

RefCard **Data Visualisation**



Comment bien représenter ses données ?

La data visualisation nous permet de créer un visuel qui transmet de manière efficace le sens d'une donnée. Ce visuel doit prendre en compte la donnée mais également la perception humaine. Parmi les 3 modes de visualisation (interactif, présentation, storytelling), celui que nous rencontrons le plus fréquemment est la présentation. C'est pour ce mode de visualisation que cette carte vous aidera à choisir le type de visuel le plus adapté à vos données afin de transmettre clairement votre message.

Déterminer le type de vos données

Vos données peuvent avoir différentes valeurs. Pour les représenter sous forme de graphique, ces données vont devenir des variables mathématiques. Il existe différents types de variables.

	Discrète ¹ (pas de valeur intermédiaire)	Continue ² (valeurs intermédiaires possibles)
Ordonnée (les valeurs sont comparables)	<ul style="list-style-type: none"> • ordinale³ (ex. taille : S, M, L, XL, ...) • quantitative⁴ (ex. compteur : 1, 2, 3, ...) 	<ul style="list-style-type: none"> • domaine⁵ (ex. altitude, température) • quantitative
Non ordonnée (les valeurs ne sont pas comparables)	<ul style="list-style-type: none"> • nominale⁶ (ex. forme : <input type="checkbox"/>, <input type="radio"/>, <input type="triangle-up"/>) • catégorie⁷ (ex. nationalité) 	<ul style="list-style-type: none"> • valeur cyclique⁸ (ex. directions, teinte)

¹ Une variable discrète représente un ensemble de valeurs que l'on peut énumérer

² Une variable continue représente une infinité des valeurs, formant un ensemble continu.

³ Une variable ordinale représente un ensemble de valeurs naturellement ordonné

⁴ Une variable quantitative représente un ensemble de valeurs permettant de quantifier la variable.

⁵ Un domaine représente un ensemble de valeurs appartenant à un domaine précis.

⁶ Une variable nominale représente un ensemble de valeurs correspondant à des noms.

⁷ Une catégorie représente un ensemble de valeurs appartenant à une catégorie que l'on peut nommer

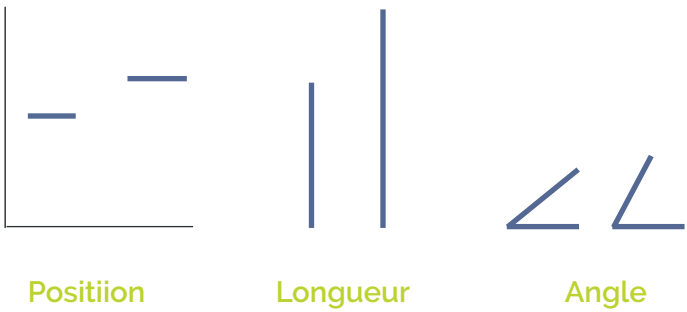
⁸ Une valeur cyclique représente un ensemble de valeurs fini et ordonné. Si nous parcourons l'ensemble de ces valeurs, nous parcourons à nouveau les mêmes valeurs et cela indéfiniment.

Selon le type de variable, la façon de la représenter n'aura pas le même impact. Voici le résultat d'études révélant l'efficacité de différents types de représentation pour chaque type de valeurs, triés par efficacité perceptuelle.

Représenter des variables quantitatives

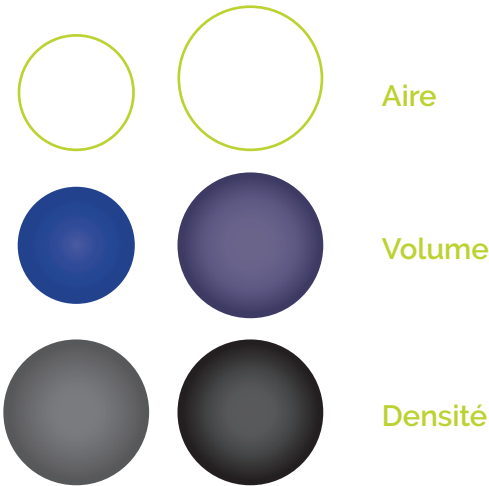
- 1. position
- 2. longueur
- 3. angle / inclinaison
- 4. aire
- 5. volume
- 6. couleur / densité / saturation / teinte

Cela signifie que lorsqu'une valeur est représentée par une position, nous comprenons mieux sa valeur quantitative que si elle avait été représentée par une couleur.



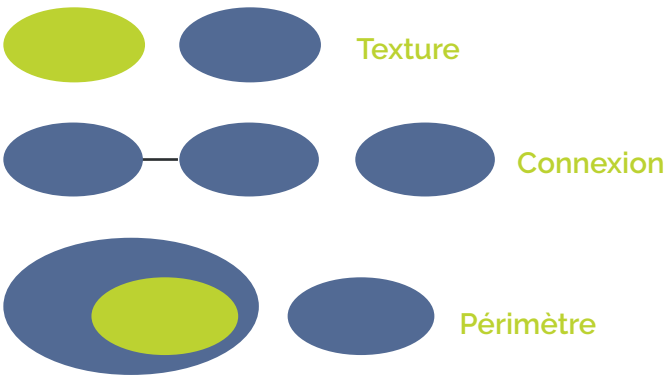
Représenter des variables ordinales

- 1. position
- 2. densité / saturation / teinte
- 3. texture
- 4. connexion
- 5. périmètre
- 6. longueur
- 7. angle / inclinaison
- 8. aire
- 9. volume



Représenter des variables nominales

- 1. position
- 2. teinte
- 3. texture
- 4. connexion
- 5. périmètre
- 6. densité
- 7. saturation
- 8. forme
- 9. longueur
- 10. angle / inclinaison
- 11. aire
- 12. volume

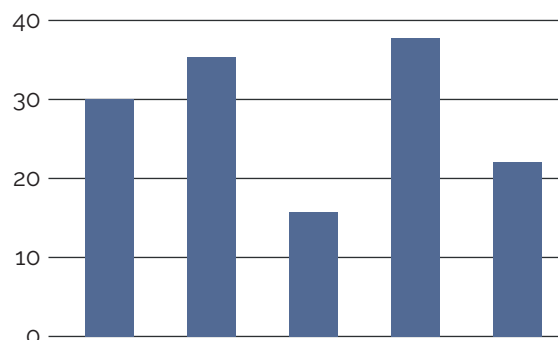


Représenter des données à l'aide de graphiques

Lorsque nous utilisons des graphiques, nous avons vu qu'il faut privilégier certaines formes géométriques plutôt que d'autres. Voici donc quelques graphiques et ce qu'ils mettent en avant.

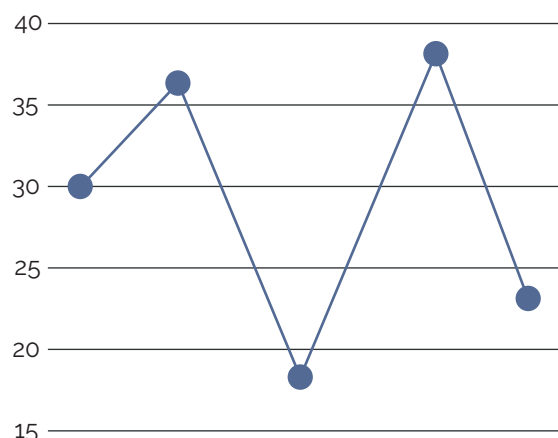
Histogramme

Ce type de graphique est généralement utilisé avec en abscisse des variables discrètes et indépendantes et en ordonnée des variables quantitatives. Visuellement, cela est bénéfique car nous affichons la variable avec une **position** (le haut de la barre) ainsi que par une **longueur** (la taille de la barre).



Lignes

Ce type de graphique est similaire au précédent car la hauteur des points correspond au sommet des colonnes. Nous bénéficions donc de la **position** mais pas de la longueur, ce qui nous permet de discerner les variables quantitatives. En revanche, en abscisse et en ordonnée, il est important d'avoir des variables quantitatives continues car nous relient les points par une ligne, ce qui implique une continuité de valeur entre 2 points. Il faut donc éviter d'utiliser un graphique de type lignes lorsque l'on veut afficher des variables discrètes, car cela donnerait l'impression qu'il existe des valeurs intermédiaires alors que ce n'est pas le cas.



Nuage de points

Si nous retirons les lignes du graphique précédent, nous obtenons un nuage de points. Celui-ci nous permet plus de flexibilité. Un graphique de type ligne affiche une fonction ce qui implique d'afficher une variable dépendante évoluant suivant une variable indépendante. Il existe donc une valeur dépendante pour chaque valeur indépendante. Le nuage de points permet d'avoir des variables quantitatives indépendantes en abscisse et en ordonnée indiquées par leur **position** horizontale et verticale. De plus, la **densité** serait mise en avant dans le cas où les points se regrouperaient en nuage.

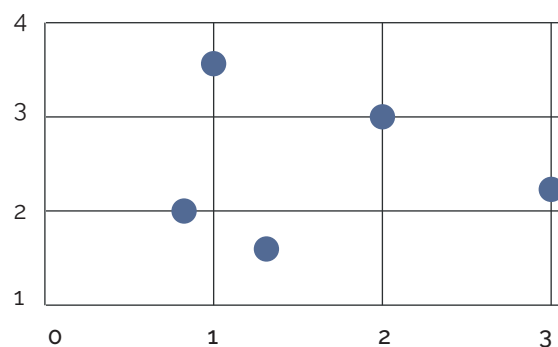


Diagramme de Gantt

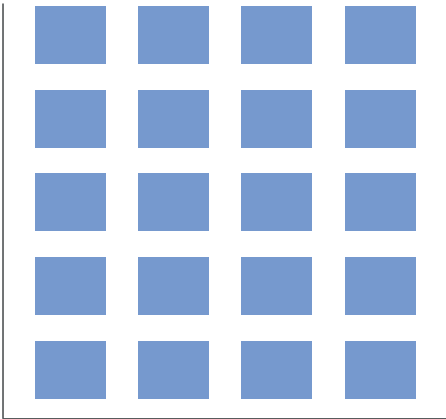
Le diagramme de Gantt ressemble à un histogramme latéral, sauf que les bases d'objets ne correspondent pas à l'un des axes comme dans un histogramme. Chaque axe représente une variable indépendante. Ce type de graphique est habituellement utilisé pour indiquer les étapes d'un projet dans le temps (en général sur l'axe des abscisses). En ordonnée, on retrouve une variable discrète non ordonnée. Souvent des tâches à effectuer. On peut ainsi montrer qu'une tâche ne peut démarrer avant une autre ou que deux tâches peuvent être réalisées en même temps. Ce diagramme comme l'histogramme bénéficie de la **position** et de la **longueur**.



Tableau

Le tableau est utilisé avec 2 variables discrètes ou nominales. Elles sont indépendantes, l'une ne dépend pas nécessairement de l'autre. La **position** est l'élément mis en avant dans ce type de graphique.

nb : si l'on regarde attentivement, on peut voir des points aux intersections. Il est important de prendre en compte la perception visuelle pour s'assurer de mettre en évidence uniquement ce qui nous intéresse.



En résumé

Voici un tableau qui vous aidera à déterminer le graphique à utiliser (parmi ceux présenté) selon les données sur lesquelles vous souhaitez communiquer.

En général, nous avons au moins une variable indépendante. Si vous avez une variable indépendante et une autre variable dépendante, dans la plupart des cas, vous utiliserez un histogramme.

Vous pouvez utiliser un graphique de type ligne uniquement dans le cas où vous avez une variable quantitative continue dépendante et une variable quantitative continue indépendante, car ce type de graphique sous-entend qu'il y a des valeurs intermédiaires horizontalement et verticalement.

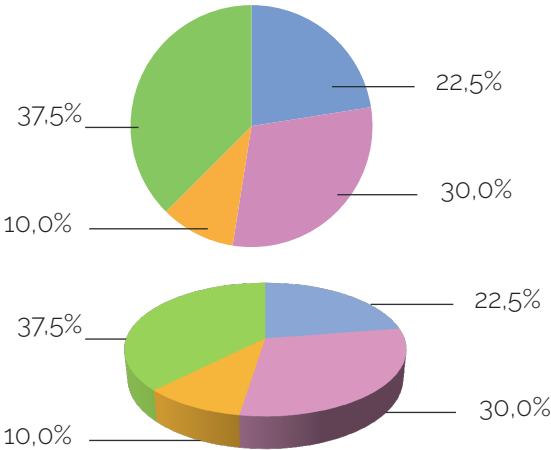
Dépendante	Quantitative continue	Histogramme	Ligne
	Quantitative discrète	Histogramme	Histogramme
Indépendante	Quantitative continue	Diagramme de Gantt	Nuage de points
	Nominale ou quantitative discrète	Tableau	Diagramme de Gantt
		Nominale ou quantitative discrète	Quantitative continue
Indépendante			

Prendre en compte la perception

Lors du choix d'une représentation graphique quelle qu'elle soit, il est important de garder à l'esprit la façon dont l'oeil perçoit les éléments.

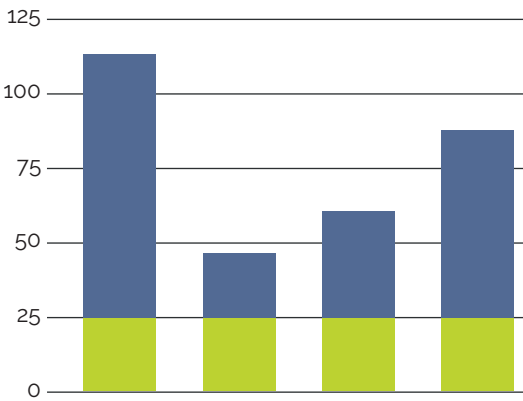
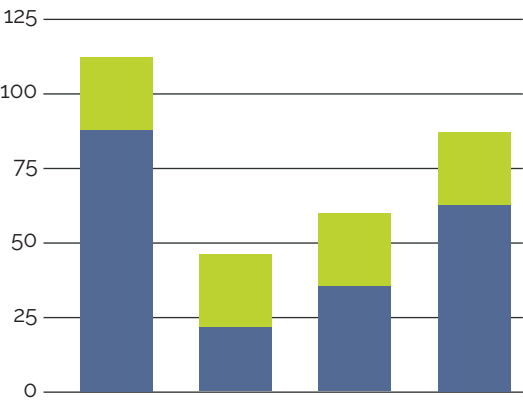
2 ou 3 dimensions ?

Le graphique en secteurs est utilisé pour représenter les proportions relatives d'une variable quantitative dépendante. Ces pourcentages sont reportés sous forme d'**angle**. Lorsqu'il est mis en 3 dimensions, la perspective rend les angles plus difficiles à lire.



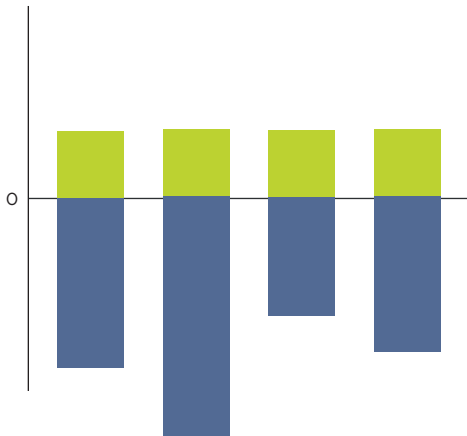
L'ordre a une importance

Il est plus facile de voir que la valeur en vert est constante sur l'histogramme empilé de droite que sur celui de gauche où il faudra faire un effort visuel pour s'en apercevoir.



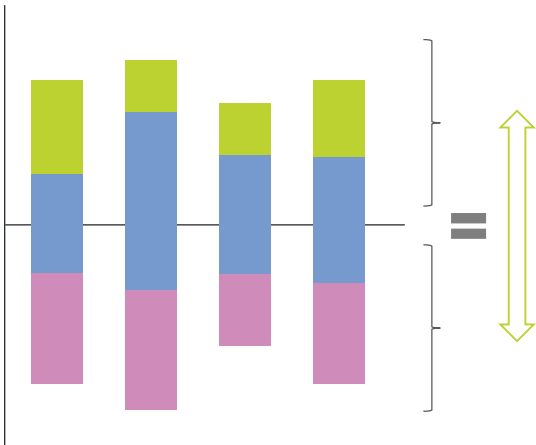
Faire une bonne première impression

Ici, nous bénéficions de la **position** et de la **longueur**, mais un histogramme empilé divergent ne peut être utilisé qu'avec deux variables. Et il est important de noter que la variable du dessous subit une connotation négative.



Faciliter la lecture

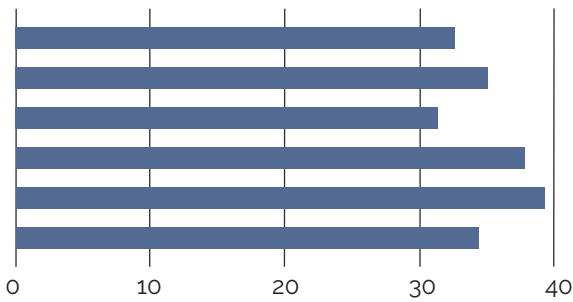
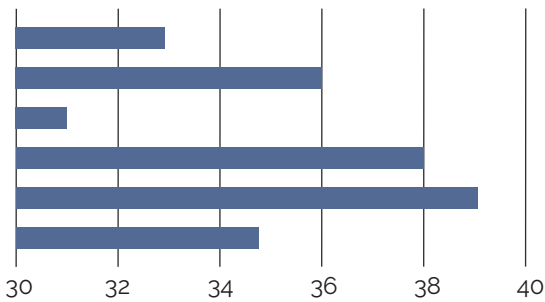
Ce graphique indique uniquement la **longueur** et fonctionne avec plusieurs variables. En revanche, la tendance est assez obscure à déterminer. Dans le même esprit, mélanger plusieurs type de graphes sur le même graphique peut dérouter. Essayez de rester simple.



L'utilité du zéro

Nous avons ici le même graphique. La différence est que sur celui de gauche, l'axe n'inclut pas la valeur zéro. On remarque alors un écart très net entre la troisième valeur affichée et la cinquième. Cet écart paraît beaucoup moins important lorsque le zéro est pris en compte. Ce type de graphique comme l'histogramme utilise la **longueur** et la **position** pour représenter les données.

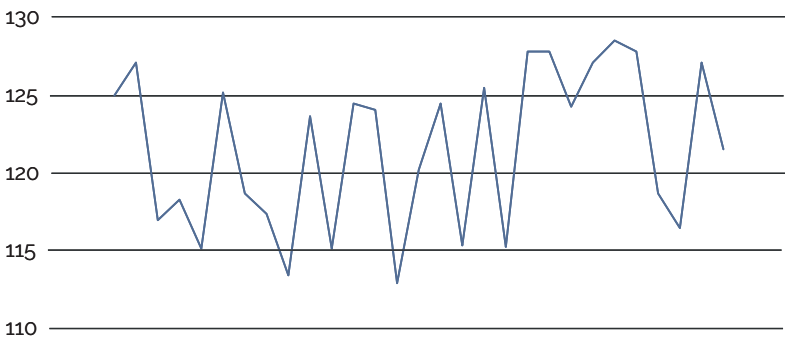
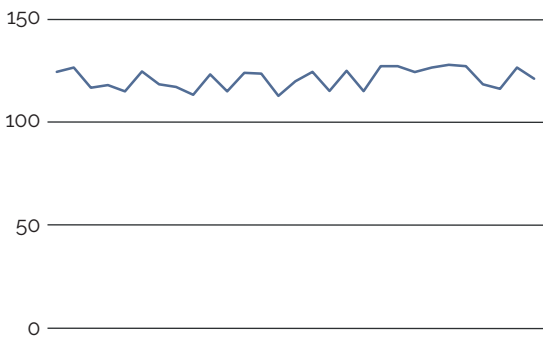
Pour que ces dernières soient interprétées de la manière la plus exacte possible, il faudrait dans la majorité des cas inclure le zéro. Ainsi la longueur de chaque barre serait proportionnelle à la valeur affichée.



Et l'importance de l'exclure

Ici encore nous avons les mêmes données représentées. Cette fois, le graphique de gauche prend en compte le zéro et celui de droite l'exclut. Un graphique de type ligne n'a pas besoin d'afficher la valeur zéro dans la plupart des cas,

car il se base sur la **position** et met en avant les changements de valeurs de la variable, ce qui est plus facilement lisible sur le graphique de droite.



Conclusion

Lorsque vous utilisez la data visualisation pour faire parler vos données, choisir le bon support n'est parfois qu'une question de détails. Mais ce sont ces détails qui feront la différence et rendront vos graphes plus percutants.



Cécile HUI-BON-HOA | Ingénieure Full Stack JS - SOAT

Débutant ma carrière en tant qu'ingénieure back-end, j'ai ensuite évolué vers un profil fullstack me permettant de comprendre et maîtriser la donnée à chaque étape.

Manipuler au quotidien une grande volumétrie de données m'a encouragé à m'appuyer sur la data visualisation pour créer des vues de données adaptées. Que ce soit un storytelling pour présenter, un support graphique pour convaincre ou un outil sur-mesure pour gérer les informations, la data visualisation est désormais un métier incontournable du Big Data.



www.soat.fr - blog.soat.fr

Sequana 1
89, quai Panhard et Levassor 75013 PARIS
contact@soat.fr | 01.44.75.42.55