE													
		POINTS			LIGNES			ZONES					
XY 2 DIMENSIONS DU PLAN Z	TAILLE												
	VALEUR												
	LES VARIABLES DE SÉPARATION DES IMAGES												
	GRAIN												
	COULEUR												
	ORIENTATION												
	FORME												

La nature et le type de la mesure détermine le type de carte

Type d'implantation	Nature des données								
	Qualitative				Quantitative				
	Nominale		Ordinale		Relative		Absolue		
Ponctuelle	Forme	Couleur	Taille	Valeur	Valeur	Couleur	Texture	Taille	
Linéaire	Forme	Couleur	Taille	Valeur	Couleur	Valeur	Couleur	Taille	
Zonale	Couleur	Texture	Valeur	Couleur	Valeur	Couleur	Taille	Points comptables	
			Texture	Grain	Texture	Grain			

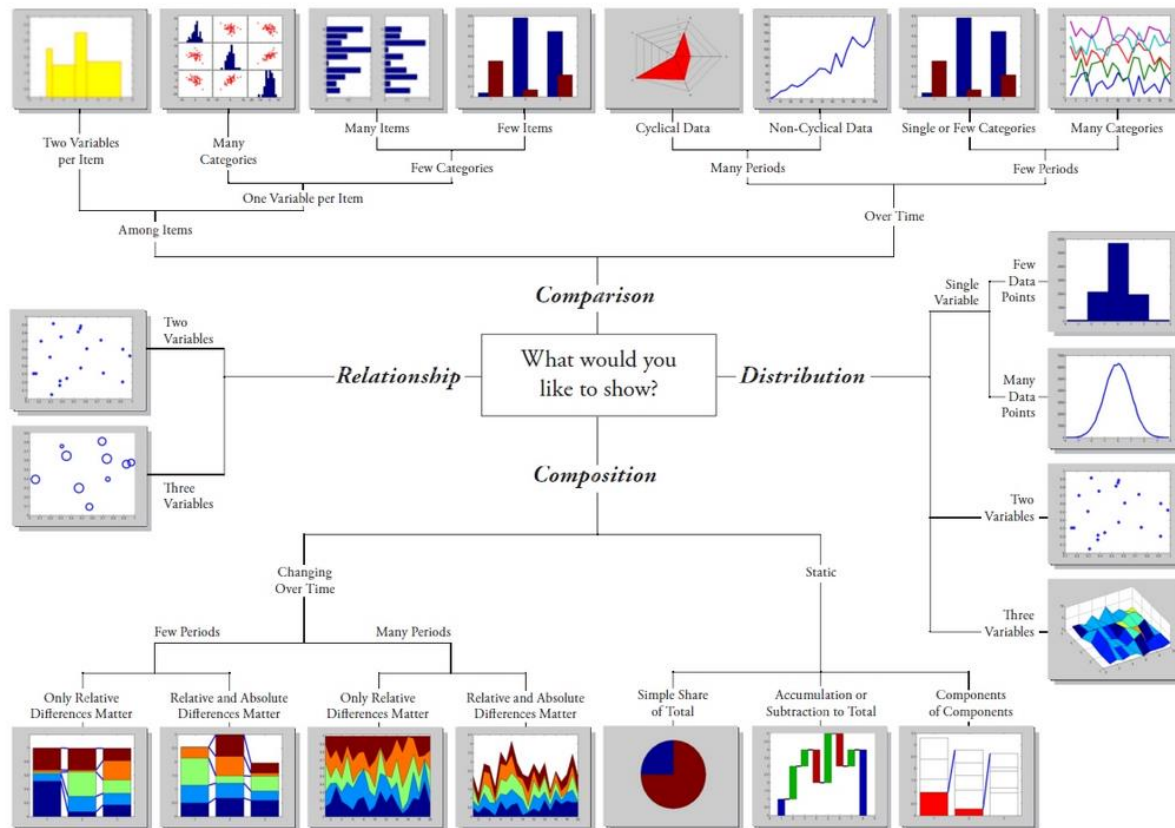
Source : C. Zanin et M.-L. Tremelo (2003), *Savoir faire une carte. Aide à la conception et à la réalisation d'une carte thématique*. Belin, Paris.

Utiliser le graphique adapté : *Visual Thinking "Codex"*



Les suggestions de graphique (A. Abela)

Chart Suggestions—A Thought-Starter



Modified with permission -Doug Hull
 blogs.mathworks.com/videos
 hull@mathworks.com 2009
 www.ExtremePresentation.com
 © 2009 A. Abela — a.abelag@gmail.com

Et le site interactif <https://www.data-to-viz.com/>

Choisir les couleurs

<https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>

- schémas séquentiels, divergents ou qualitatifs selon la nature des données
- les résultats doivent tenir compte de l'utilisation finale de la carte
- systèmes de spécification de couleur avec des nombres généralement écrits en hexadécimal et décimal

The screenshot displays the ColorBrewer 2.0 interface. On the left, the 'Number of data classes' is set to 3. Under 'Nature of your data', 'sequential' is selected. The 'Pick a color scheme' section shows the 'BuGn' sequential scheme. The 'Only show' section has 'colorblind safe', 'print friendly', and 'photocopy safe' checked. The 'Context' section has 'roads', 'cities', and 'borders' checked. The 'Background' section has 'solid color' selected. A color palette on the right shows three classes: #e5f5f9, #99d8c9, and #2ca25f. A tooltip for 'BuGn class 2' shows RGB: 153, 216, 201; CMYK: 40.0, 15.0; and HEX: #99d8c9. The map on the right shows France with these three colors applied to its regions. The bottom of the interface includes the copyright notice '© Cynthia Brewer, Mark Harrower and The Pennsylvania State University' and the 'axismaps' logo.

Travailler l'habillage

- Ajouter au graphique un titre clair, des intitulés lisibles, des étiquettes, une note de lecture ...
- En faire le maximum avec le logiciel qui génère le graphique *assurer la reproductibilité*.
- Sinon, l'exporter vers un logiciel de dessin au *format vectoriel* .svg ou .pdf pour pouvoir le retravailler.

Présenter le graphique

S'assurer de transmettre une information correcte mais aussi facilement accessible. Pour cela, l'image doit être *Nette, Economique, Séparative* et *Hiérarchique*.

Tester, tester et re tester le graphique ...

La séduction des images

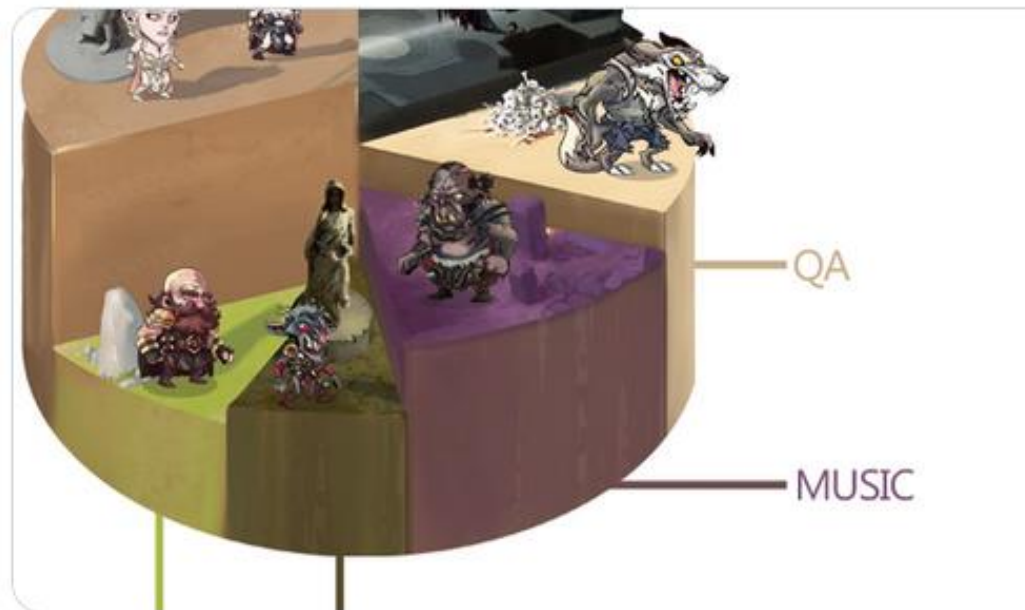
↳ Data Viz Fail a retweeté



Ω @Worthless_Bums · 4 sept. 2015

THIS IS THE WORST BUDGET PIE CHART IN EXISTENCE. WHAT.

WHAT IS THIS ABOMINATION.



5

5

8



En conclusion

- La dataviz = *outils* mais tout le monde n'est pas formé à la *sémiologie graphique*
- Son utilisation nécessite de la *rigueur* et de la *méthode* (attention aux logiciels)
- Il y a *des moyens* meilleurs que d'autres pour mettre en valeur des structures dans les données
- La sémiologie graphique est une source pour la recherche en représentation graphique
- La question des *métadonnées* (transparence, traçabilité, reproductibilité) est cruciale