

**— MANUEL DE —**


# **DATA**

## **VISUALISATION**

---

L'éditeur remercie Karen Bastien, Blandine Berthe, Denis Delbecq, Luigi Digirolamo, Jules Grandin, Nicols Guillerat, Romain Imbach, Gurvan Kristanadaja et Patrice Mitrano pour leurs contributions à cet ouvrage.

Conception de la couverture et de la maquette intérieure :  
Maud Warg avec la contribution de Jean-Marie Lagnel

<p>Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.</p> <p>Le Code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements</p>	 <p><b>DANGER</b> LE PHOTOCOPIAGE TUE LE LIVRE</p>	<p>d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).</p>
--	---	---

© Dunod, 2017, 2021

11 rue Paul Bert, 92240 Malakoff  
www.dunod.com

ISBN : 978-2-10-080902-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2<sup>o</sup> et 3<sup>o</sup> a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

— MANUEL DE —

# DATA

## VISUALISATION

---

**JEAN-MARIE  
LAGNEL**



**DUNOD**

# SOMMAIRE

## 6 ● AVANT-PROPOS

## A/ PRÉREQUIS

### ● CHAPITRE 1

#### 10 : QU'EST-CE QUE LA DATAVISUALISATION ?

- 11 : 1. Présentation
- 15 : 2. Les différents types de datavisualisation
- 18 : 3. Spécificités : papier, applications, sites web

### ● CHAPITRE 2

#### 26 : DE LA CONCEPTION À LA LIVRAISON

- 27 : 1. Les différentes étapes
- 32 : 2. La veille et les outils

## B/ CONCEPTION

### ● CHAPITRE 3

#### 38 : BRIEF ET ORGANISATION DES ÉLÉMENTS

- 39 : 1. Un brief pour bien cerner la commande
- 41 : 2. Identifier et sélectionner les données
- 45 : 3. Quels formats choisir ?
- 50 : 4. Analyse du brief

### ● CHAPITRE 4

#### 52 : CONCEPTION D'UNE DATAVISUALISATION

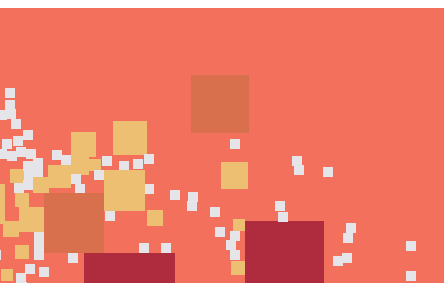
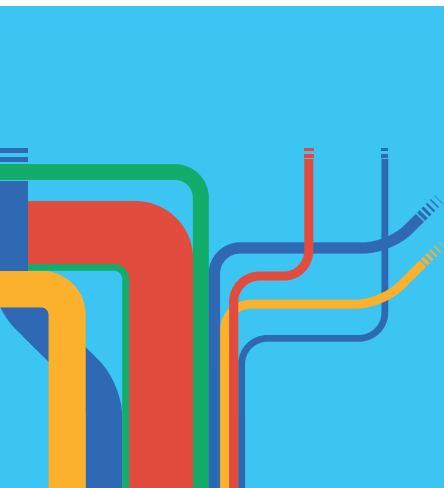
- 53 : 1. Récolter, analyser et organiser les données
- 58 : 2. Contextualiser l'information
- 62 : 3. Comment éviter de perdre le lecteur ?

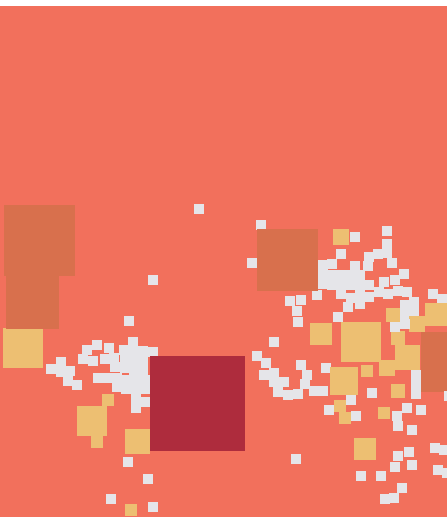
## C/ MISE EN FORME

### ● CHAPITRE 5

#### 70 : LES GRANDES FAMILLES DE REPRÉSENTATIONS GRAPHIQUES DES DONNÉES

- 71 : 1. Classement par catégories graphiques
- 94 : 2. Classement par applications





119 ● **CHAPITRE 6**  
**ENVIRONNEMENT : COMMENT HABILLER UNE DATA ET FAIRE DE BONS CHOIX GRAPHIQUES**

- 120 1. Choix des couleurs employées  
 125 2. Choix des polices employées  
 125 3. Habillage graphique et iconographie  
 141 4. Simplifier un graphique pour optimiser la lecture

146 ● **CHAPITRE 7**  
**RESSOURCES**

- 147 1. Banques d'images vectorielles : les pictogrammes  
 148 2. Plug-ins pour optimiser la production  
 154 3. L'editing

## D/ MISE EN ŒUVRE ET RÉALISATION

158 ● **CHAPITRE 8**  
**DATAVISUALISATION STATIQUE OU ANIMÉE**

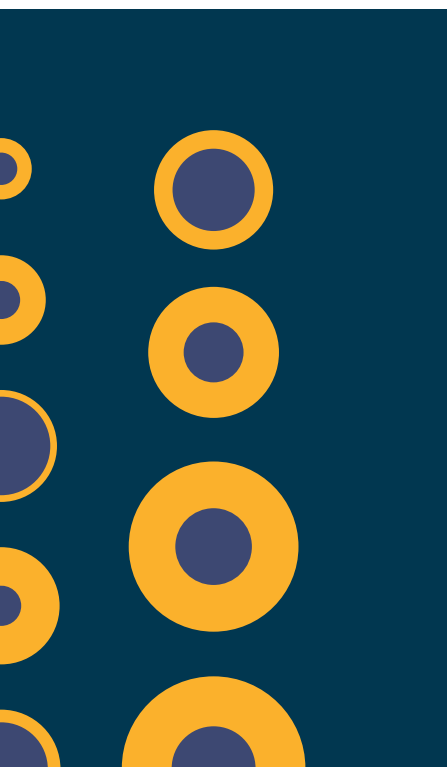
- 159 1. Datavisualisation statique pour le print  
 165 2. Les outils pour créer une datavisualisation animée  
 182 3. Les bases de données

190 ● **CHAPITRE 9**  
**CRÉER UNE ROSE DE NIGHTINGALE AVEC ILLUSTRATOR**

- 191 1. Première méthode  
 194 2. Seconde méthode

197 ● **CHAPITRE 10**  
**CINQ PROJETS À LA LOUPE**

- 214 CONCLUSION  
 215 POUR ALLER PLUS LOIN  
 217 TWITTOGRAPHIE  
 219 INDEX  
 222 CRÉDITS ICONOGRAPHIQUES



# AVANT-PROPOS

Cette nouvelle édition a été actualisée et enrichie de nouveaux exemples, projets et études de cas. Ce manuel est destiné à toutes les personnes souhaitant s'initier à la datavisualisation : étudiants, salariés au sein de services de communication ou manipulant des statistiques, graphistes indépendants ou simplement curieux des possibilités de mise en forme des données. Les lecteurs ayant déjà une expérience et une pratique professionnelle de la datavisualisation n'apprendront peut-être pas grand-chose ici. Ce manuel a pour but, avant tout, de faire découvrir le monde de la datavisualisation, d'en poser les fondamentaux, de donner un aperçu des possibilités et des outils d'aide à la conception. Pour guider le lecteur, nous nous appuyerons sur des exemples concrets et des études de cas d'acteurs reconnus dans le domaine. Nous essaierons de ne pas rentrer dans un détail trop technique, mais nous donnerons les clés, avec des pas-à-pas, pour guider le débutant afin qu'il se lance puis poursuive selon ses besoins son apprentissage. Un index en fin d'ouvrage permet de retrouver les occurrences significatives des notions clés dans le texte.

Différents encadrés viennent enrichir le texte : les exemples ancrent le propos dans la réalité et

livrent un avis professionnel à travers des anecdotes et des expériences vécues ; la méthodologie donne au lecteur des clés pour aborder certaines étapes d'une manière logique et efficace ; les nombreuses études de cas montrent différentes applications de datavisualisation à travers des projets réels. Enfin, les encadrés historiques présentent les précurseurs dans le domaine : en effet, l'histoire des données et de la statistique est jalonnée d'innovations graphiques. Parmi les principaux précurseurs de la datavisualisation, citons par exemple John Snow en Angleterre, Charles Joseph Minard en France, Otto Neurath en Autriche et Willard Cope Brinton aux États-Unis. La place nous manque dans cet ouvrage pour être exhaustif en citant tous les pionniers de la visualisation et détailler leurs travaux, mais nous présenterons les plus emblématiques. Le lecteur découvrira notamment que certaines formes de visualisations qui semblent aujourd'hui, lorsqu'on les découvre pour la première fois, nouvelles, originales ou radicales ne sont en fait souvent que des réappropriations de formes très anciennes, « simplement » adaptées au Web et rentrées dans une autre dimension grâce aux possibilités offertes par l'interactivité.

## 1 REMERCIEMENTS

L'auteur souhaite remercier les personnes lui ayant apporté avec leurs remarques, leurs expertises et leurs contributions, une aide précieuse tout au long de l'écriture de cette deuxième édition.

- Karen Bastien et François Prosper, cofondateurs de l'agence WeDoData. Karen Bastien a contribué à cet ouvrage en rédigeant les études

de cas « Notre (r)évolution carbone, 1990-2010 », « Profession journaliste » et « Sur la piste des œuvres antiques ». François Prosper a, quant à lui, apporté des remarques précieuses sur les textes de l'auteur.

- Blandine Berthe pour ses illustrations et l'étude de cas « Glyphosate : difficile de s'en passer ».

- Luigi Digirolamo, ancien chef du service infographie des *Échos*, pour la rédaction du cas pratique «Création d'une datavisualisation fixe pour *Les Échos*».

- Denis Delbecq, ancien chercheur et journaliste, ancien rédacteur en chef de l'émission *La Terre au carré* (France Inter) passé, entre autres, par *La Recherche*, *Sciences & Avenir*, *Le Monde*, *Libération*, *Heidi.news*, pour la rédaction du cas pratique «Workshop Visualiser la science» et de la partie de l'article «Short is beautiful» traitant des textes.

- Jules Grandin, chef du service infographie des *Échos*, pour le making-off sur la série d'été d'infographies «photoviz».

- Nicolas Guillerat, data designer, pour l'étude de cas issue de son beau livre *Infographies de la Seconde Guerre mondiale* (Perrin).

- Romain Imbach, chef adjoint du service infographie du *Monde*, pour l'étude de cas «Confinée, L'Île-de-France s'est engourdie».

- Patrice Mitrano, responsable du service cartographie de Sciences Po, pour son texte sur le Covid-19 présentant l'outil «Khartis».

- Gurvan Kristanadjaja, journaliste à *Libération*, pour sa contribution à l'ouvrage sur la partie «Les données» ainsi qu'à des dataviz citées.

- Clément Lefoll, journaliste pigiste, pour la traduction d'une partie des articles sur les données de

Quartz: «Les problèmes que vous devrez résoudre vous-même».

L'auteur souhaite également remercier Sylvain Lebrin pour sa contribution à la partie traitant des bases de données et du développement HTML. Et Cyril Courgeau, cartographe, pour ses remarques sur les parties traitant de la cartographie, ainsi que Duane Bekaert, développeur, pour son article sur «R». Merci également à Bénédicte Huchet pour sa relecture.

Enfin, l'auteur remercie les personnes et sociétés suivantes pour leurs autorisations de publications dans l'ouvrage: Alyson Hurt, Anne-Marie Drossart, Antanas Marcelionis, Ben Schneiderman, Bertrand Roubaud, Benoît Rouhier, Clara Alberto, Edward Tufte, Erik Laurijssen, Fabienne Mirbeau, Feilding Cage, Gregor Aisch, Hani Lim, Julius Tröger, Jacqueline Galante, Jan Willem Tulp, Jean-Louis Picot, Jean-Marie Le Guen, Johan Hufnagel, Johnny Christmas, Judith Barret-Chevrel, Jules Grandin, Laurent Bianco, Maryvonne Marconville, Mark Fischetti, Martin Grandjean, Mike McCormick, Nathalie Robert, Nicola Hellmann, Nigel Holmes, Oliver Uberti, Olivier Hodasava, Peter Orntoft, Rebecca Byers, Rebecca Dove, Renée Rigdon, Scott Adams, Sébastien Natroll, Sergio Peçanha, SimpleTherapy Team, Smäll Team, Sophie Coisne, Stanford Kay, The Noun Project Team et Tim Wallace.

## 2 À L'ÈRE DU BIG DATA

Avec l'avènement du Web, de l'*open data* et des objets connectés, une masse gigantesque de données sont devenues facilement accessibles. Rien qu'en utilisant son smartphone, chaque individu génère quotidiennement un grand nombre de données exploitables. Nous prenons des photos qui sont géolocalisées, nous partageons sur les réseaux sociaux, etc. Mais comment exploiter ces données brutes pour les comprendre et en tirer

des informations? C'est le rôle de la datavisualisation qui, en les représentant sous forme visuelle pour les rendre rapidement lisibles, nous permet de les explorer et de les interpréter.

Afin de faciliter l'accès aux datavisualisations interactives évoquées dans le texte, l'ouvrage est parsemé de QRcodes qui vous permettront de vous rendre rapidement et facilement aux ressources en ligne.





# PRÉREQUIS



1

# QU'EST-CE QUE LA DATAVISUALISATION ?



*«La plus grande valeur d'une image,  
c'est quand elle nous oblige à remarquer  
ce que nous ne nous attendions pas à voir.»*

JOHN TUKEY, *EXPLORATORY DATA ANALYSIS* (1977)

# 1 PRÉSENTATION

## DÉFINITION

Il n'y a pas encore de définition officielle de la datavisualisation, et chaque spécialiste y va de sa propre description. Nous pourrions la définir de la façon suivante: la datavisualisation sert à présenter rapidement les données et à les rendre lisibles. Elle permet d'analyser et de faciliter la compréhension des données en les retranscrivant efficacement sous forme visuelle.

La visualisation nous aide à comprendre les données plus rapidement en les mettant en images. Une datavisualisation réussie rend simple et assimilable un sujet complexe. On la génère automatiquement, à l'aide d'outils, de logiciels ou à partir de code HTML, et elle est généralement constituée d'une base de données relativement importante. Elle permet de créer une représentation rendant les valeurs lisibles graphiquement. Dans quel but? La datavisualisation fait « parler les données » et peut même raconter une histoire. Familièrement appelée « dataviz » par les professionnels, elle favorise la sérendipité et les facultés cognitives. Elle va rendre possible et faciliter l'exploration dans la profondeur des données et permettra de visualiser des tendances, des aberrations, des corrélations, etc. La visualisation ne se limite pas aux chiffres, elle peut être utilisée pour analyser des textes, des discours, faire ressortir des faits marquants en les contextualisant (tel discours a été prononcé à telle période, dans tel contexte). On utilise également les termes de visualisation de données ou de data design. L'évolution d'Internet, la profusion des datas, avec notamment l'*open data* qui se généralise, et l'interactivité ont porté la datavisualisation dans une nouvelle dimension. Le lecteur, qui par le passé était un simple spectateur passif, est devenu aujourd'hui grâce à la dataviz interactive un lecteur actif qui explore l'information comme il l'entend, en

faisant ses propres choix de consultation, dans un ordre qu'il définit lui-même selon ses centres d'intérêt.

Démonstration du principe de la dataviz avec l'exemple du quartet d'Anscombe (voir figure 1.1). Tandis qu'un tableau classique fait apparaître tous les chiffres des quatre groupes au même niveau, les rendant difficilement parlants, une simple représentation graphique en nuage de points permet de retranscrire les différences entre les données. Ce qui est invisible dans le tableau prend ainsi forme avec le graphique.

## DATAVISUALISATION OU INFOGRAPHIE ?

Depuis quelques années, la datavisualisation est à la mode et elle a pris tellement d'importance sur Internet et sur les supports papier que l'infographie a du mal à conserver sa place, et l'on fait parfois une confusion entre ces deux formes de représentation graphique. Un débat anime le Web sur le thème de la fin des infographies, qui seraient remplacées par des datavisualisations. Or, même si une quantité importante de projets bricolés, de qualité très médiocre, ont pu contribuer à nuire à l'image de l'infographie, les deux vecteurs d'informations sont complémentaires et, lorsqu'ils sont réalisés de manière professionnelle et à bon escient, ont chacun leur place et peuvent même se compléter et se combiner.

Qu'est-ce qui distingue une infographie d'une datavisualisation? Ce n'est pas la même chose: une infographie, c'est un graphique d'information conçu pas à pas à l'aide d'un logiciel comme Illustrator, qui raconte une histoire (voir figure 1.2). (Attention, il ne s'agit pas ici du terme générique d'infographie qui, dans d'autres secteurs, comme le jeu vidéo ou la modélisation 3D, fait référence simplement au graphisme conçu sur ordinateur.) L'infographie n'est pas



Figure 1.1. Quartet d'Anscombe.

une simple illustration, car elle se compose de texte et d'images. La définition de l'infographie pourrait donc être: ensemble d'informations mises en scène graphiquement. L'infographie n'est pas forcément composée de chiffres. Elle va présenter et exposer des informations de façon didactique (et parfois ludique) ou montrer un processus. Elle peut également être analytique ou descriptive, par exemple: un écorché technique détaillant une centrale nucléaire, un avion, une construction architecturale... Elle peut être composée uniquement de diagrammes ou, au contraire, principalement d'illustrations, mais comprendra toujours un titre, parfois une légende si nécessaire, et du texte. L'infographie peut être minimaliste, comme celles employées pour des journaux économiques, ou très fortement illustrée. Les deux approches, ainsi que les combinaisons, sont possibles: tout dépend du but et du type de public recherché. La science, le sport et l'économie se prêtent tout particulièrement au traitement infographique, notamment

dans la presse et les médias. Mais il n'y a pas de limite dans les applications et les sujets.

Une série de pictogrammes accompagnée de légendes ne constitue pas forcément une infographie. Dans bien des cas, il s'agit simplement d'une mise en forme de visuel avec du texte, d'une simple maquette réalisée sous InDesign ou Quark XPress. Pour que ce soit une infographie, il faut que la mise en scène des éléments apporte une autre dimension et délivre visuellement un message, une information. Par exemple, si la taille des pictos est représentée proportionnellement à leur valeur, on crée ainsi des niveaux de lecture et une hiérarchie. Ce qui permet de guider le regard et de visualiser les différences entre les catégories. On peut alors parler d'infographie.

Contrairement à l'infographie, une data-visualisation est le plus souvent générée automatiquement à l'aide d'un outil ou de code. La dataviz est souvent constituée d'une base de données. On représente graphiquement un jeu de données complexes afin de le rendre