

Data-journalisme

MMI 2 – TP#4 S4



Danielo JEAN-LOUIS

Datavisualisation - Objectifs

- Transmettre des données complexes via des graphiques et des légendes
 - Permet de prendre une décision rapide
- Appelé également "dataviz" ou "visualisation de données"

Source(s) :

- <https://fr.venngage.com/blog/qu-est-ce-qu-une-infographie/>

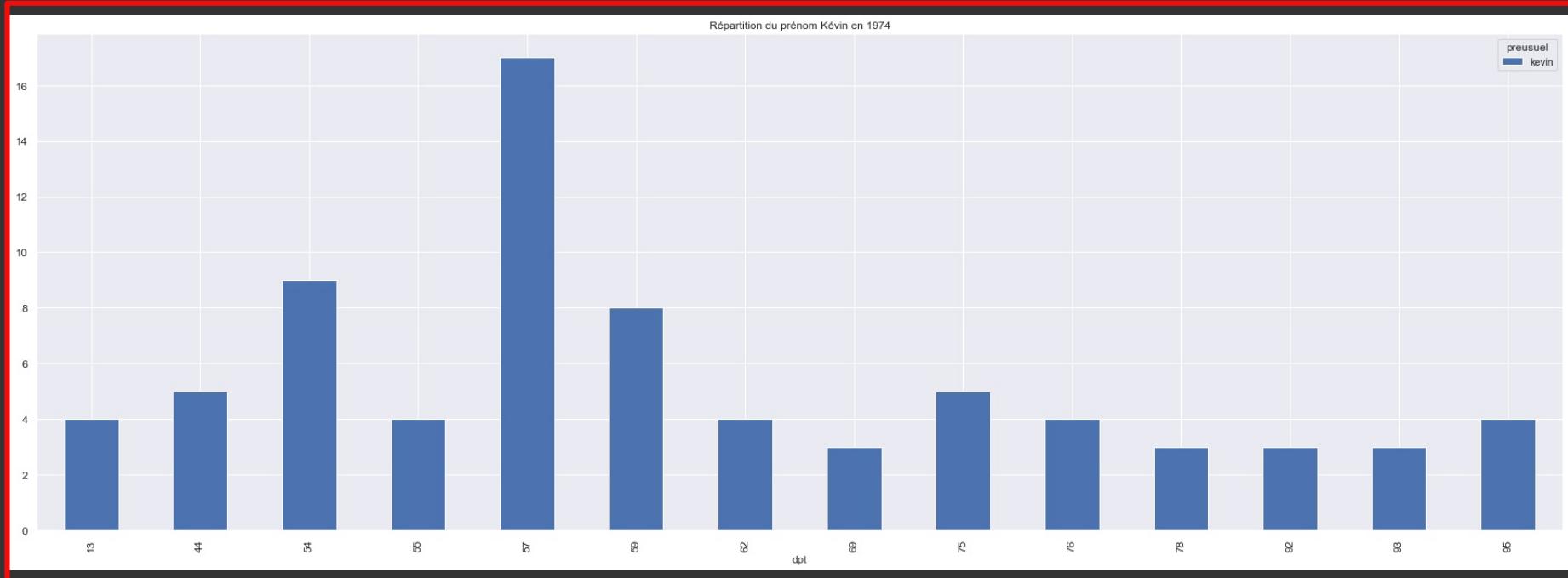
Datavisualisation - Objectifs

- Les gens lisent moins, il leur faut une accroche, une synthèse d'informations
- **Une infographie est un assemblage de dataviz, chiffres et illustrations**
- Peut créer une histoire (storytelling)

Source(s) :

- <https://fr.venngage.com/blog/qu-est-ce-qu-une-infographie/>

Datavisualisation - Objectifs



Datavisualisation (graphique + légendes)

Infographie

- A ne pas confondre avec la datavisualisation
- Ensemble de dataviz, illustrations et autres textes
- Sert à décrire un processus en le simplifiant

Source(s) :

- <https://fr.venngage.com/blog/qu-est-ce-qu-une-infographie/>

Infographie

- Peut raconter une histoire (story-telling)
 - Peut prendre un parti
- Ne pas oublier de mettre vos sources

Source(s) :

- <https://fr.vennngage.com/blog/qu-est-ce-qu-une-infographie/>

Infographie - Exemples

IPM INTEGRATED PROJECT MANAGEMENT
CHICAGO ST. LOUIS SAN FRANCISCO BOSTON

WANTS YOU

as a Project Management Consultant

BECAUSE YOU HAVE

- Leadership
- Creative problem solving
- Ability to quickly adapt to new environments
- Bachelor's or higher
- Knowledge and application of project management
- Excellent communication skills
- Experience in one or more of our industries

A LITTLE BIT ABOUT US

- Our PMCs work on-site to help companies on critical projects
- 50% TRAVEL COMMITMENT
- Proudly SERVING INDUSTRIES: Life Sciences, Food & Beverage, Consumer & Industrial Products, Healthcare
- OFFICES: San Francisco, BOSTON, ST. LOUIS, Chicago

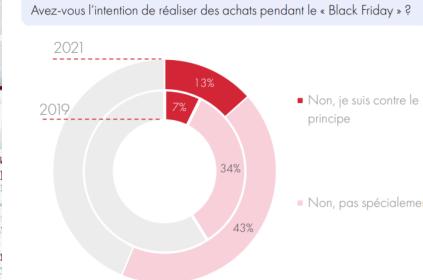
Survey says ...

Our staff was asked a series of questions about working at IPM. Here's what they said:

- I am proud to tell others I work here
- I feel good about the ways we contribute to the community
- This is a friendly place to work
- People here are given a lot of responsibility
- Management shows appreciation for good work and extra effort

Interested In Joining Our Team?

Learn more about Integrated Project Management and apply to become a Project Management Consultant on our website: www.ipmcinc.com



13 % des Français se disent opposés au principe du Black Friday.
C'est près de 2 fois plus qu'en 2019.

Let Them Eat CAKE

CAKE

A Guide to Wedding Cake Size and Servings

CAKE SIZES AND SERVINGS

Size	Servings (w/o Top)	Servings (w/ Top)
4"	20	14
6"	40	26
8"	46	40
10"	100	86
12"	210	196
14"		

CAKES

1939

RESIDENCE
Wayne Manor: \$3,800*
Batcave: \$5,600

CLOTHING
Suit: \$10

VEHICLES
Batmobile: \$600

2013

RESIDENCE
Wayne Manor: \$444,000
Batcave: \$54.4 million

CLOTHING
Kevlar body armor: \$300,000
Polysatin cape: \$3,000

VEHICLES
Batmobile: \$18 million
Batplane: \$60 million
Batcycle: \$15 million

ride.com
planningtip.com

Un excellent graphique est celui qui donne au spectateur le plus grand nombre d'idées avec le moins d'encre possible, dans le plus petit espace.

Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.

Edward R. Tufte, Professeur de statistiques à l'université de Yale

Diagrammes en bâtons (Bar chart)

- Très facile à lire
- Permet de comparer des données
- Existe aussi sous la forme horizontale, groupée ou encore groupée et empilée

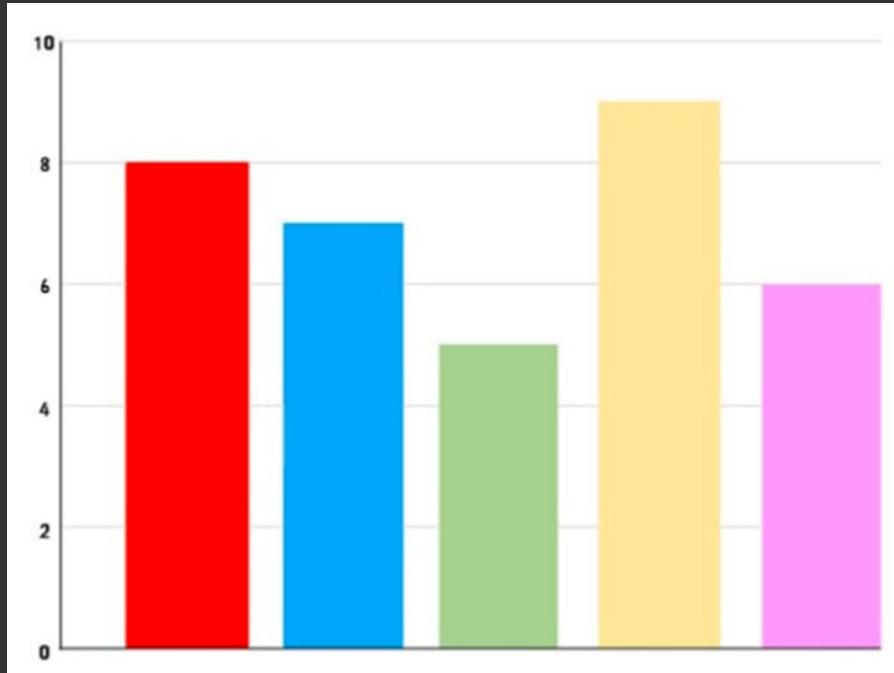
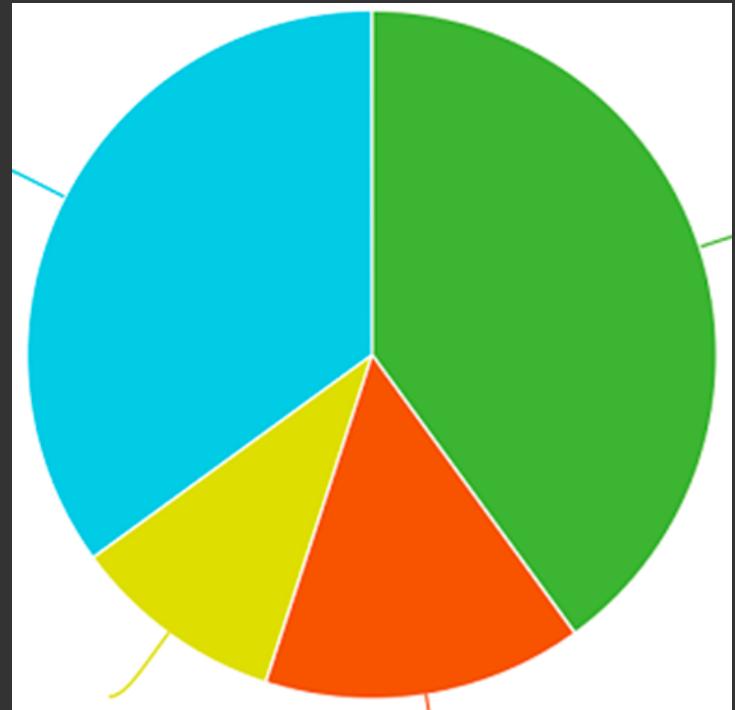


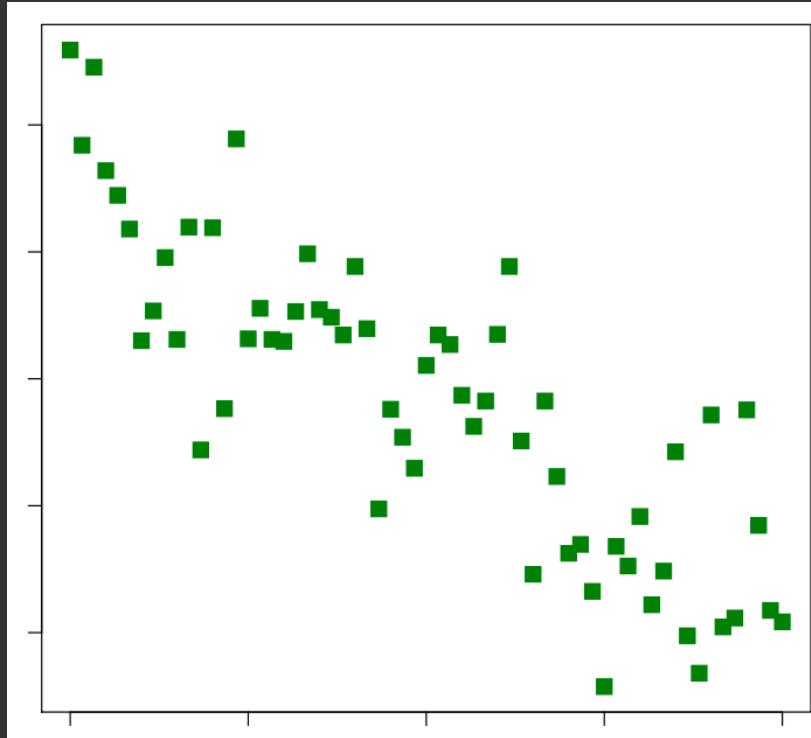
Diagramme circulaire (Pie chart)

- Appelé aussi "camembert"
- **Peut devenir très vite illisible**
- Limiter le nombre d'entrées
- Permet de représenter la composition/répartition de quelque chose
- Peut être représenté sous forme de bâtons



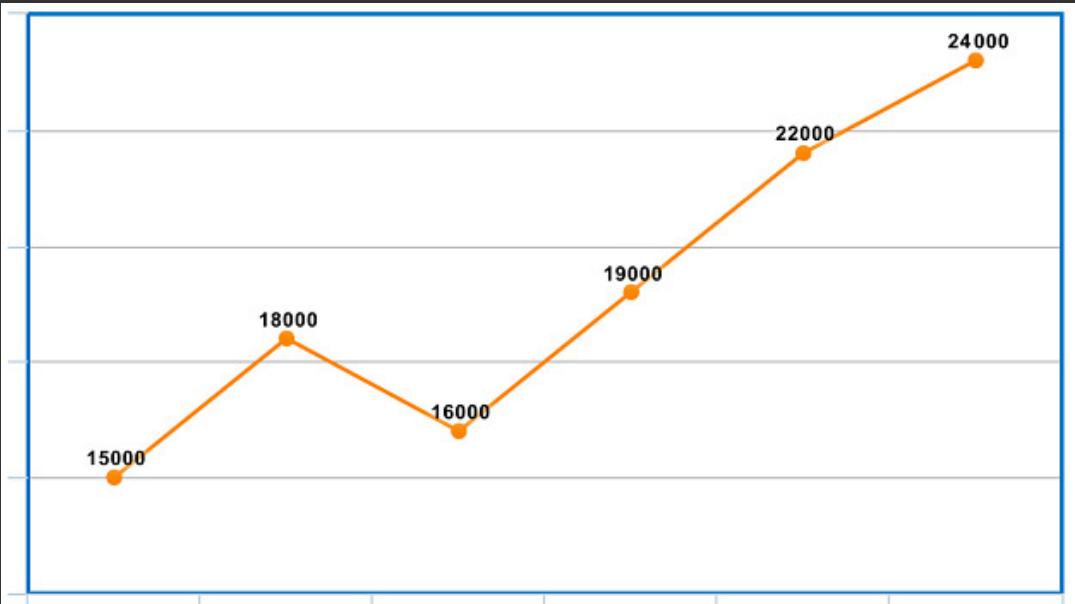
Nuage de points (scatter plot)

- Peut représenter une corrélation entre deux variables
- Les deux variables doivent être numériques
- Souvent utilisé avec une régression linéaire



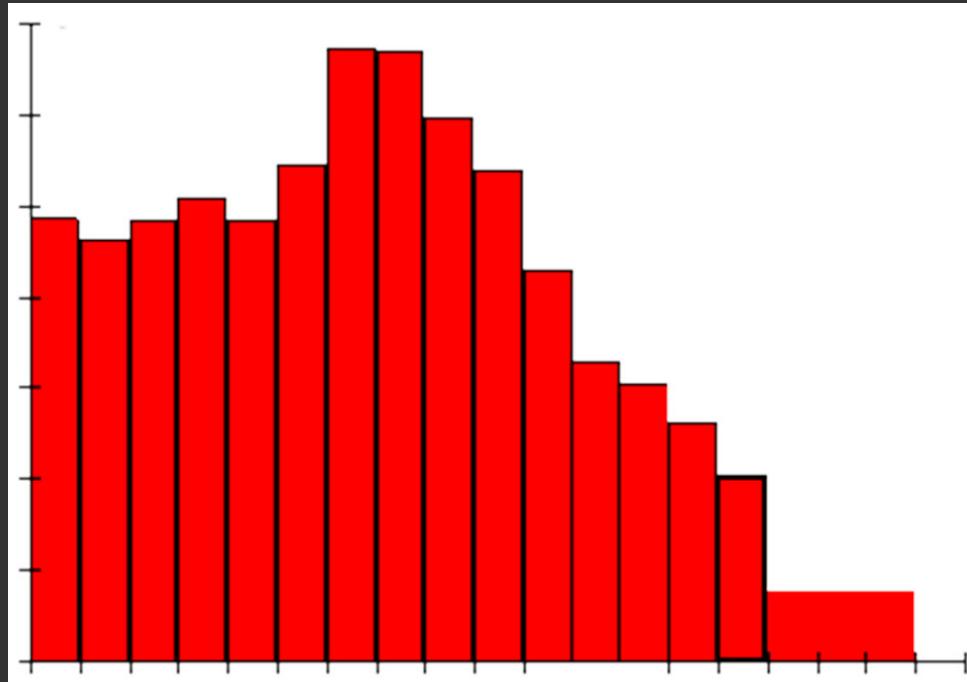
Courbe (line chart)

- Facile à lire
- Visualise l'évolution d'une catégorie
- Adapté aux données temporelles
- Possibilité d'avoir plusieurs courbes sur le même graphique

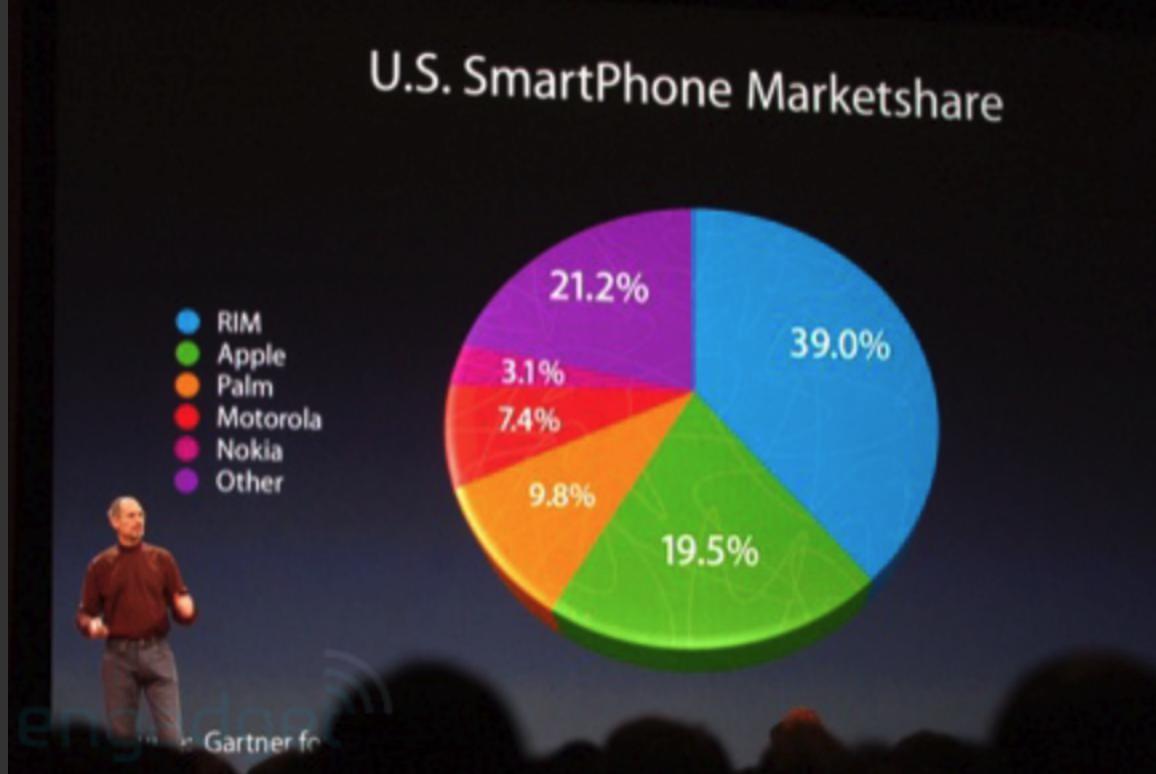


Histogramme (histogram)

- A ne pas confondre avec le diagramme en bâtons
- Montre la fréquence de distribution d'une valeur
- Les données doivent être continues



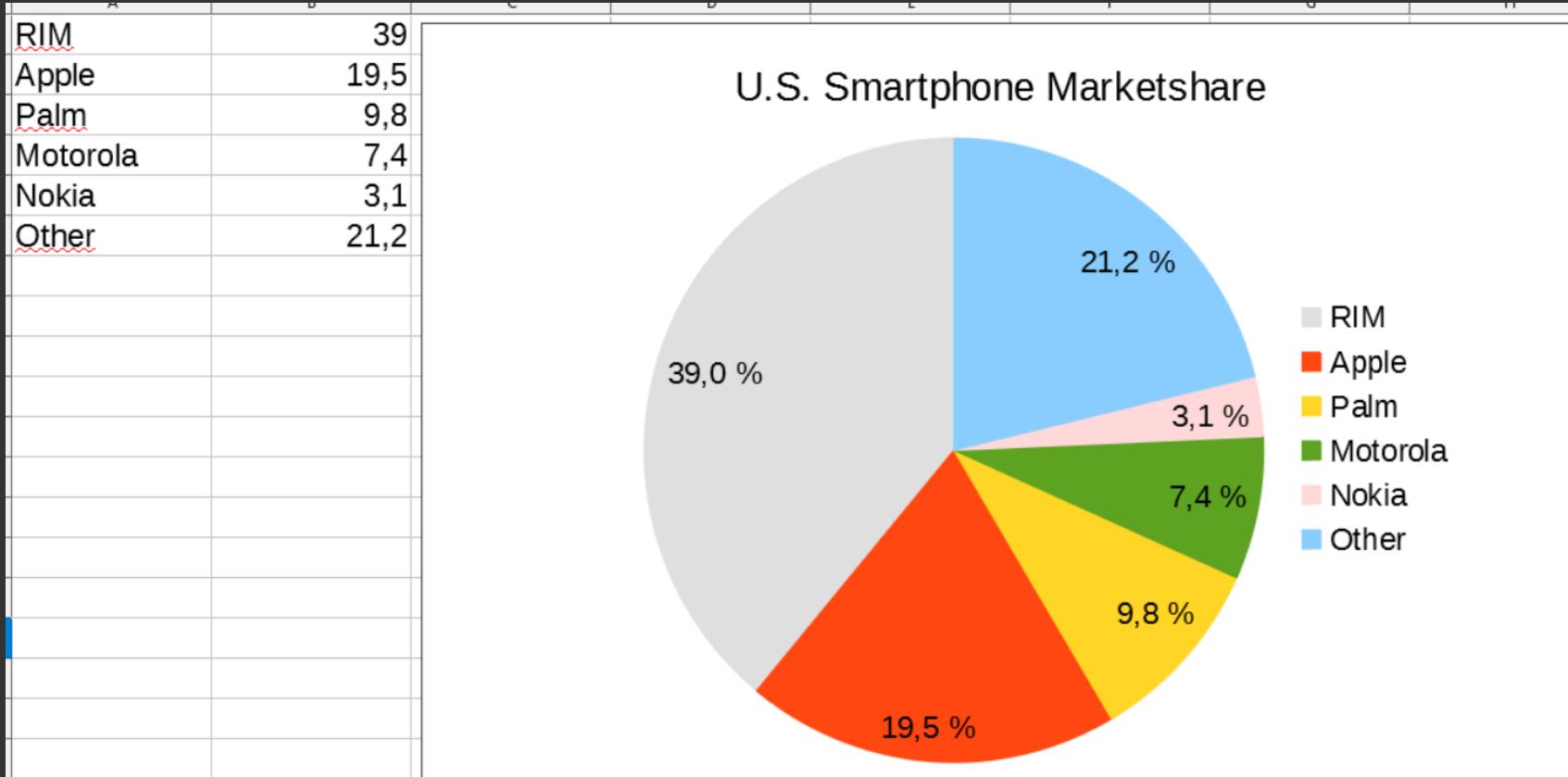
Datavisualisation – Attention à la manipulation



Keynote d'Apple le 9 juin 2008

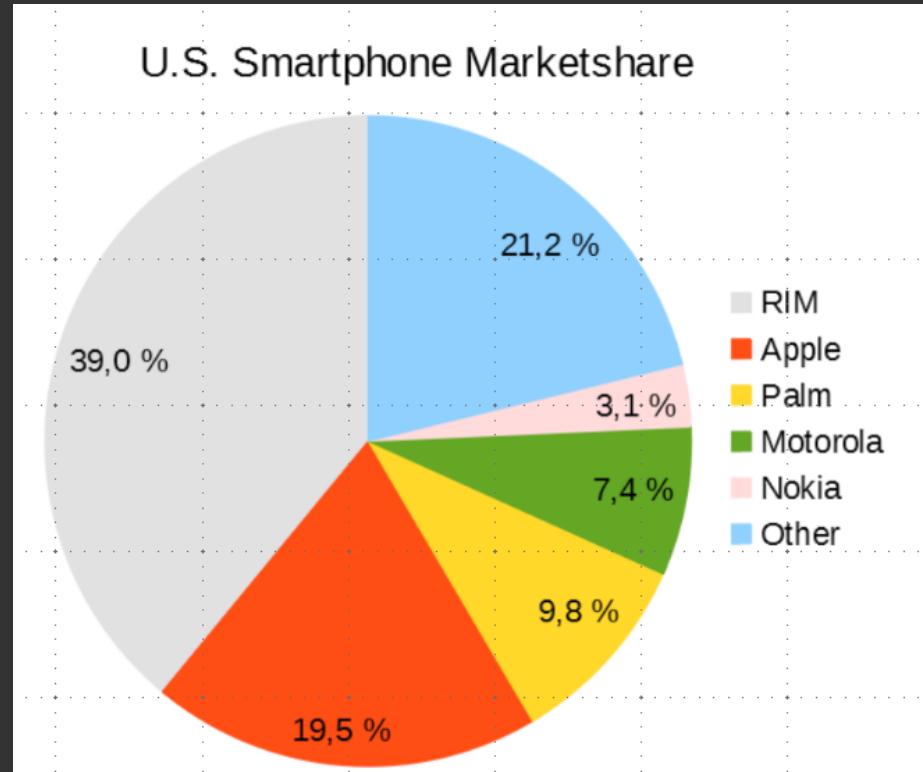
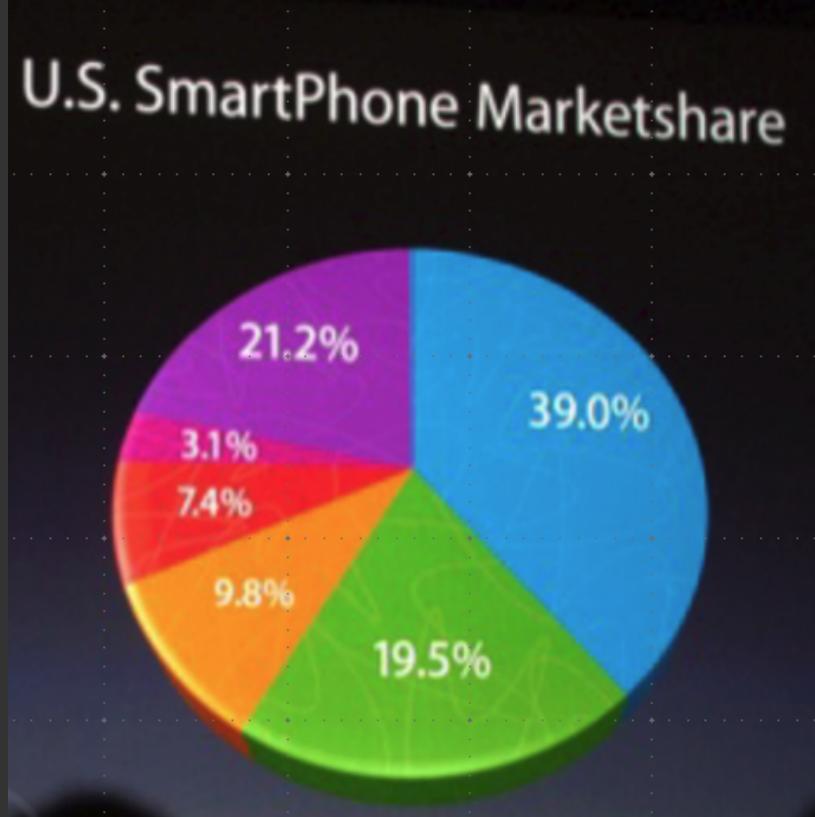
Quel est le problème de la dataviz ci-dessus ?

Datavisualisation – Attention à la manipulation



Voici la vraie représentation du graphique

Datavisualisation – Attention à la manipulation



En jouant sur la perspective, Apple donne l'impression d'être le numéro 2 des smartphones aux États-Unis alors qu'il est en fait troisième

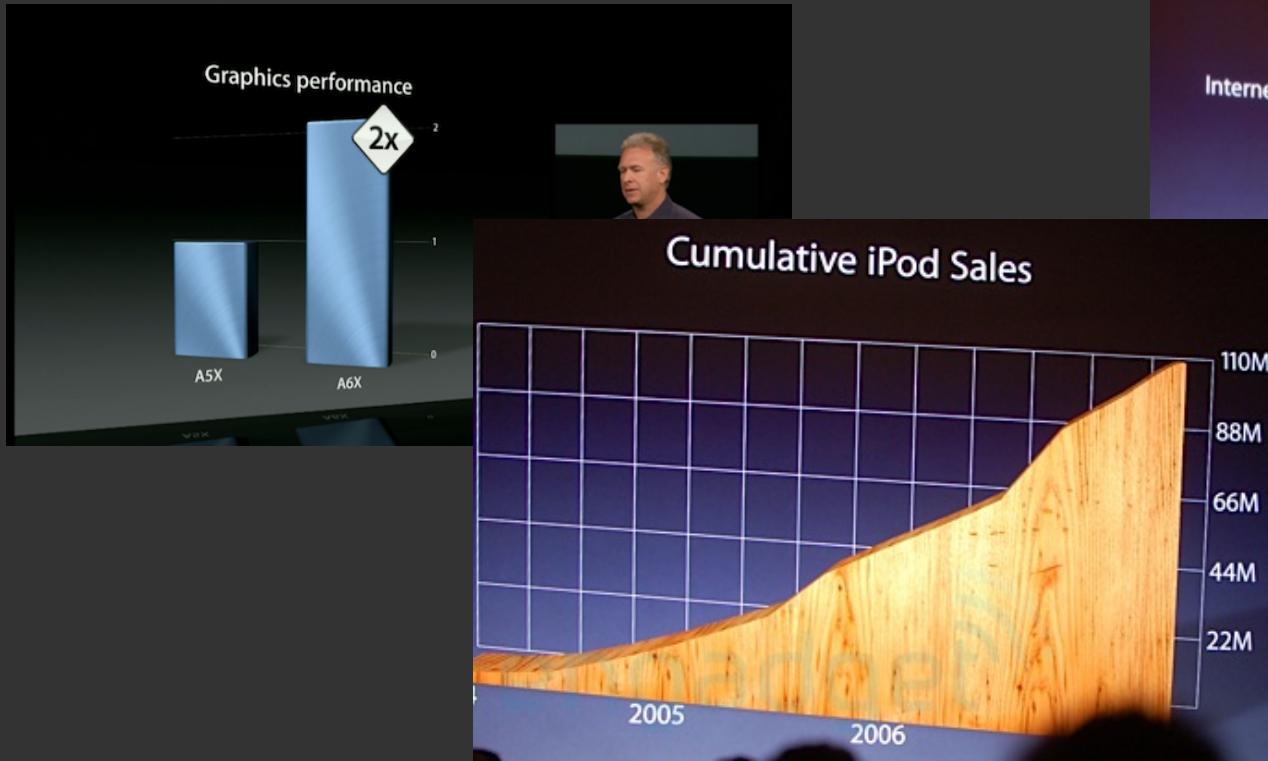
Datavisualisation – Mensonges communs

- Problème d'échelle / proportion
 - Ex : L'axe Y ne commence pas à 0
 - Ex : On se focalise sur une partie de l'axe X
- Données manquantes / troncature des axes
- L'utilisation de la 3D
- ...

Source(s) :

- <https://www.smalsresearch.be/comment-mentir-avec-un-graphique>
- https://apmeplorraine.fr/IMG/pdf/ac_lycee_pv136.pdf
- <https://www.youtube.com/watch?v=xZBrfxUo7kM>
- https://en.wikipedia.org/wiki/Misleading_graph - anglais
- <https://www.youtube.com/watch?v=RK0TSg6BU5s> - anglais

Datavisualisation – Mensonges communs

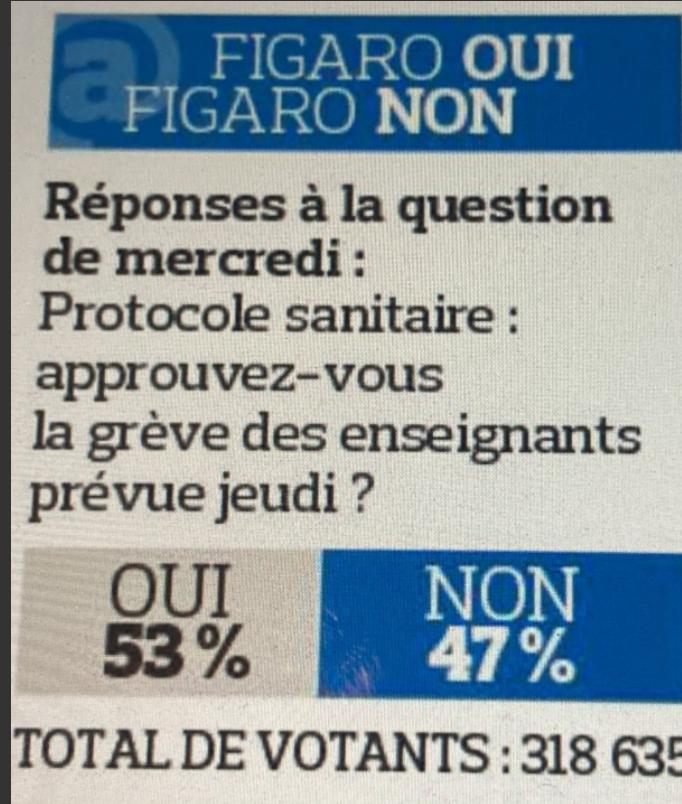


Le bruit graphique (chartjunk en anglais), la spécialité d'Apple

Source(s) :

- https://fr.wikipedia.org/wiki/Bruit_graphique

Datavisualisation – Mensonges communs



Le bruit graphique (chartjunk en anglais) en action dans le Figaro

Source(s) :

- https://fr.wikipedia.org/wiki/Bruit_graphique
- https://twitter.com/DEFAKATOR_Off/status/1481592487063998466/photo/1

Datavisualisation – Facteur de mensonge

- Concept inventé par Edward Tufte
- Définit le facteur de "non-respect des proportions" d'un graphique
- Lie factor en anglais

Source(s) :

- <https://www.smalsresearch.be/comment-mentir-avec-un-graphique>
- https://apmeplorraine.fr/IMG/pdf/ac_lycee_pv136.pdf
- <https://www.youtube.com/watch?v=xZBrfxUo7kM>
- https://en.wikipedia.org/wiki/Misleading_graph - anglais

Datavisualisation – Facteur de mensonge

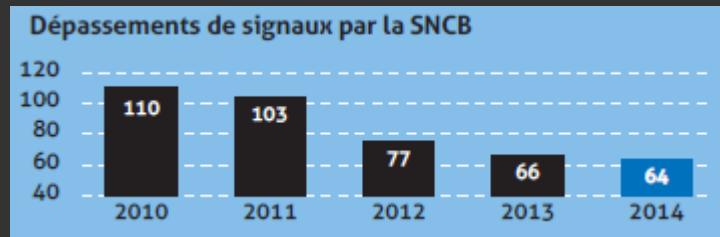
$$\text{facteur de mensonge} = \frac{\frac{|valeur\ 2(\text{graphique}) - valeur\ 1(\text{graphique})|}{valeur\ 1(\text{graphique})}}{\frac{|valeur\ 2(\text{jeu de données}) - valeur\ 1(\text{jeu de données})|}{valeur\ 1(\text{jeu de données})}}$$

- Doit être compris entre 0,95 et 1,05
 - Si le facteur est hors bornes → Graphique mensonger
- Nécessite de mesurer (avec une règle) pour les valeurs graphiques

Source(s) :

- <https://www.smalsresearch.be/comment-mentir-avec-un-graphique>
- https://apmeplorraine.fr/IMG/pdf/ac_lycee_pv136.pdf
- <https://www.youtube.com/watch?v=xZBrfxUo7kM>
- https://en.wikipedia.org/wiki/Misleading_graph - anglais

Datavisualisation – Facteur de mensonge



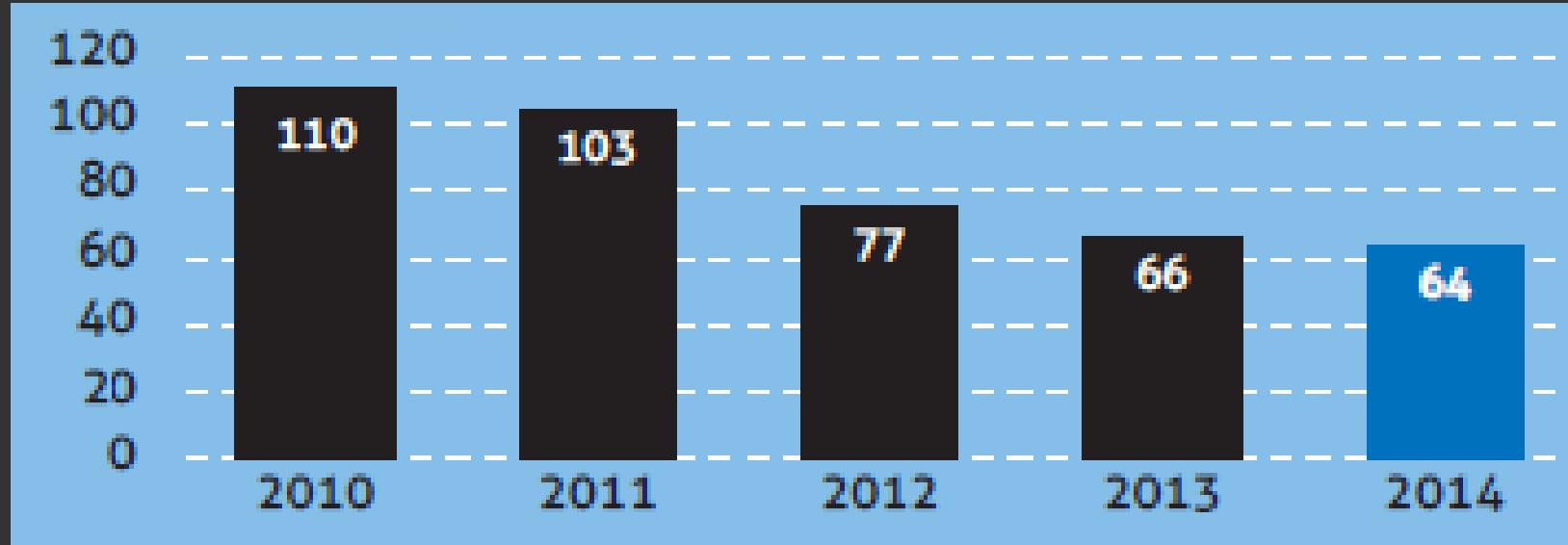
Taille des barres

- Barre 110 : 1 cm
- Barre 44 : 0,4 cm

$$\text{facteur de mensonge} = \frac{\frac{|valeur\ 2(graphique) - valeur\ 1(graphique)|}{valeur\ 1(graphique)}}{\frac{|valeur\ 2(jeu\ de\ données) - valeur\ 1(jeu\ de\ données)|}{valeur\ 1(jeu\ de\ données)}}$$

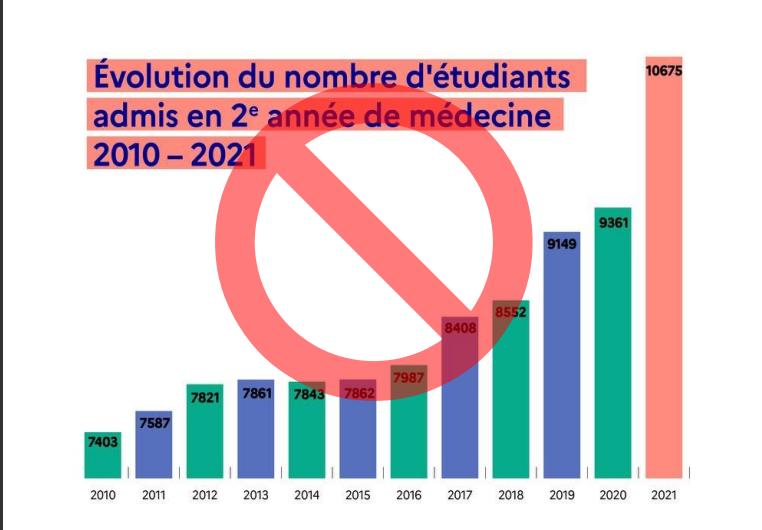
Calculez le facteur de mensonge de la courbe ci-dessus.

Datavisualisation – Facteur de mensonge

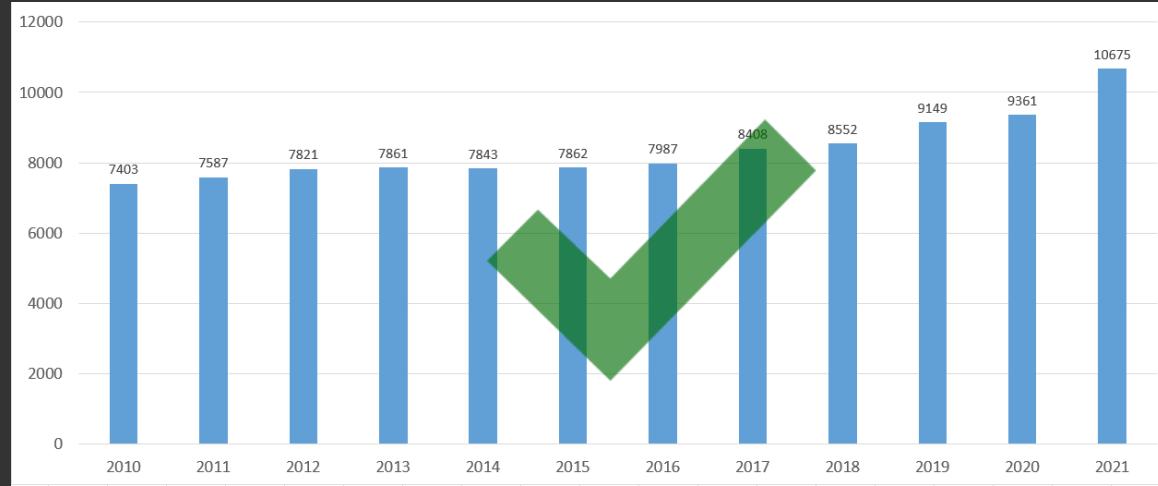


Voici le graphique corrigé
Ici le facteur de mensonge est proche de 1, c'est correct

Datavisualisation – Facteur de mensonge



Ici le facteur de mensonge est de ~19,46 !



Ici le facteur de mensonge est de ~1,04

Source(s) :

- <https://twitter.com/vidalfrederique/status/1390735956014256131>
- <https://mobile.twitter.com/batiste17180/status/1390832048189939712> - Le graphique corrigé

Datavisualisation - Règles

- Ne pas oublier les légendes
- Éviter les mensonges
 - Problèmes d'échelles / axes
 - Problèmes d'inclinaisons (voir cas Apple)
- Penser aux contrastes des couleurs
 - Éviter les couleurs trop similaires
- Éviter les fioritures → **Allez à l'essentiel**

Source(s) :

- <https://www.smalsresearch.be/comment-mentir-avec-un-graphique>
- https://apmeplorraine.fr/IMG/pdf/ac_lycee_pv136.pdf
- <https://www.youtube.com/watch?v=xZBrfxUo7kM>
- https://en.wikipedia.org/wiki/Misleading_graph - anglais
- <https://www.youtube.com/watch?v=RK0TSg6BU5s> - anglais
- <https://www.linkedin.com/pulse/edward-tufte-s-six-principles-graphical-integrity-radhika-raghu> - anglais

En résumé

- Plus d'exemples et de cas d'utilisations :
 - Quand utiliser quel graphique :
 - <https://www.data-to-viz.com/> - anglais
 - <http://www.infographicsblog.com/chart-suggestions-a-thought-starter-andrew-abela/>
- anglais / Source du graphique de la slide précédente
 - <https://gramener.github.io/visual-vocabulary-vega/#>
 - <https://python-graph-gallery.com/> – anglais
 - Exemples d'erreurs communes et à éviter – anglais
 - <https://venngage.com/blog/bad-infographics/>
 - <https://viz.wtf/>

Datavisualisation – Outils - Datawrapper

1 Télécharger des données ✓ 2 Vérifier et décrire ✓ **3 Visualisez** 4 Publiez et intégrez

Type de graphique Affiner Annoter Layout

Diagramme en barres Séparer les barres Barres empilées Barres regroupées
Barres à puces Tracé en points Tracé de la série Tracé de la flèche
Graphique en colonnes Graphique à barres groupées Graphique en barres empilées Lignes
Diagramme par zone Diagramme circulaire Graphique en anneau Diagrammes circulaires multiples
Anneaux multiples Graphique en nuage de points Anneau de choix Tableau

[Insérez le titre ici]
Apple sales by product in shipped units, 2000 to 2017.

Apple stopped to report iPod sales at the end of 2014.
Source: Apple • Crée avec Datawrapper

Astuce: Si la visualisation ne correspond pas à vos attentes, essayez de transposer les données

SIZE: 600 x 400 TEST POUR PERSONNES: DANTONINNES, Norm, Deut, Prot, Trit, Achr

The screenshot shows the Datawrapper interface. It's a step-by-step process: 1. Télécharger des données (Download data), 2. Vérifier et décrire (Check and describe), 3. Visualisez (Visualize), and 4. Publiez et intégrez (Publish and integrate). Step 3 is highlighted in red. On the left, there's a grid of icons representing different chart types: bar charts, line charts, pie charts, etc. The 'Lignes' (Lines) icon is selected. In the center, there's a line chart titled 'Apple sales by product in shipped units, 2000 to 2017'. The chart tracks three products: iPhone, iPad, and Mac. The iPhone shows the most significant growth, especially after 2010. The Mac shows lower sales, peaking around 2010. The iPad shows moderate growth. At the bottom, there's a note that Apple stopped reporting iPod sales at the end of 2014. The bottom right corner includes options for adjusting the size of the visualization and testing it for different types of users (DANTONINNES, Norm, Deut, Prot, Trit, Achr).

- Gratuit
- Nécessite inscription pour enregistrer ses créations
- Gestion de la colorimétrie pour l'accessibilité

Source(s) :

- <https://app.datawrapper.de/chart/5KRJy/visualize>

Datavisualisation – Outils - flourish

Data

	A	B	C	D	E	F	G	H	I
1	Date	Auvergne-Rhône-Alpes	Bourgogne-Franche-Comté	Bretagne	Centre-Val de Loire	Corse	Grand Est	Hauts-de-France	Île-de-France
2	3/4/20	49	16	23	0	0	38	65	55
3	3/5/20	64	39	29	2	3	77	74	76
4	3/6/20	75	78	32	11	3	136	97	89
5	3/7/20	102	129	40	16	5	250	173	104
6	3/10/20	249	123	83	19	42	489	249	468
7	3/11/20	264	155	94	20	51	578	263	492
8	3/12/20	329	219	117	30	64	699	349	577
9	3/13/20	416	272	155	33	85	911	397	721
10	3/14/20	454	339	182	54	106	1085	466	949
11	3/15/20	523	412	207	71	115	1378	507	1209

Preview Data

Saved

Upload data file

Data

Labels/time REQUIRED

A column of names or times

A

Values

One or more columns of numbers

B-I

Charts grid

If specified and "Grid of charts" view is on, creates a separate mini chart for each value found in the column.

Info for custom popups

One or more columns of information to include in custom popups

Auvergne-Rhône-Alpes Bourgogne-Franche-Comté Bretagne Centre-Val de Loire Corse Grand Est Hauts-de-France Île-de-France

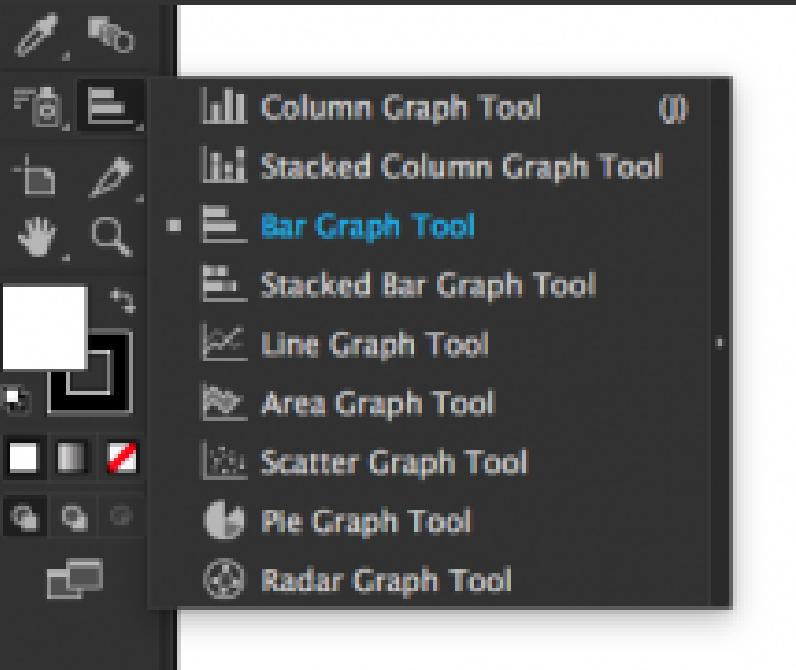


- Gratuit et 100 % en ligne
- Nécessite une inscription

Source(s) :

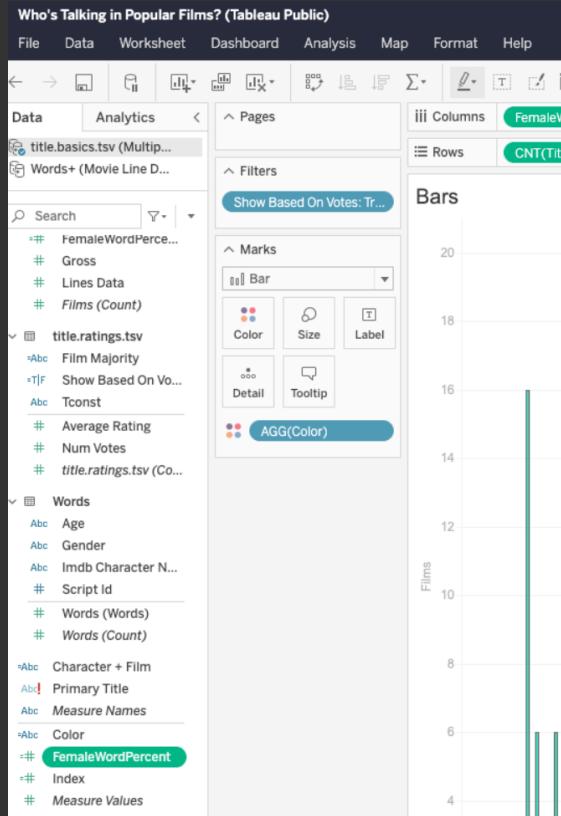
- <https://flourish.studio>

Datavisualisation – Outils - Illustrator



- Logiciel payant
- Limité en terme de choix de dataviz
- Propose plus de libertés en terme de design / mise en page
- Permet de faire des infographies

Datavisualisation – Outils – Tableau Public



- En ligne et freemium
- Permet de faire des infographies interactives (sans code)
- **A ne pas confondre avec tableau**
 - Même entreprise différent produit
- **Rend le projet public, accessible à tous**

Source(s) :

- <https://public.tableau.com/s/>

Datavisualisation – Outils – En vrac

Ne nécessite pas de développer

- <https://charts.livegap.com/#TypesofCharts> – Gratuit
- <https://gallery.keshif.me/VisTools> – Outil d'aide à la décision
- <https://creately.com/> - Outil en ligne freemium
- <https://infogram.com/> - Outil en ligne freemium

Nécessite de développer

- <https://d3js.org/> - Bibliothèque javascript très puissante
- <https://www.chartjs.org/> - Bibliothèque javascript
- seaborn / matplotlib – <https://seaborn.pydata.org/> - Bibliothèque Python
- <https://docs.bokeh.org/> - Bibliothèque Python et javascript

Exemples de datavisualisations

- <https://www.reddit.com/r/Infographics/>
- <https://www.informationisbeautiful.net/visualizations/based-on-a-true-story/>
 - <https://informationisbeautiful.net/>

Quelque soit l'outil utilisé pour vos graphiques, il faudra certainement nettoyer, filtrer, extraire vos données. N'oubliez pas pandas.

Datavisualisation - Étapes

1. Acquisition des données
2. Transformation
3. Filtrage
4. Exploration
5. Représentation
6. Affinage
7. Interagir (facultatif)

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation - Acquisition des données

- Webscraping
- Sites de données ouvertes
 - data.gouv.fr
 - insee.fr
 - ec.europa.eu/eurostat/web/main/data/database
 - ...
- Flux RSS
- Sondages

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Sondages - Attention

- Toujours vérifier :
 - Le panel (ensemble des personnes sondées)
 - Les questions posées
 - Elles peuvent parfois porter à confusion et donc fausser les résultats
 - Les données brutes

Source(s) :

- <https://www.franceinter.fr/emissions/secrets-d-info/secrets-d-info-18-mars-2017>

Notez que vous devrez parfois croiser des jeux de données, donc travailler avec plusieurs jeux en même temps

Datavisualisation - Transformation

- Homogénéité :
 - Unités
 - Casse de chaînes de caractères
- Outils utilisables :
 - pandas
 - Google Refine (sans programmation)

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation - Filtrages

- Suppression des données :
 - Incorrectes
 - Absurdes
 - Manquantes
 - Inutiles

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation - Exploration

- Calcul d'indicateurs :
 - Moyenne
 - Médiane
 - Mode
 - Écart-type
 - ...

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation - Représentation

- Réalisation de dataviz
- Choix des composantes :
 - Qu'est-ce que je choisis d'afficher ?

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation - Affinage

- Phase d'amélioration
- Correction d'éventuelle erreurs sur le filtrages ou dataviz
- Revenir à la phase "filtrage" en cas de problèmes

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

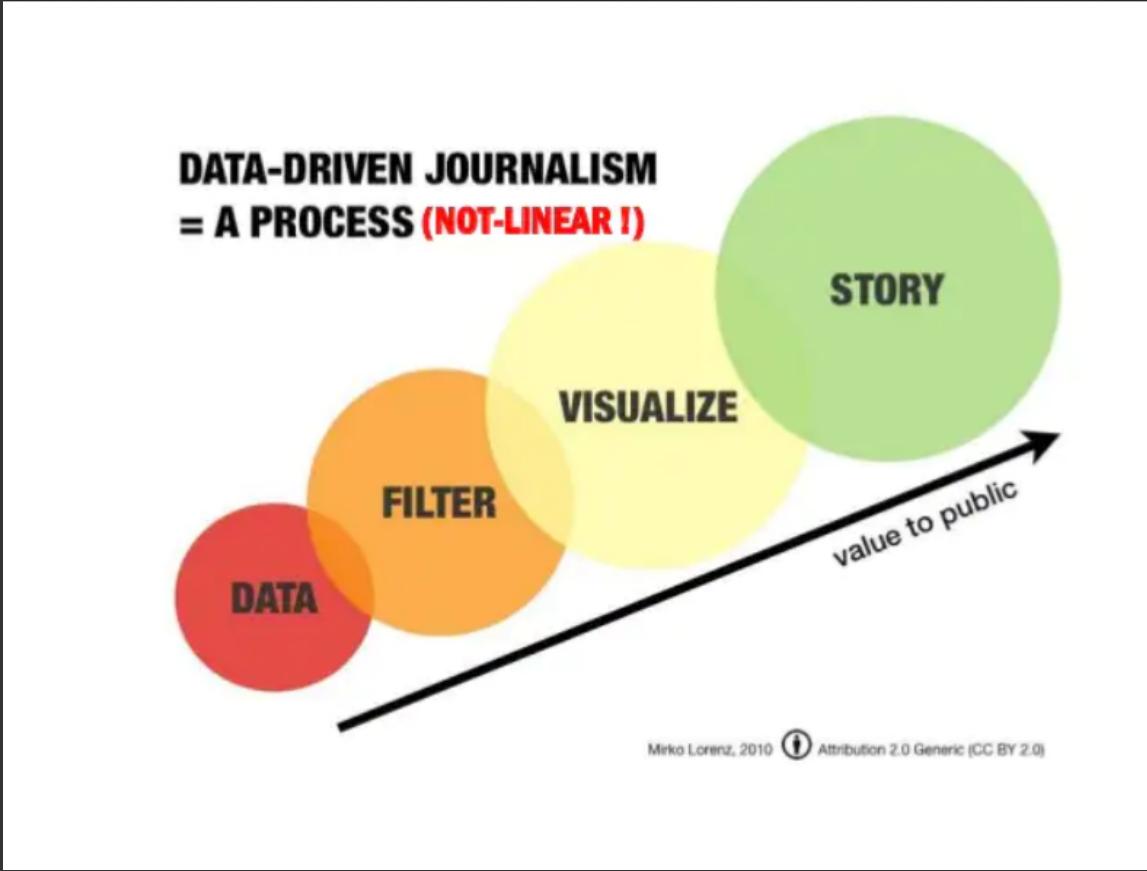
Datavisualisation - Interagir

- Phase facultative
- Permet à l'utilisateur d'interagir avec la dataviz. Outils comme :
 - plotly
 - D3
 - ...

Source(s) :

- <https://www.youtube.com/watch?v=xZBrfxUo7kM>

Datavisualisation – Le processus en résumé



Source(s) :

- <http://mirkolorenz.com/> - Mirko Lorenz

