Visualiser les données

avec la bibliothèque Plotly

Visualiser

Qu'est-ce qu'une bonne visualisation

?

Invariants & culture

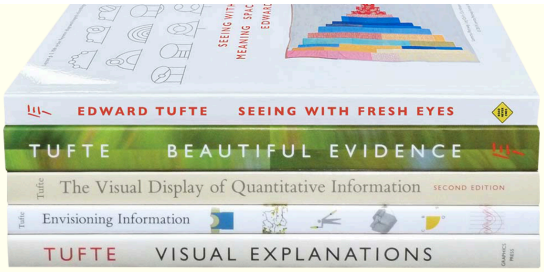
- L'être humain est un animal visuel
  - Détection de formes
  - Identification des couleurs
- Des codes culturels
  - Rouge = danger/important
  - etc.

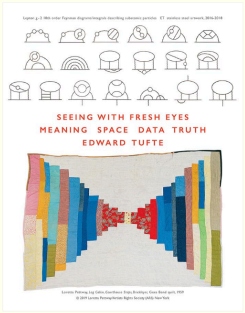
Un domaine vaste

À la croisée entre design & statistiques

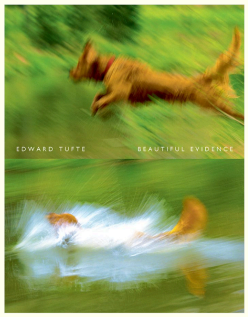
Edward Tufte is a statistician and artist, and Professor Emeritus of Political Science, Statistics, and Computer Science at Yale University. He wrote, designed, and self-published 5 classic books on data visualization and has been described by *The New York Times* as the "Leonardo da Vinci of data," and *Bloomberg* as the "Galileo of graphics."

He is the founder of Graphics Press, ET Modern Gallery/Studio, and Hogpen Hill Farms – a 234-acre tree farm and sculpture park in northwest Connecticut, which will show his artworks and remain open space in perpetuity.

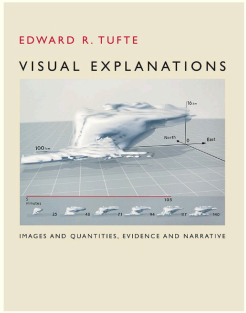




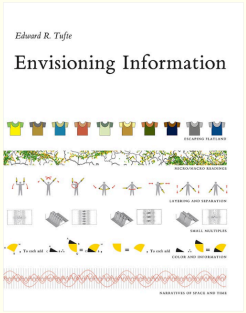
Seeing with Fresh Eyes: Meaning, Space, Data, Truth
Edward R. Tufte



Beautiful Evidence
Edward R. Tufte
2006



Visual Explanations: Images and Quantities, Evidence and Narrative



Envisioning Information
Edward R. Tufte
1990

De nombreuses recommandations

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

Ten Simple Rules for Better Figures

Nicolas P. Rougier[1,2,3]*, Michael Droettboom[4], Philip E. Bourne[5]

1 INRIA Bordeaux Sud-Ouest, Talence, France, **2** LaBRI, UMR 5800 CNRS, Talence, France, **3** Institute of Neurodegenerative Diseases, UMR 5293 CNRS, Bordeaux, France, **4** Space Telescope Science Institute, Baltimore, Maryland, United States of America, **5** Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America

Scientific visualization is classically defined as the process of graphically displaying scientific data. However, this process is far from direct or automatic. There are so many different ways to represent the same data: scatter plots, linear plots, bar plots, and pie charts, to name just a few. Furthermore, the same data, using the same type of plot, may be perceived very differently depending on who is looking at the figure. A more accurate definition for scientific visualization would be a graphical interface between people and data. In this short article, we do not pretend to explain everything about this interface; rather, see [1,2] for introductory work. Instead we aim to provide a basic set of rules to improve figure design and to explain some of the common pitfalls.

Rule 1: Know Your Audience

Given the definition above, problems arise when how a visual is perceived differs significantly from the intent of the conveyer. Consequently, it is important to identify, as early as possible in the design process, the audience and the message the visual is to convey. The

Rule 2: Identify Your Message

A figure is meant to express an idea or introduce some facts or a result that would be too long (or nearly impossible) to explain only with words, be it for an article or during a time-limited oral presentation. In this context, it is important to clearly identify the role of the figure, i.e., what is the underlying message and how can a figure best express this message? Once clearly identified, this message will be a strong guide for the design of the figure, as shown in Figure 2. Only after identifying the message will it be worth the time to develop your figure, just as you would take the time to craft your words and sentences when writing an article only after deciding on the main points of the text. If your figure is able to convey a striking message at first glance, chances are increased that your article will draw more attention from the community.

Rule 3: Adapt the Figure to the Support Medium

A figure can be displayed on a variety of media, such as a poster, a computer monitor, a projection screen (as in an oral

must be kept simple and the message must be visually salient in order to grab attention, as shown in Figure 3. It is also important to keep in mind that during oral presentations, figures will be video-projected and will be seen from a distance, and figure elements must consequently be made thicker (lines) or bigger (points, text), colors should have strong contrast, and vertical text should be avoided, etc. For a journal article, the situation is totally different, because the reader is able to view the figure as long as necessary. This means a lot of details can be added, along with complementary explanations in the caption. If we take into account the fact that more and more people now read articles on computer screens, they also have the possibility to zoom and drag the figure. Ideally, each type of support medium requires a different figure, and you should abandon the practice of extracting a figure from your article to be put, as is, in your oral presentation.

Rule 4: Captions Are Not Optional

Whether describing an experimental setup, introducing a new model, or

Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833.

All the rules

1. Know Your Audience
2. Identify Your Message
3. Adapt the Figure to the Support Medium
4. Captions Are Not Optional

5. Do Not Trust the Defaults
6. Use Color Effectively
7. Do Not Mislead the Reader
8. Avoid “Chartjunk”
9. Message Trumps Beauty
10. Get the Right Tool

Des choix à tous le niveaux

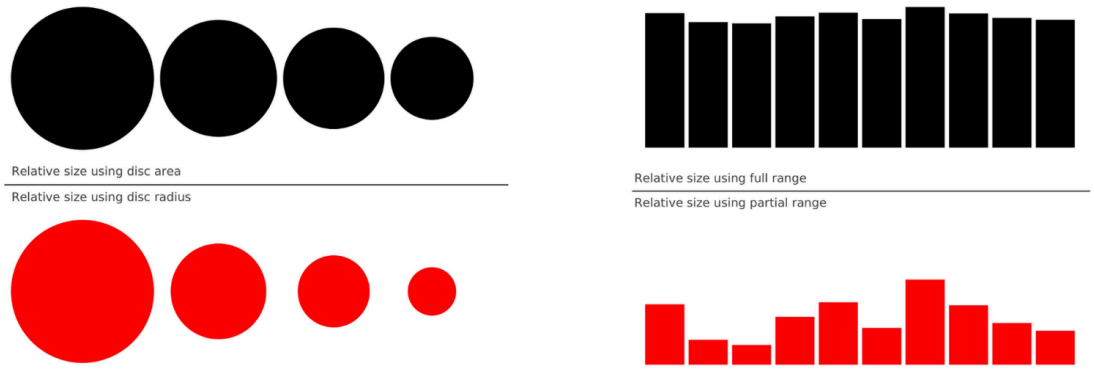
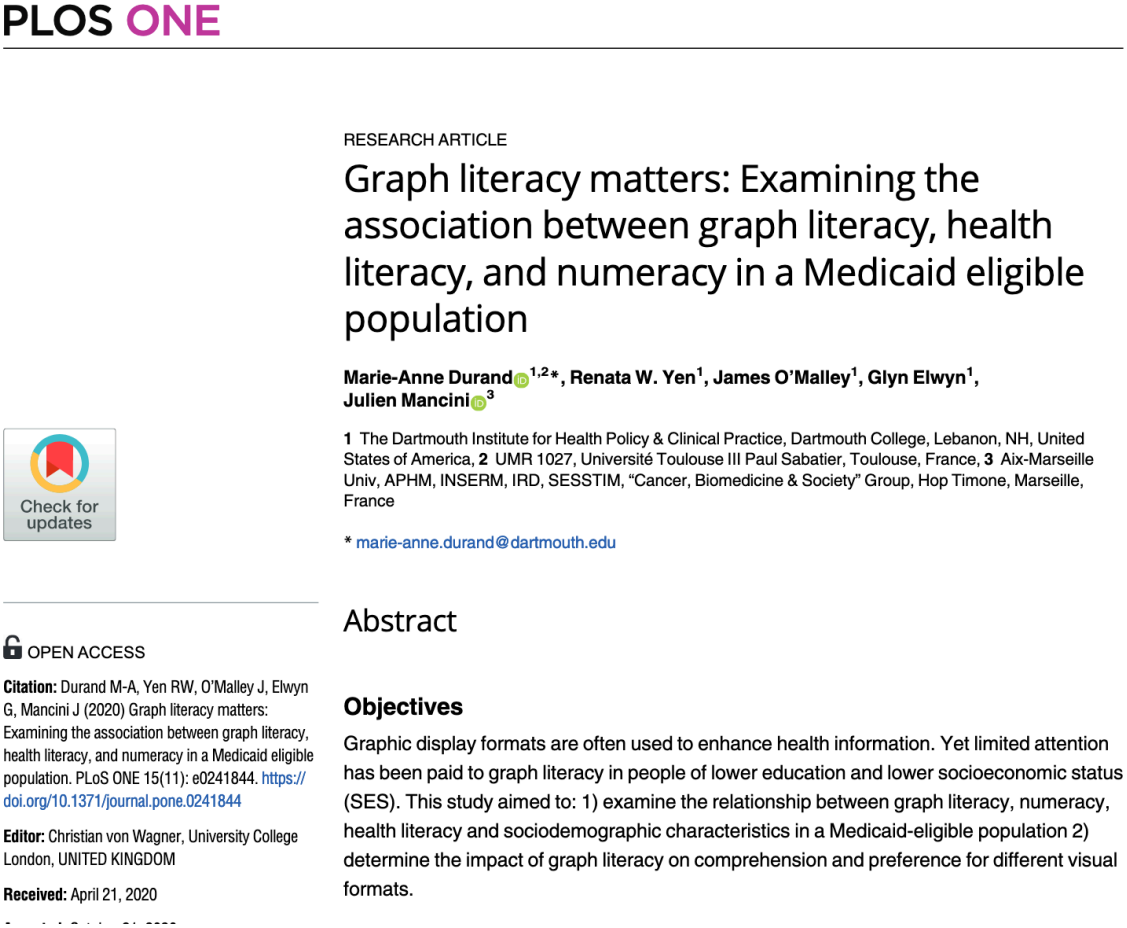


Figure 6. Do not mislead the reader. On the left part of the figure, we represented a series of four values: 30, 20, 15, 10. On the upper left part, we used the disc area to represent the value, while in the bottom part we used the disc radius. Results are visually very different. In the latter case (red circles), the last value (10) appears very small compared to the first one (30), while the ratio between the two values is only 3:1. This situation is actually very frequent in the literature because the command (or interface) used to produce circles or scatter plots (with varying point sizes) offers to use the radius as default to specify the disc size. It thus appears logical to use the value for the radius, but this is misleading. On the right part of the figure, we display a series of ten values using the full range for values on the top part (y axis goes from 0 to 100) or a partial range in the bottom part (y axis goes from 80 to 100), and we explicitly did not label the y-axis to enhance the confusion. The visual perception of the two series is totally different. In the top part (black series), we tend to interpret the values as very similar, while in the bottom part, we tend to believe there are significant differences. Even if we had used labels to indicate the actual range, the effect would persist because the bars are the most salient information on these figures.
doi:10.1371/journal.pcbi.1003833.g006

Des conséquences quotidiennes



PLOS ONE

RESEARCH ARTICLE

Graph literacy matters: Examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population

Marie-Anne Durand[1,2]*, **Renata W. Yen**[1], **James O'Malley**[1], **Glyn Elwyn**[1], **Julien Mancini**[3]

1 The Dartmouth Institute for Health Policy & Clinical Practice, Dartmouth College, Lebanon, NH, United States of America, **2** UMR 1027, Université Toulouse III Paul Sabatier, Toulouse, France, **3** Aix-Marseille Univ, APHM, INSERM, IRD, SESSTIM, "Cancer, Biomedicine & Society" Group, Hop Timone, Marseille, France

* marie-anne.durand@dartmouth.edu

Check for updates

OPEN ACCESS

Citation: Durand M-A, Yen RW, O'Malley J, Elwyn G, Mancini J (2020) Graph literacy matters: Examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population. PLoS ONE 15(11): e0241844. https://doi.org/10.1371/journal.pone.0241844

Editor: Christian von Wagner, University College London, UNITED KINGDOM

Received: April 21, 2020

Abstract

Objectives

Graphic display formats are often used to enhance health information. Yet limited attention has been paid to graph literacy in people of lower education and lower socioeconomic status (SES). This study aimed to: 1) examine the relationship between graph literacy, numeracy, health literacy and sociodemographic characteristics in a Medicaid-eligible population 2) determine the impact of graph literacy on comprehension and preference for different visual formats.

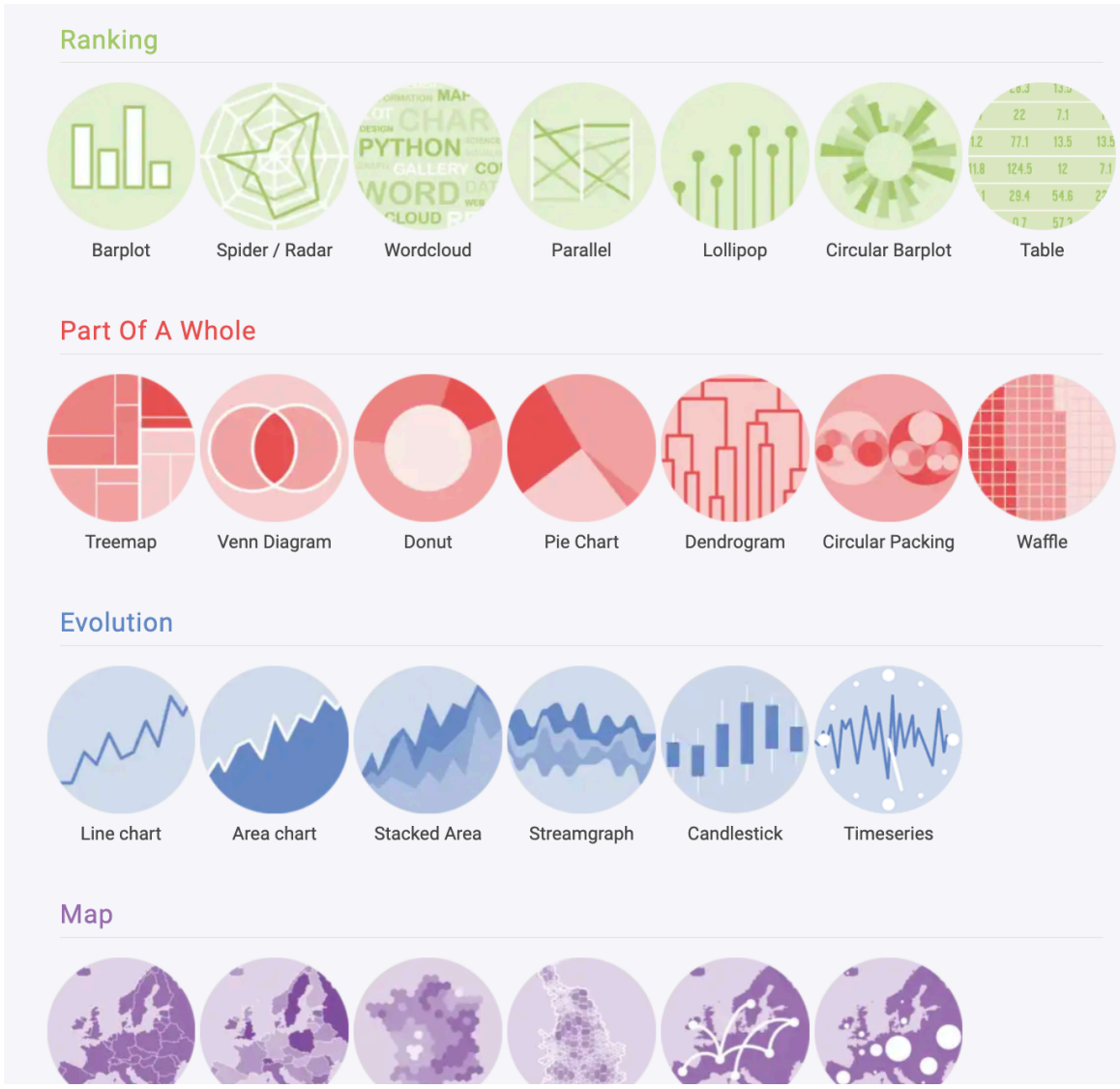
Une diversité de pratiques

En fait les visualisations couvrent beaucopp de situations différentes

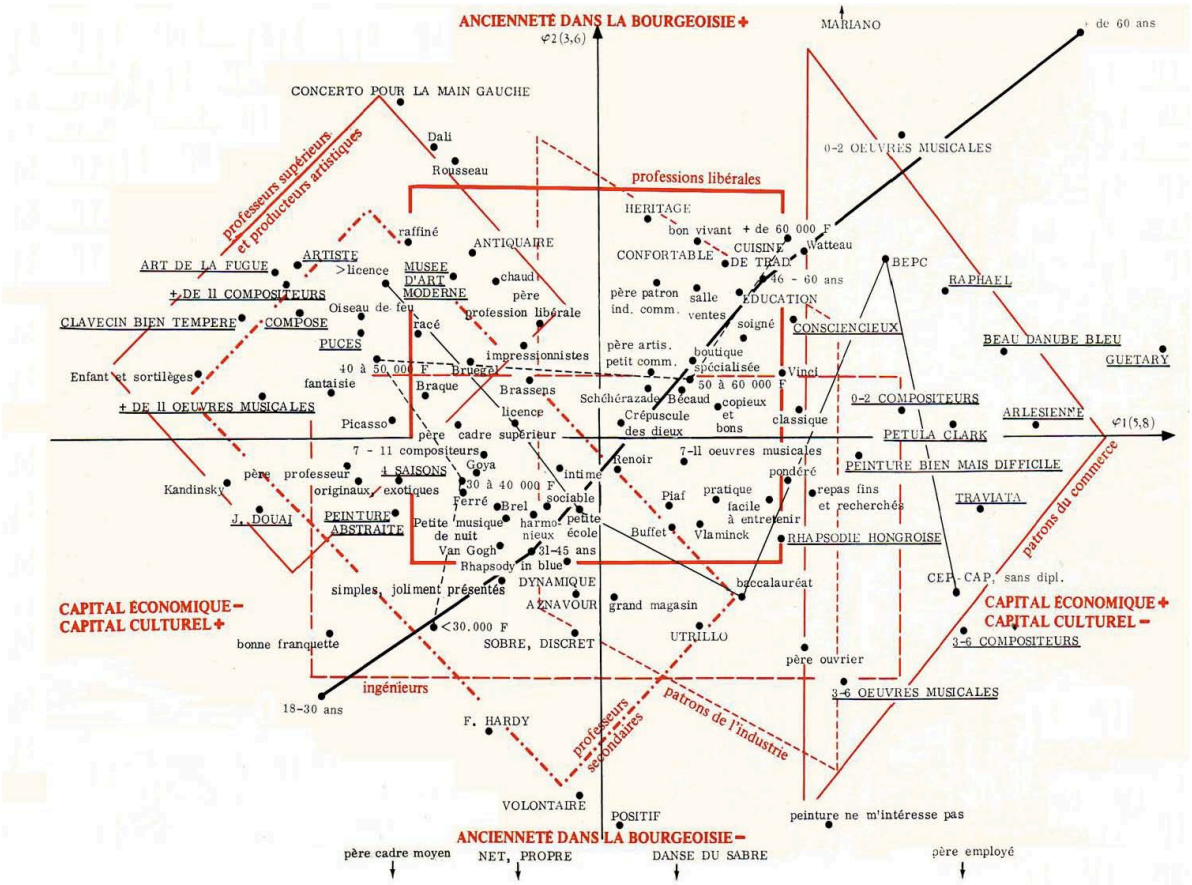
- De la vue d'artiste
- A la visualisation issue des données
- Et tous les intermédiaires

Une diversité de visualisations

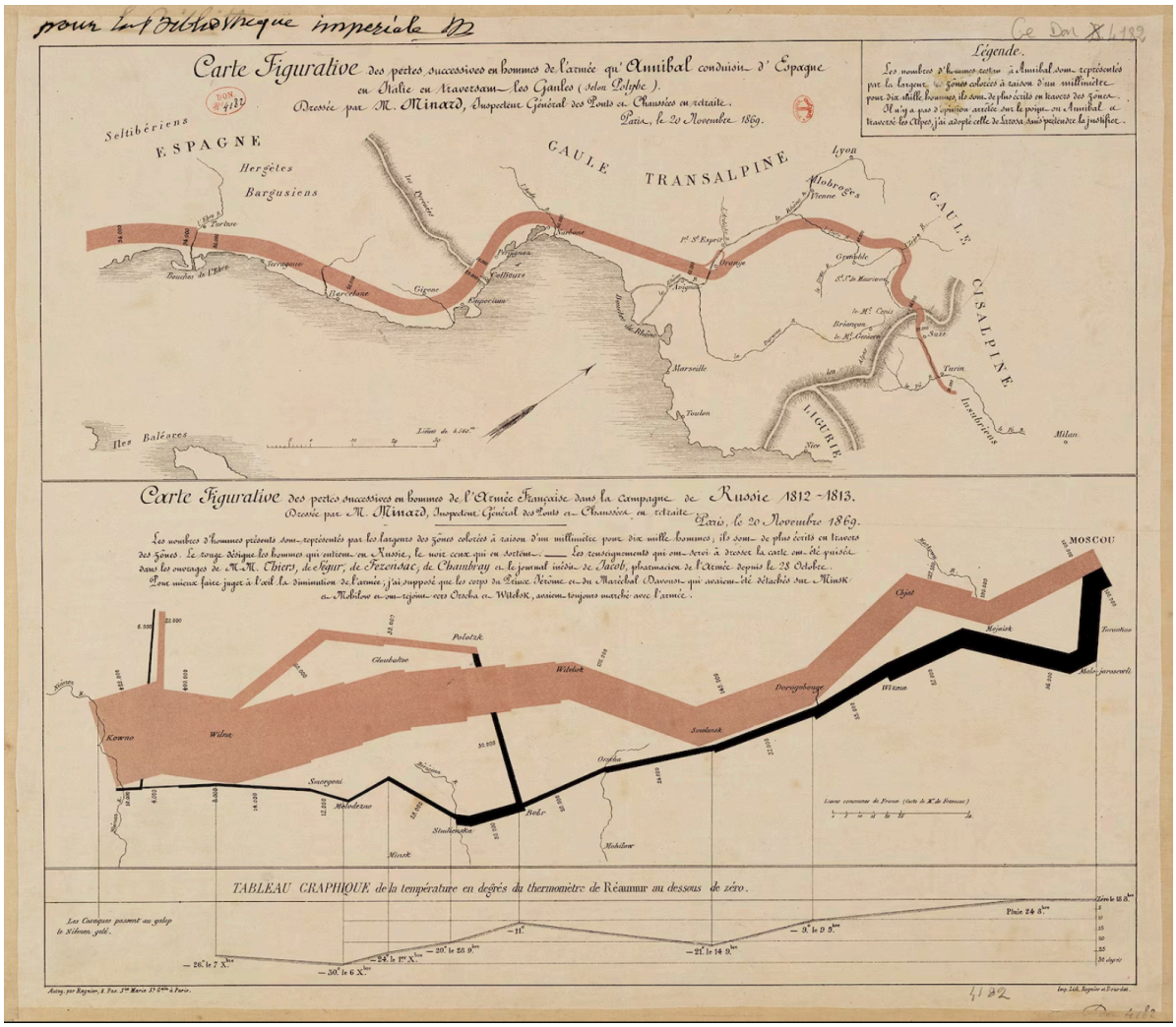
Python graph gallery


Ranking
Barplot
Spider / Radar
Wordcloud
Parallel
Lollipop
Circular Barplot
Table
Part Of A Whole
Treemap
Venn Diagram
Donut
Pie Chart
Dendrogram
Circular Packing
Waffle
Evolution
Line chart
Area chart
Stacked Area
Streamgraph
Candlestick
Timeseries
Map

Un peu de Bourdieu



Le “meilleur graphique”



Pour en savoir plus...

Réseau retravaillé

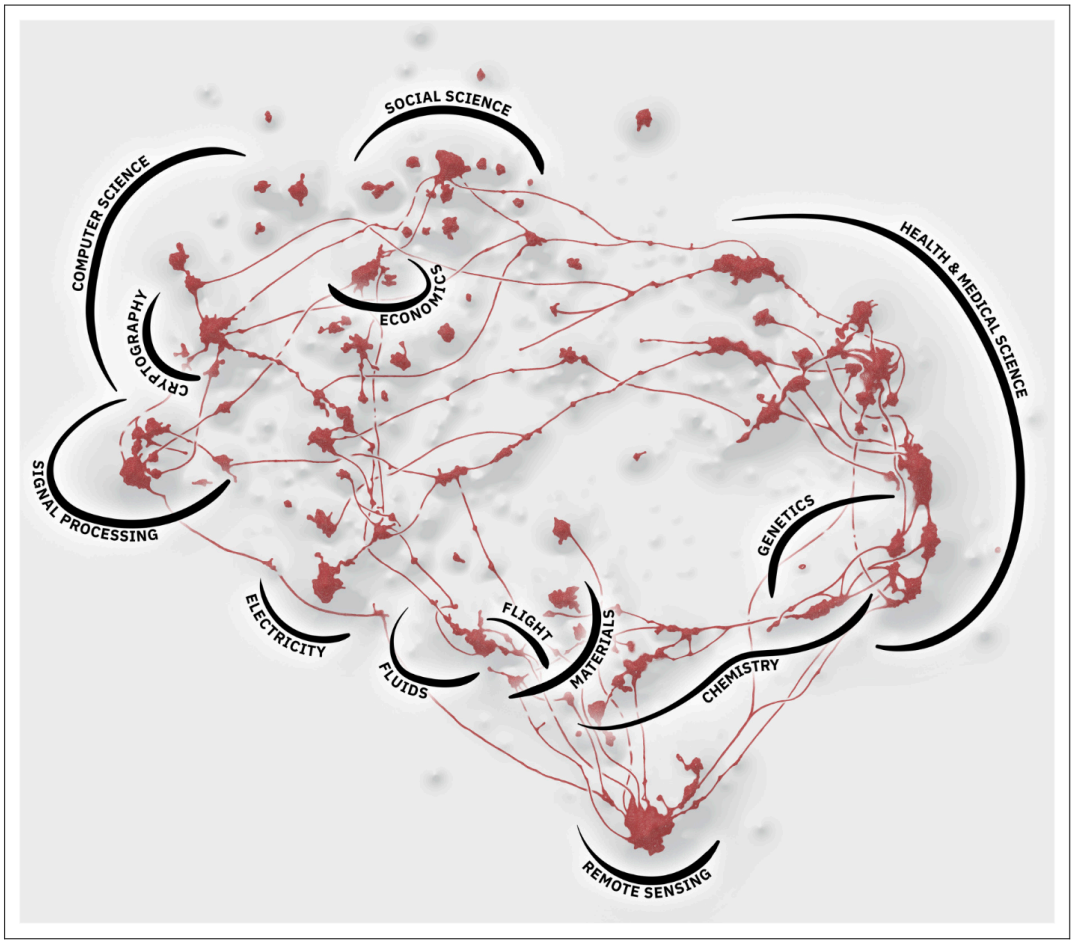
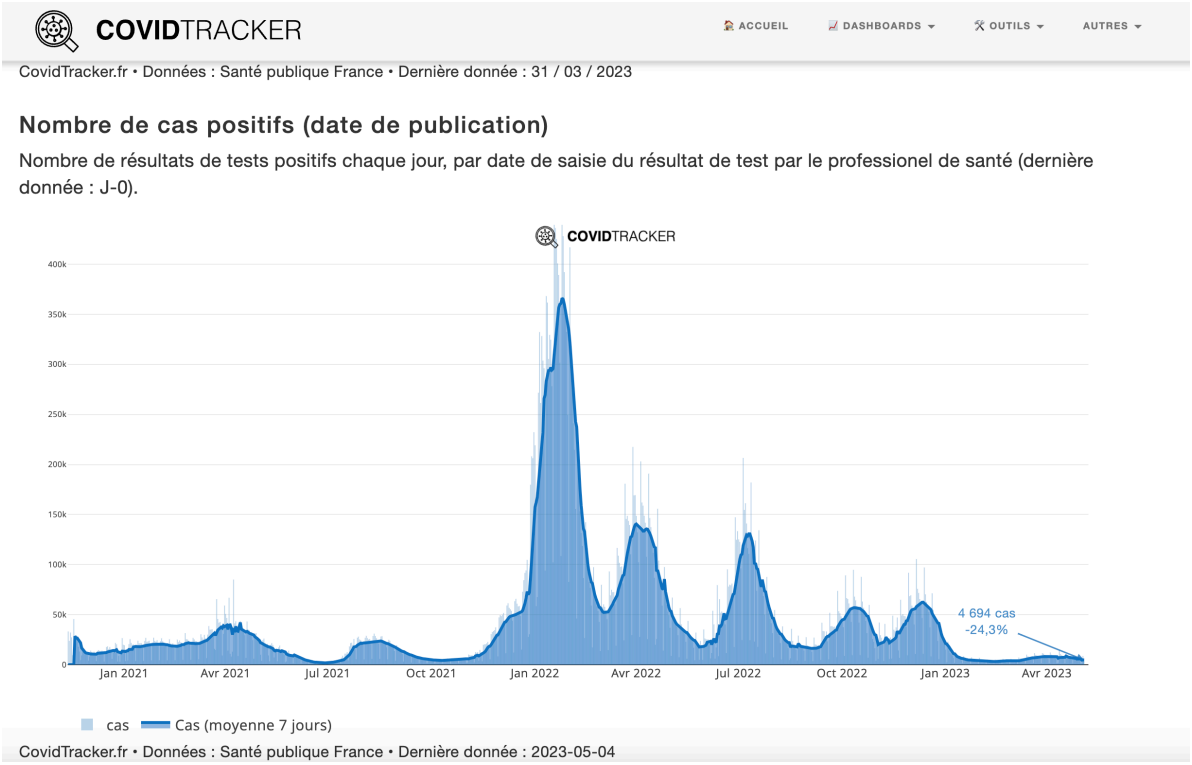
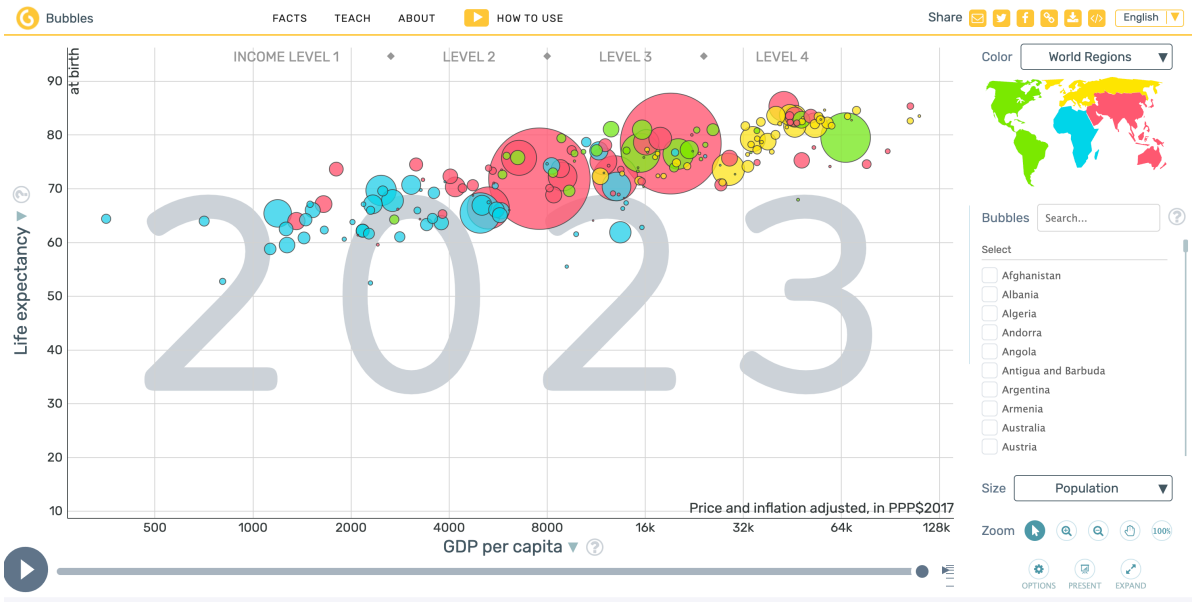


Figure 9. The main annotated areas of the map are highlighted in black. They broadly correspond to scientific fields. We have intentionally kept the borders of the areas vague, allowing them to overlap and transition into one another where the landmarks and bridges dictate it. It is also worth noting that the areas do not entirely cover the map, as some landmarks do not fit in this thematic breakdown. For instance, the "social science" cluster (on top, see also Figure 2) loosely encompasses clusters about sociological subfields ("social network analysis"), techniques used in sociology but not only ("principal component analysis"), sociology of algorithms ("socially responsible AI"), adjacent fields ("text search systems") and some more distant fields ("fuzzy logic").

Applications : Covid Traker



Animé



Gap minder

L'importance des couleurs

Visualization Viewpoints

Editor:
Theresa-Marie Rhyne

Rainbow Color Map (Still) Considered Harmful

David Borland and Russell M. Taylor II
University of North Carolina at Chapel Hill

Research has shown that the rainbow color map is rarely the optimal choice when displaying data with a pseudocolor map. The rainbow color map confuses viewers through its lack of perceptual ordering, obscures data through its uncontrolled luminance variation, and actively misleads interpretation through the introduction of non-data-dependent gradients.

Despite much published research on its deficiencies, the rainbow color map is prevalent in the visualization community. We present survey results showing that the rainbow color map continues to appear in more than half of the relevant papers in IEEE Visualization Conference proceedings; for example, it appeared on 61 pages in 2005. Its use is encouraged by its selection as the default color map used in most visualization toolkits that we inspected. The visualization community must do better.

In this article, we reiterate the characteristics that make the rainbow color map a poor choice, provide examples that clearly illustrate these deficiencies even on simple data sets, and recommend better color maps

mercials, weather forecasts, and even the IEEE Visualization Conference 2006 call for papers, just to name a few. The problem with this wide use of the rainbow color map is that research shows that it is rarely, if ever, the optimal color map for a given visualization.[1-6] Here we will discuss the rainbow color map's characteristics of confusing the viewer, obscuring data, and actively misleading interpretation.

Confusing

For all tasks that involve comparing relative values, the color map used should exhibit perceptual ordering. A simple example of a perceptually ordered color map is the gray-scale color map. Increasing luminance from black to white is a strong perceptual cue that indicates values mapped to darker shades of gray are lower in value than values mapped to lighter shades of gray. This mapping is natural and intuitive.

The rainbow color map is certainly ordered—from a shorter to longer wavelength of light (or vice versa)—but it's not perceptually ordered. If people are given a

Exemple de l'effet des couleurs

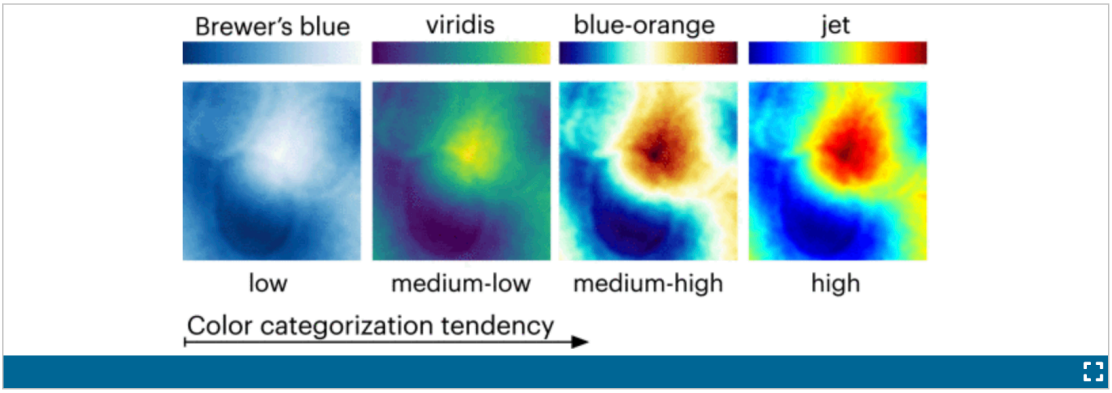


Fig. 2.
Four colormaps (used in Experiment 1) representing increasing levels of color categorization. Note the effect of categorization on the perceptibility of features in the scalar fields: the 'bean' shaped features are most visible in Brewer's *blue*, but those features become less discernible as color categorization increases. However, categorization may afford a better overview of the data's spatial extent and distribution.

Rainbow Colormaps: What are They Good and Bad for?

Importance de la composition

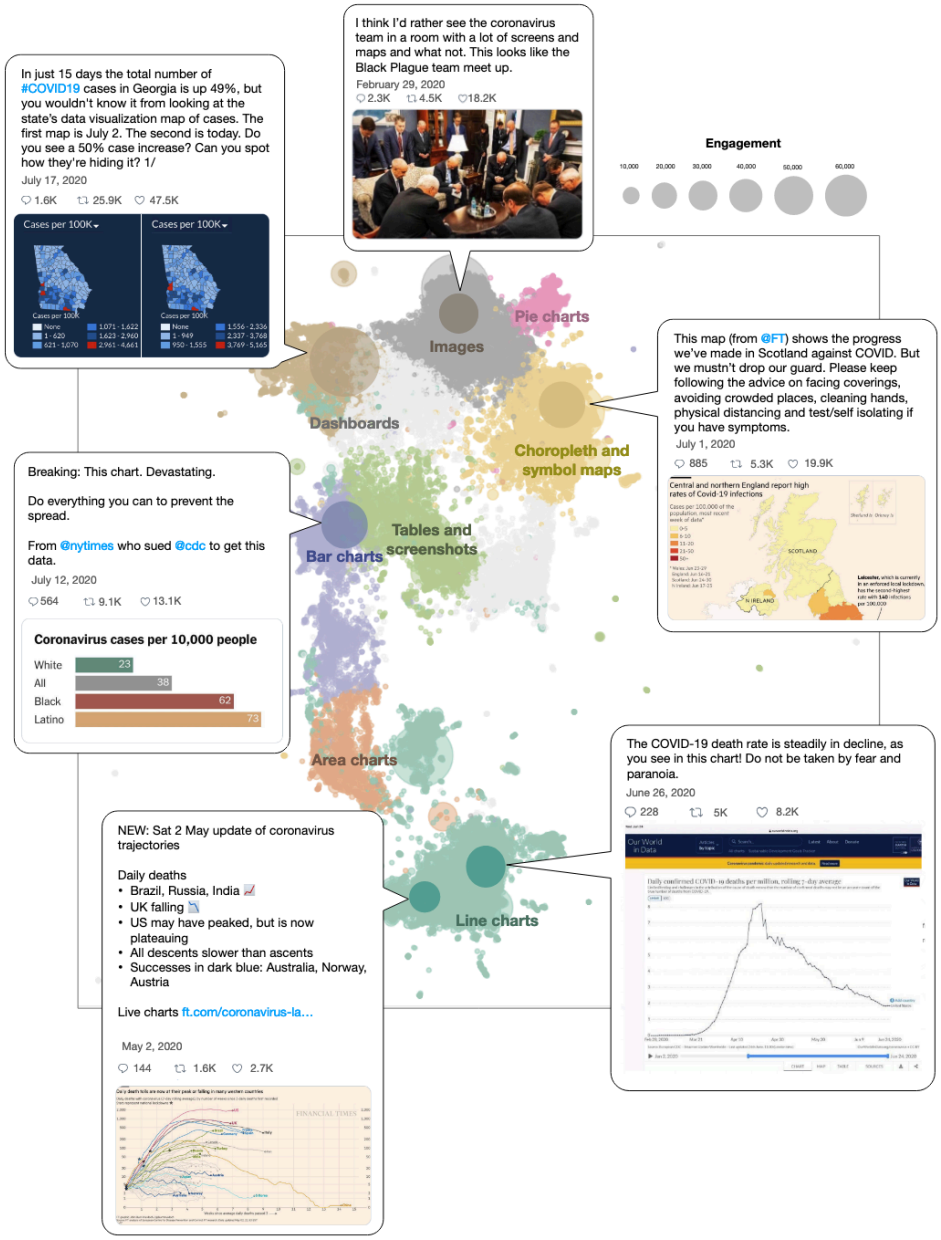
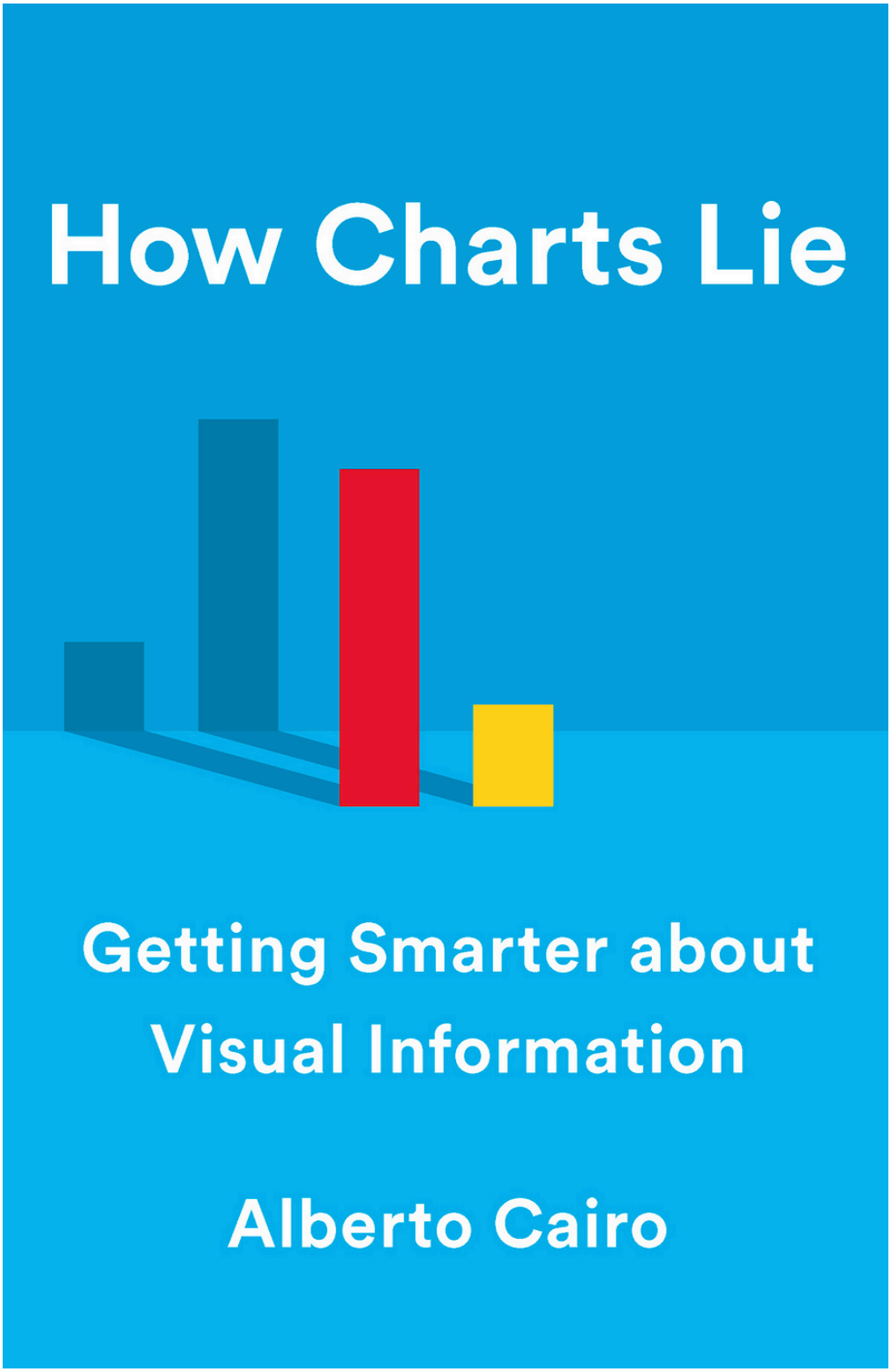


Figure 1: A UMAP visualization of feature embeddings of media found in our Twitter corpus. Color encodes labeled clusters, and size encodes the amount of engagement the media received (i.e., the sum of replies, favorites, retweets, and quote tweets).

Attention aux chartjunk



Des visualisations politiques

- Au sens large, elles véhiculent plus que des données
  - Importance dans la publication scientifique
- Au sens restreint, elles peuvent peser dans le débat

Toujours possible d'améliorer

- Colorblind
- Alt pour compléter la visualisation
- …

Une diversité d'outils

- Métiers
  - Cartes
  - Réseaux …
- Logiciels génériques
  - Tableurs …
- Programmation
  - R avec ggplot2
  - Python avec matplotlib, seaborn, plotly, …

Utiliser Python

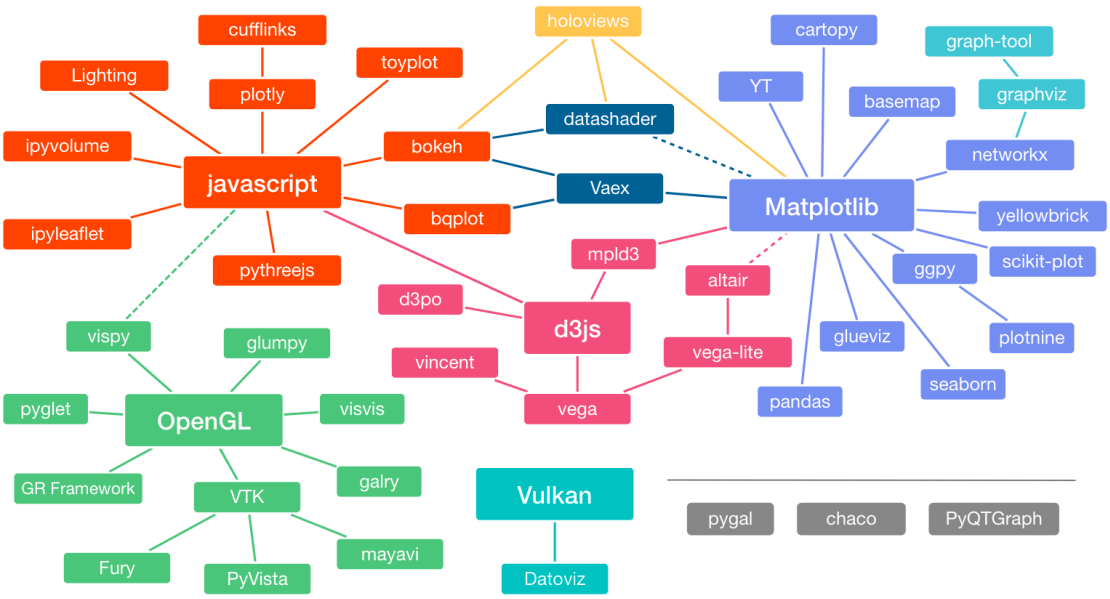


Figure 1
Python scientific visualisation landscape in 2018 (not exhaustive). Adapted from the original idea of Jake Vanderplas. **Sources:** github.com/rougier/python-visualization-landscape

Figure 1: Nicolas Rougier, *Scientific Visualization: Python + Matplotlib*, p.19

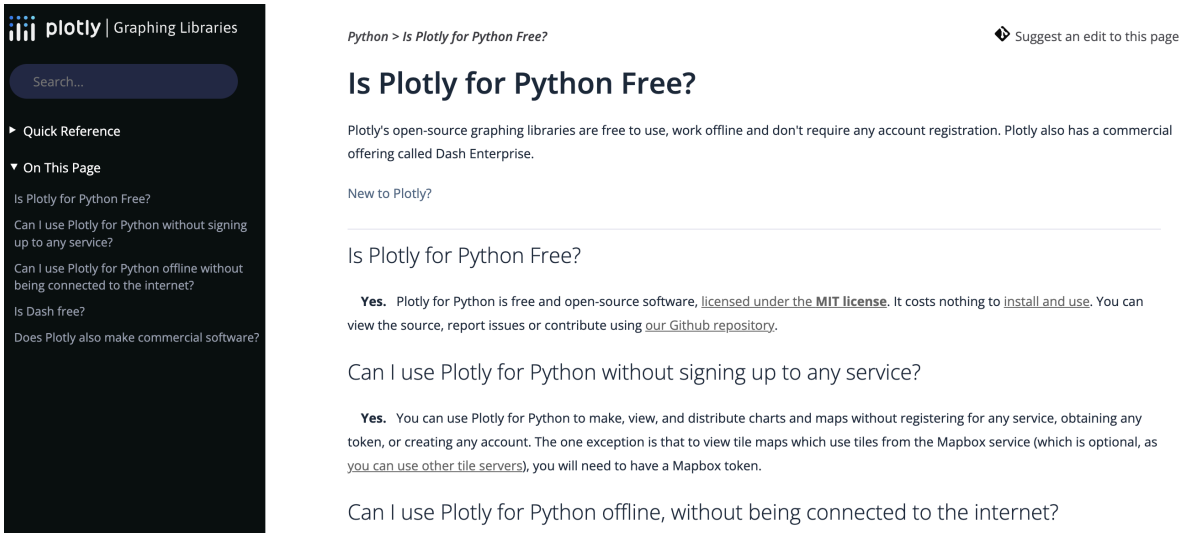
Les bases de Plotly

Introduction à Plotly

- Plotly est une bibliothèque de visualisation interactive en Python.
- Compatible avec pandas et d'autres outils analytiques.
- Prend en charge les graphiques statiques et interactifs.
- Fonctionne avec Jupyter Notebook et les applications web.

Plotly, c'est quoi ?

Un modèle économique autou de l'open source



Vue d'ensemble

Tutorial complet de Programming Historian

- Plusieurs "niveaux" dans Plotly :
  - de l'interface "haut-niveau" pour la visualisation (plotly express)
  - aux éléments "bas-niveau" pour triturer les figures
- *i.e.* : des submodules :
  - plotly.express
  - plotly.graph_objects
  - ...
- Gallerie

Installation et configuration

```
pip install plotly
```

- Importation :

```
import plotly.express as px
import plotly.graph_objects as go
```

- Compatible avec pandas pour la manipulation des données.

Graphiques avec Plotly Express

- Facile à utiliser pour des visualisations rapides.
- Exemple de scatter plot :

```
import plotly.express as px
df = px.data.iris()
fig = px.scatter(df, x='sepal_width', y='sepal_length', color='species')
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

Graphiques personnalisés avec Graph Objects

- Permet un contrôle plus précis.

```
import plotly.graph_objects as go
fig = go.Figure(data=[go.Scatter(x=[1,2,3], y=[4,5,6], mode='lines+markers')])
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

- Utilisation de plusieurs traces.

Ajout d'interactivité

- Zoom, hover, sélection dynamique.
- Ajout de sliders et boutons pour modifier les graphiques en temps réel.
- Exemple :

```
fig.update_layout(updatemenus=[
    dict(type='buttons',
         buttons=[dict(label='Afficher', method='update', args=[{'visible': True}])])
])
```

```
Unable to display output for mime type(s): text/html
```

Visualisation de données géographiques

- Exemples avec `px.choropleth` et `px.scatter_geo`.

```
df = px.data.gapminder()
fig = px.scatter_geo(df, locations="iso_alpha", size="pop", color="continent")
fig.show()
df.head()
```

```
Unable to display output for mime type(s): text/html
```

	country	continent	year	lifeExp	pop	gdpPercap	iso_alpha	iso_num
0	Afghanistan	Asia	1952	28.801	8425333	779.445314	AFG	4
1	Afghanistan	Asia	1957	30.332	9240934	820.853030	AFG	4
2	Afghanistan	Asia	1962	31.997	10267083	853.100710	AFG	4
3	Afghanistan	Asia	1967	34.020	11537966	836.197138	AFG	4
4	Afghanistan	Asia	1972	36.088	13079460	739.981106	AFG	4

Personnalisation et mise en forme

- Changer les couleurs, les étiquettes et les annotations.
- Exemple de mise en forme :

```
fig.update_layout(title='Mon Graphique', xaxis_title='Axe X', yaxis_title='Axe Y')
```

Exportation et partage

- Exporter en HTML, PNG, PDF.

```
fig.write_html("graph.html")
fig.write_image("graph.png")
```

- Intégration dans des sites web et notebooks Jupyter.

Learning by doing

Plus simple à apprendre en pratique.

Proposez une visualisation de résumé de votre corpus de données.

- Réfléchissez à ce que vous représentez
- Choix des couleurs
- Ajout d'interactivité & labels