Atelier Visualisation de données (DataViz)

Boris Mericskay

▶ **To cite this version:**

Boris Mericskay. Atelier Visualisation de données (DataViz). Maitrise. France. 2019. cel-02000916

HAL Id: cel-02000916
https://shs.hal.science/cel-02000916

Submitted on 31 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Atelier

Visualisation de données

Janvier 2019 / Master Géomatique

UNIVERSITÉ de Cergy-Pontoise

UNIVERSITÉ RENNES 2

@Boris Mericskay

CC BY SA

Objectifs

Sensibilisation aux enjeux relatifs à la visualisation de données

- Pourquoi visualiser des données ?
- Comment visualiser des données ?
- Quelles complémentarités avec la cartographie ?
- Quelles types de visualisation pour quelles données ?
- Comment diffuser ses visualisations de données ?

Introduction à diverses solutions en ligne « gratuites »

- Solutions simples d'utilisation et accessibles
- Emphase sur la dimension design
- Prise en compte de la dimension interactive des visualisation de données en ligne

Programme de l'atelier

Mardi après-midi

- Introduction à la visualisation de données (2h)
- Infographie des chiffres clef de l'UCP avec infogr.am (1h)
- Visualisation de données avec Datamatic (1h)

Mercredi matin

- Visualisations poussées avec RawsGraph et Plotly
- Aller plus loin avec Flourish
- Géodatavisualisation avec KeplerGL (1h)

Mercredi après-midi

- Projet Portfolio Web (2h)

La boite à outils de l'atelier

infogr.am

RAWGraphs

The missing link between spreadsheets and data visualization.

Flourish

plotly

datamatic

kepler.gl

En complément

Jeux de données

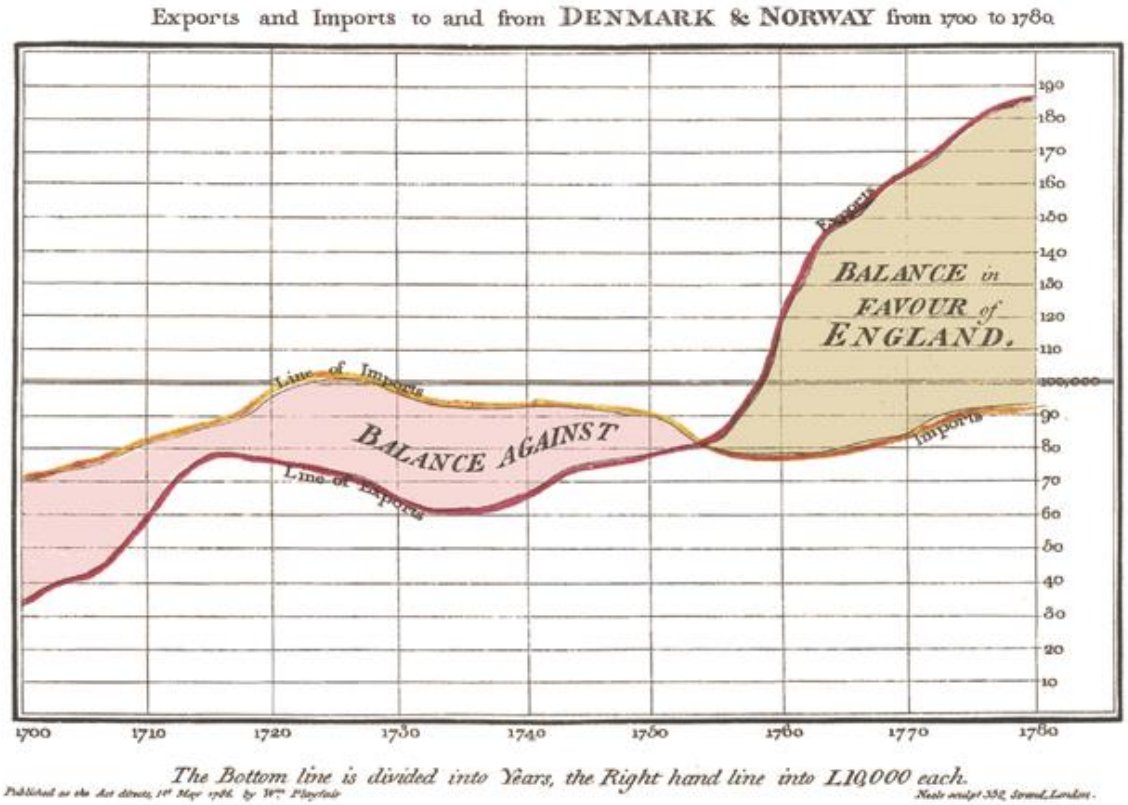
https://www.dropbox.com/s/inpxmypircixf47/datasets.zip?dl=0

- Base Insee Commune
- Communes et distances
- Budget participatif de Paris
- Catalogue de data.gouv
- Logements sociaux de Paris
- Prénoms à Rennes
- Production cinématographiques depuis 2003
- Tournage de films à Paris
- Fréquentation du site Web de Rennes métropole
- ZAC de Rennes métropole
- Trajets de BlaBlaCar
- Origines CSP des bacheliers

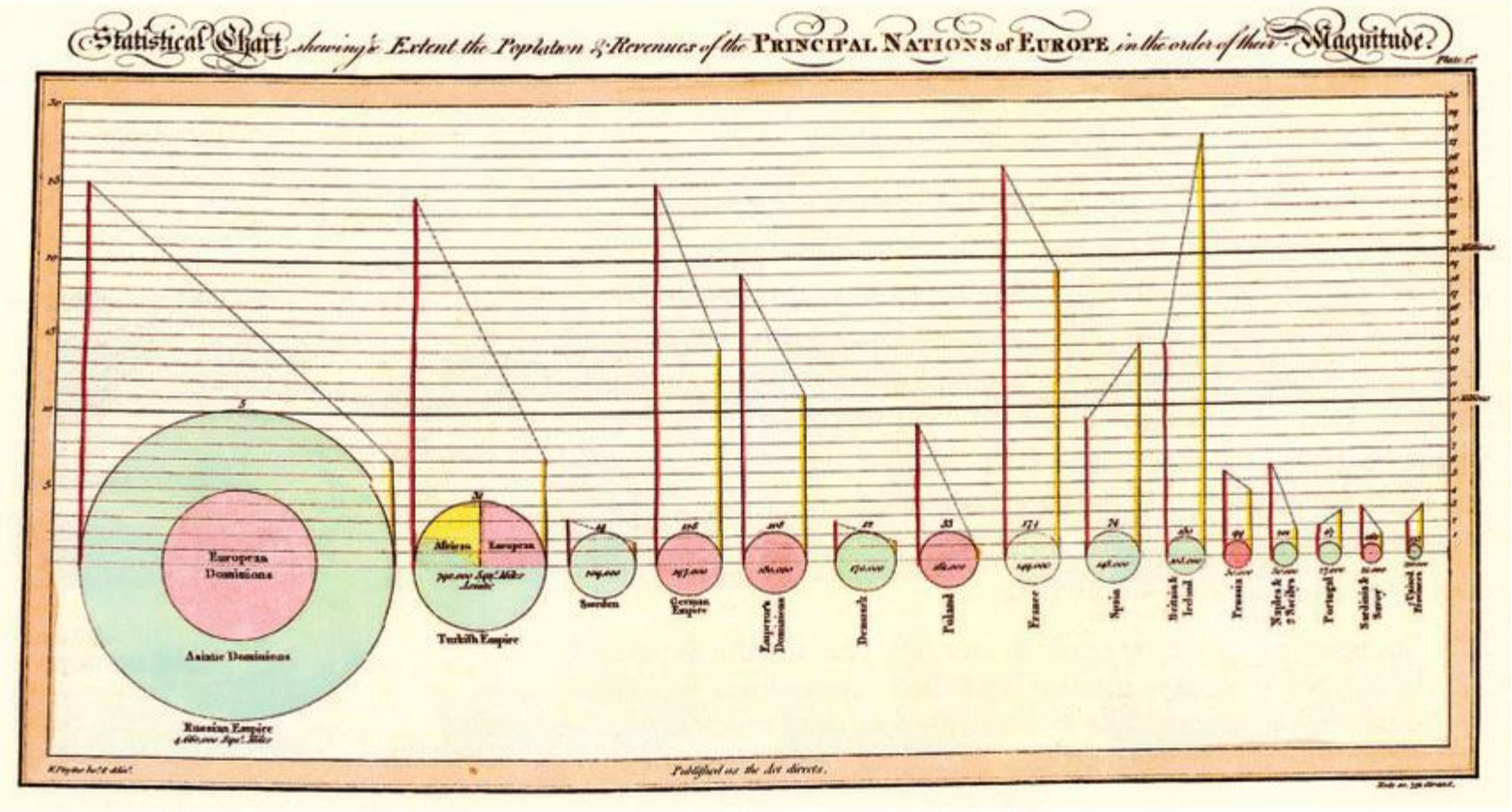
RETOUR AUX SOURCES DE LA VISUALISATION DE DONNEES

Historique

William Playfair (ingénieur et économiste écossais / 1759-1823) considéré comme l'inventeur des graphiques pour représenter les données : line plots, bar chart and pie chart

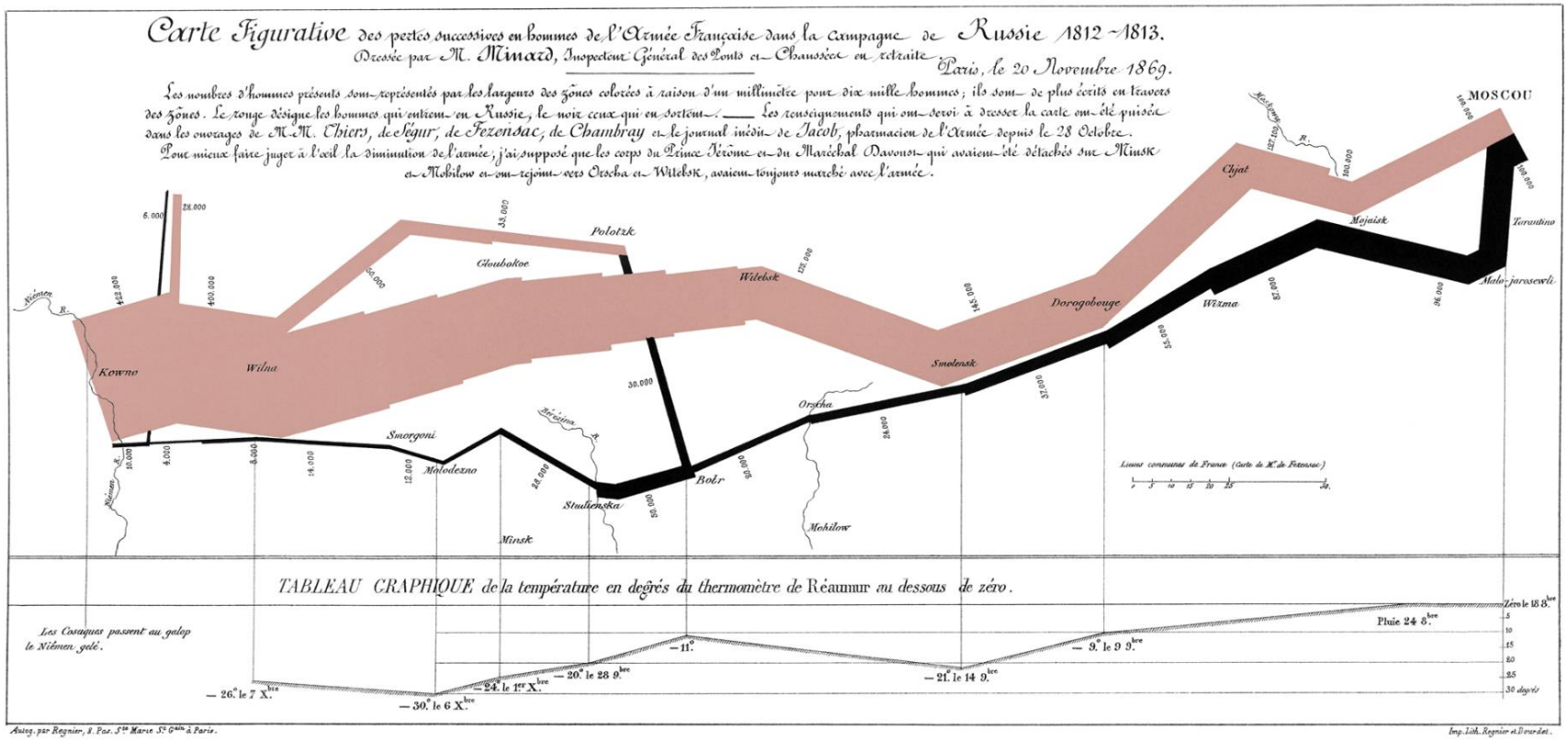


Historique

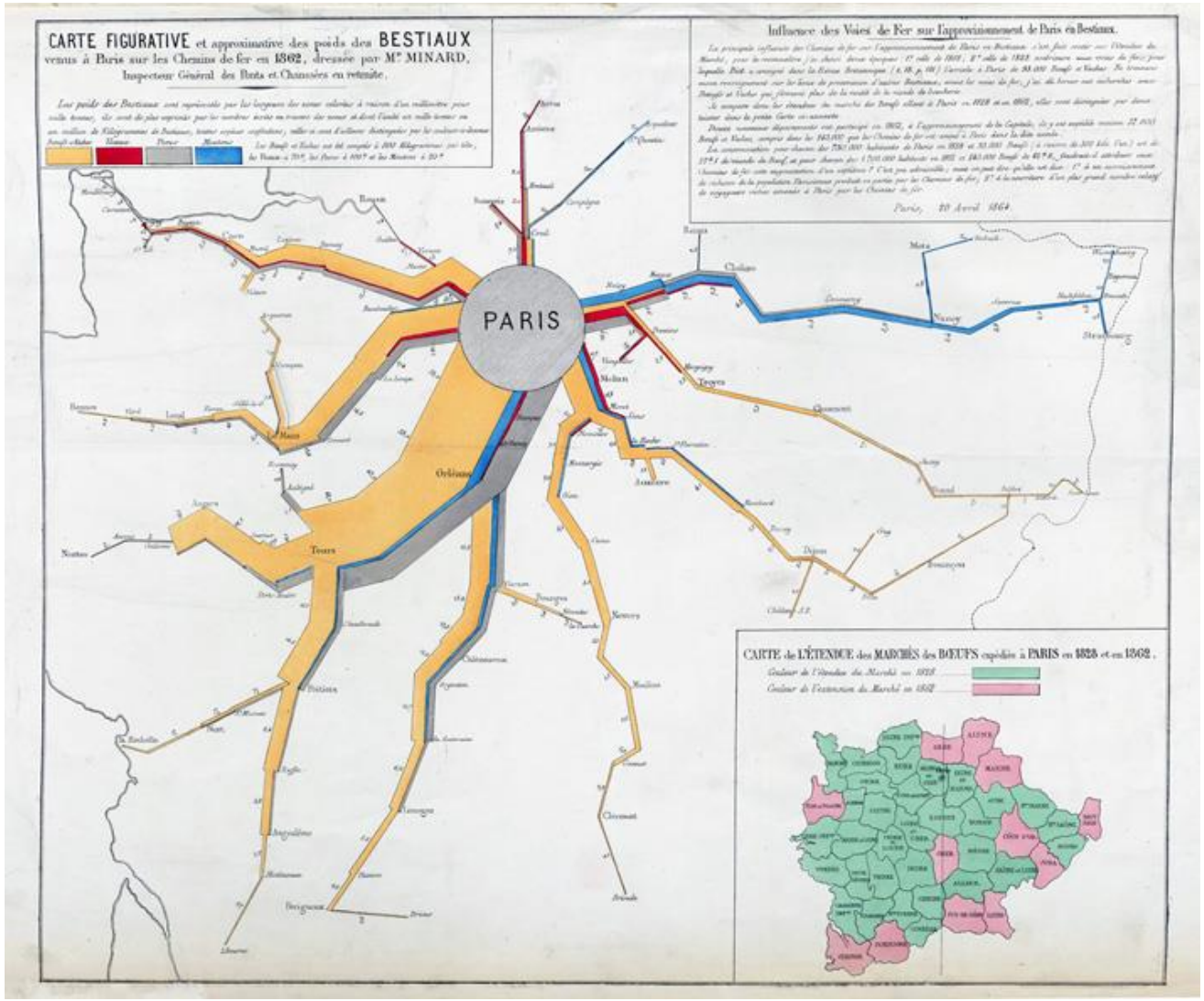

Statistical Chart shewing the Extent the Population & Revenues of the PRINCIPAL NATIONS of EUROPE in the order of their Magnitude.
European Dominions
Asiatic Dominions
Russian Empire
Turkish Empire
Sweden
German Empire
Emperor's Dominions
Denmark
Poland
France
Spain
Britain & Ireland
Portugal
Published as the Act directs.

Historique

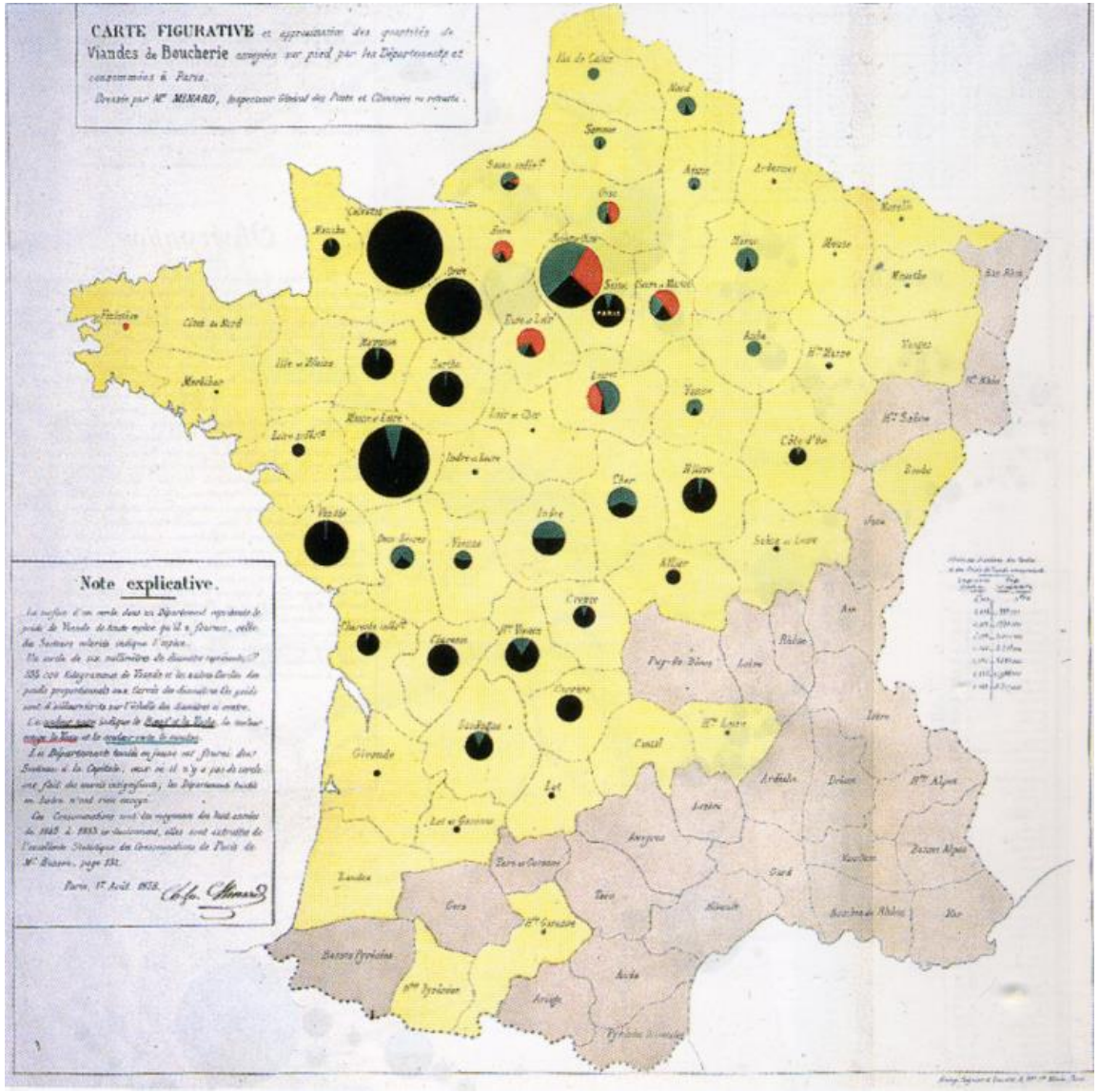
Charles Minard (ingénieur civil français / 1781-1870) a fortement contribué au domaine de l'information graphique et statistique en particulier avec l'utilisation de cartes.



Historique

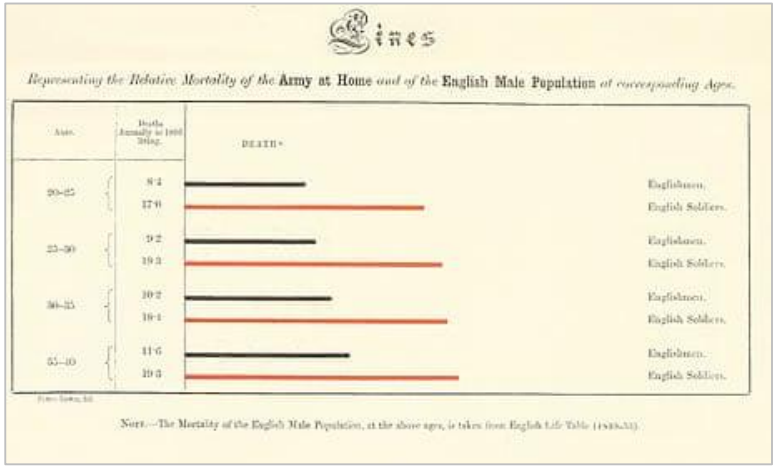

CARTE FIGURATIVE et approximative des poids des BESTIAUX venus à Paris sur les Chemins de fer en 1862, dressée par Mr MINARD, Inspecteur Général des Ponts et Chaussées en retraite.
Influence des Voies de Fer sur l'approvisionnement de Paris en Bestiaux.
PARIS
CARTE de L'ÉTENDUE des MARCHÉS des BŒUFS expédiés à PARIS en 1828 et en 1862.

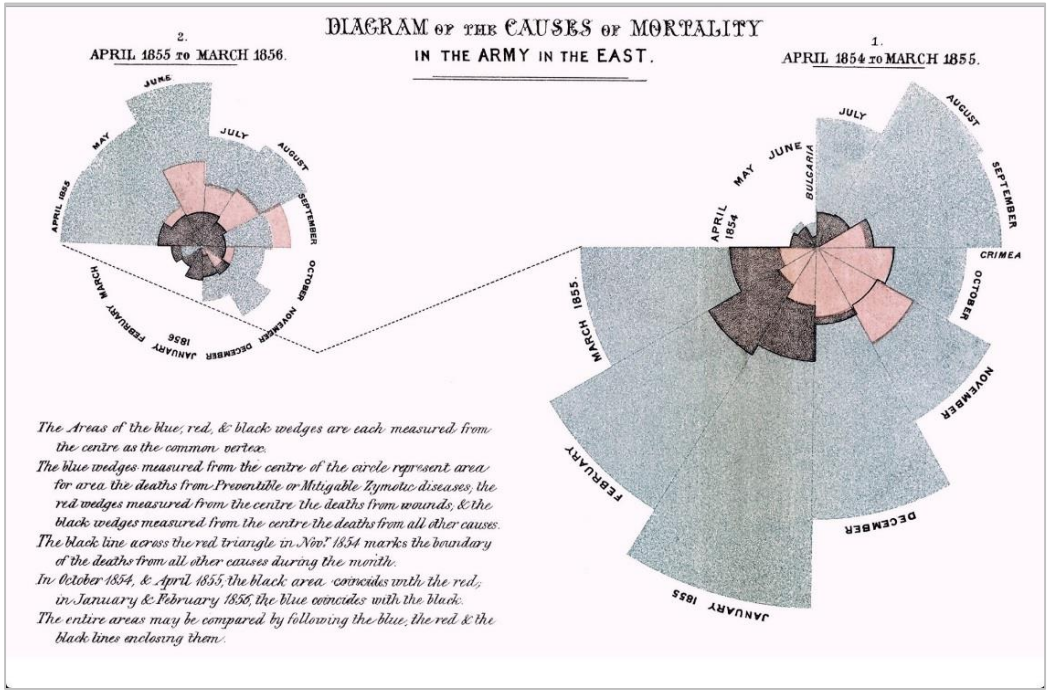
Historique


CARTE FIGURATIVE
Viandes de Boucherie
Note explicative.
PARIS

Historique

Florence Nightingale (infirmière britannique / 1820-1910) est connue comme étant la “mère” de la profession d’infirmière moderne. Elle a aussi contribué aux représentations graphiques.





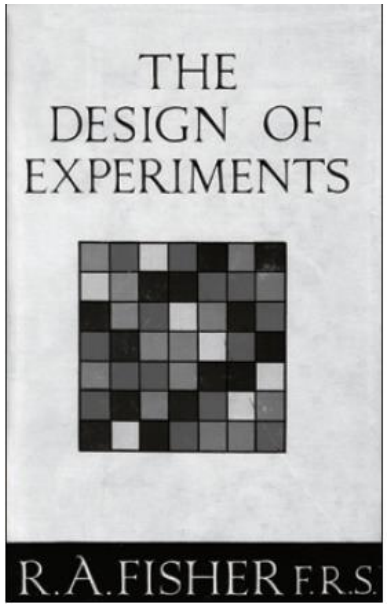
Historique

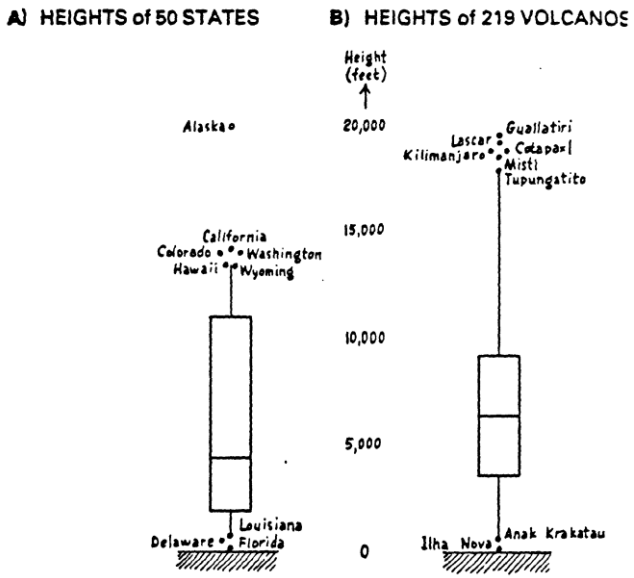
John Snow (médecin britannique / 1813–1858) connu pour avoir tracé (détecté) les sources de Choléra à Londres.

Historique

Ronald Fisher (1890-1962) et John Tukey (1915-2000) :méthodes graphiques avancées pour l'analyse des données.

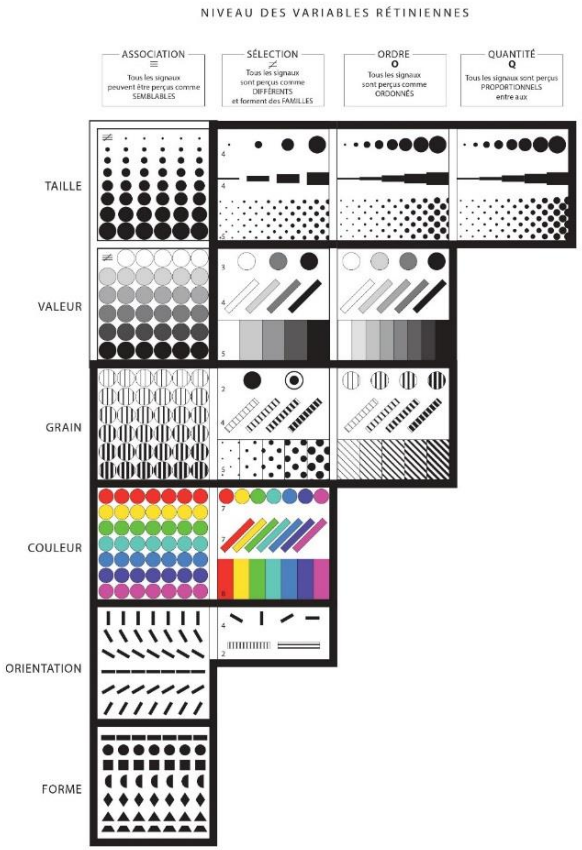
- Fisher : dessin des données pour comprendre les relations
- Tukey : promotion de l'analyse de données exploratoires, il a créé en particulier le box plot, le stem plot et le leaf plot.



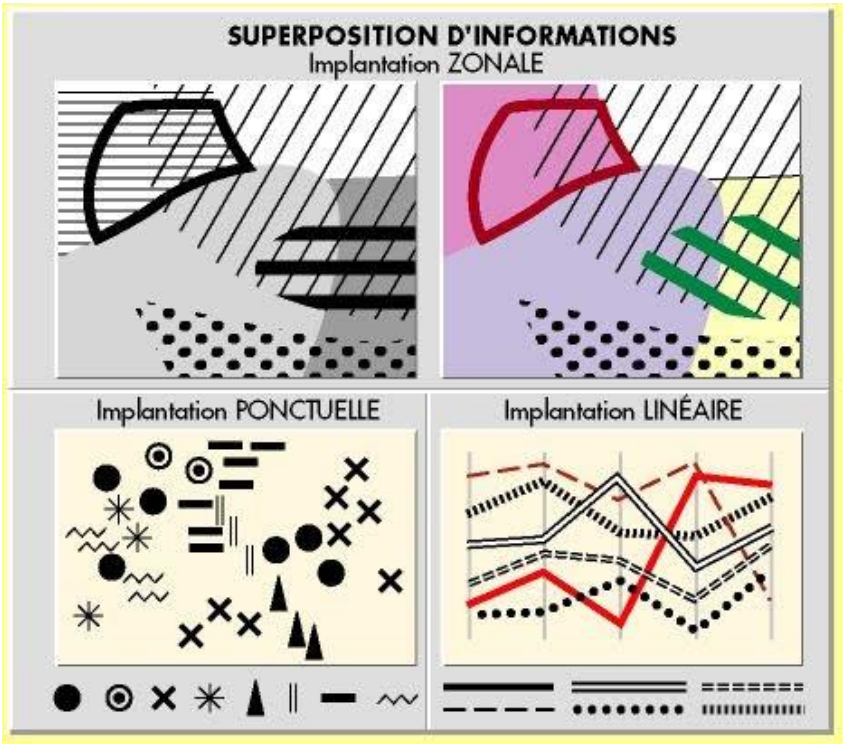


Historique

Jacques Bertin (1918-2010): *sémiologie graphique* (Système de signes pour la transmission de l'information)



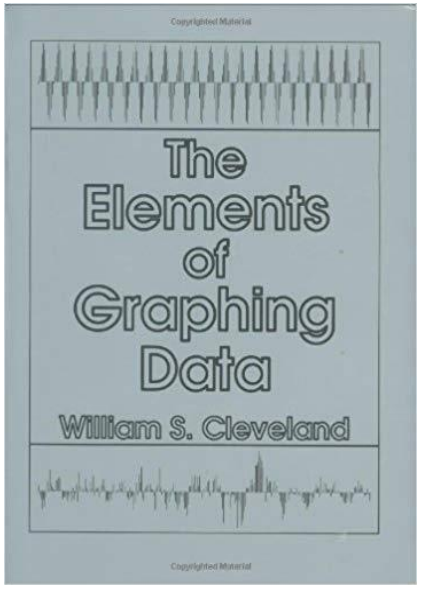
Jacques Bertin, "Sémiologie Graphique", 1973.

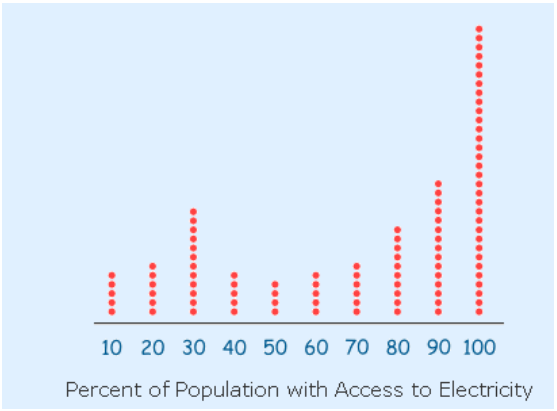


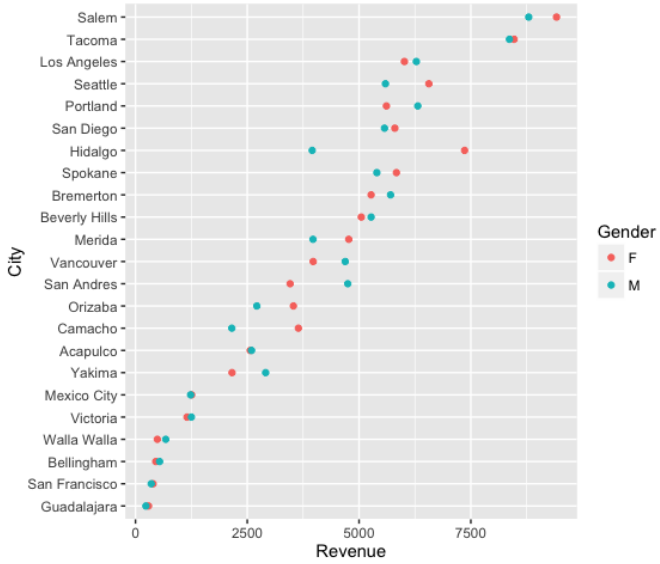
Historique

William Cleveland's (informaticien) connu pour promouvoir le dot plot comme alternative aux barres, diagrammes circulaires...

- Le dot plot permet une certaine clarté et une facilité de comparaison des données



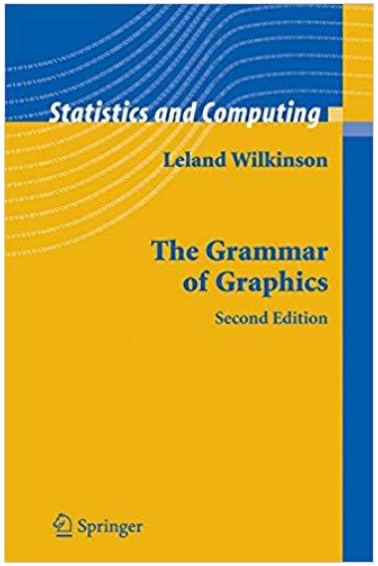


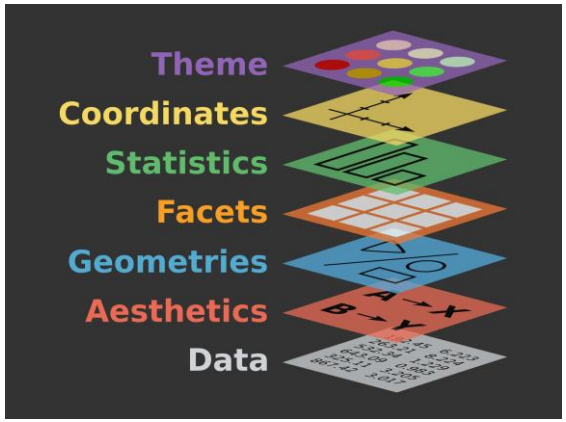


Historique

The Grammar of Graphics de Leland Wilkinson (Statisticien), a eu une influence importante sur la façon de penser les graphes

- Grammaire = règles mathématiques et esthétiques
- “Avant” on se focalisait surtout sur le coté esthétique d’un contenu statique
- L’idée générale est de **décrire** et donc **produire** un **graphique** comme un **assemblage de couches**





Historique

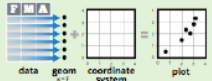
Les Graphiques avec ggplot2

Aide mémoire

RStudio

Les Bases

ggplot2 est basé sur "grammar of graphics", le principe est que vous pouvez construire tous les graphiques à partir d'un même petit nombre d'éléments : un **jeu de données**, un ensemble de **geoms** (repères visuels) qui représentent les points de données et un **système de coordonnées**.



Pour afficher les valeurs de données, il faut utiliser les variables du jeu de données en tant que propriétés esthétiques du geom dans size, color, x et y.

Les graphs se construisent avec **ggplot()** ou **qplot()**

propriétés esthétiques | données | geom

qplot(x = cty, y = hwy, color = cyl, data = mpg, geom = "point")
génère un graphique complet à partir des données, du geom et des propriétés esthétiques passés en paramètres et intègre de nombreux paramètres par défaut très utiles.

ggplot(data = mpg, **aes**(x = cty, y = hwy))
initialise un graphique à compléter en ajoutant des calques. Il n'y a pas de calques par défaut, mais cela permet plus de contrôle que qplot().

données

```
ggplot(mpg, aes(hwy, cty)) +
  geom_point(aes(color = cyl)) +
  geom_smooth(method = "lm") +
  coord_cartesian() +
  scale_color_gradient() +
  theme_bw()
```

ajout de calques avec+
calque = geom + stat par défaut + calque spécifique
éléments complémentaires

On ajoute un calque à un graphique avec une fonction **geom_*()** ou **stat_*()**. Chacun génère un geom, un ensemble de propriétés esthétiques et un claque statistique.

last_plot()
Renvoie le dernier graphique

ggsave("plot.png", width = 5, height = 5)
Sauvegarde dans l'espace de travail le dernier graphique affiché dans un fichier "plot.png" de dimension 5' x 5'. Le type de fichier généré dépend directement de l'extension de nom de fichier indiquée.

Geoms - Utiliser un geom pour représenter les points de données, utiliser les propriétés esthétiques du geom pour représenter des variables. Chaque fonction renvoie un calque.

Une variable

Continue

a <- ggplot(mpg, aes(hwy))

- a + **geom_area(stat = "bin")** — x, y, alpha, color, fill, linetype, size — b + geom_area(aes(y = ..density..), stat = "bin")
- a + **geom_density(**kernel = "gaussian"**)** — x, y, alpha, color, fill, linetype, size, weight — b + geom_density(aes(y = ..county..))
- a + **geom_dotplot()** — x, y, alpha, color, fill
- a + **geom_freqpoly()** — x, y, alpha, color, linetype, size — b + geom_freqpoly(aes(y = ..density..))
- a + **geom_histogram(**binwidth = 5**)** — x, y, alpha, color, fill, linetype, size, weight — b + geom_histogram(aes(y = ..density..))

Discrète

b <- ggplot(mpg, aes(fl))

- b + **geom_bar()** — x, alpha, color, fill, linetype, size, weight

Éléments graphiques

c <- ggplot(map, aes(long, lat))

- c + **geom_polygon(**aes(group = group)**)** — x, y, alpha, color, fill, linetype, size

d <- ggplot(economics, aes(date, unemploy))

- d + **geom_path(**lineend="butt", linejoin="round', linemitre=1**)** — x, y, alpha, color, linetype, size
- d + **geom_ribbon(**aes(ymin=unemploy 900, ymax=unemploy + 900)**)** — x, ymax, ymin, alpha, color, fill, linetype, size

e <- ggplot(seals, aes(x = long, y = lat))

- e + **geom_segment(**aes(xend = long + delta_long, yend = lat + delta_lat)) — x, xend, y, yend, alpha, color, linetype, size
- e + **geom_rect(**aes(xmin = long, ymin = lat, xmax= long + delta_long, ymax = lat + delta_lat)) — xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size

Deux variables

X Continue, Y Continue

f <- ggplot(mpg, aes(cty, hwy))

- f + **geom_blank()** (Utile pour étendre les limites)
- f + **geom_jitter()** — x, y, alpha, color, fill, shape, size
- f + **geom_point()** — x, y, alpha, color, fill, shape, size
- f + **geom_quantile()** — x, y, alpha, color, linetype, size, weight
- f + **geom_rug(**sides = "bl"**)** — alpha, color, linetype, size
- f + **geom_smooth(**model = lm**)** — x, y, alpha, color, fill, linetype, size, weight
- f + **geom_text(**aes(label = cty)) — x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

X Discrète, Y Continue

g <- ggplot(mpg, aes(class, hwy))

- g + **geom_bar(stat = "identity")** — x, y, alpha, color, fill, linetype, size, weight
- g + **geom_boxplot()** — lower, middle, upper, x, ymax, ymin, alpha, color, fill, linetype, shape, size, weight
- g + **geom_dotplot(**binaxis = "y", stackdir = "center"**)** — x, y, alpha, color, fill
- g + **geom_violin(**scale = "area"**)** — x, y, alpha, color, fill, linetype, size, weight

X Discrète, Y Discrète

h <- ggplot(diamonds, aes(cut, color))

- h + **geom_jitter()** — x, y, alpha, color, fill, shape, size

Distribution bivariée continue

i <- ggplot(movies, aes(year, rating))

- i + **geom_bin2d(**binwidth = c(5, 0.5)**)** — xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size, weight
- i + **geom_density2d()** — x, y, alpha, colour, linetype, size
- i + **geom_hex()** — x, y, alpha, colour, fill size

Fonction continue

j <- ggplot(economics, aes(date, unemploy))

- j + **geom_area()** — x, y, alpha, color, fill, linetype, size
- j + **geom_line()** — x, y, alpha, color, linetype, size
- j + **geom_step(**direction = "hv"**)** — x, y, alpha, color, linetype, size

Marge d'erreur

df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
k <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))

- k + **geom_crossbar(**fatten = 2**)** — x, y, ymax, ymin, alpha, color, fill, linetype, size
- k + **geom_errorbar()** — x, ymax, ymin, alpha, color, linetype, size, width (also **geom_errorbarh()**)
- k + **geom_linerange()** — x, ymin, ymax, alpha, color, linetype, size
- k + **geom_pointrange()** — x, y, ymin, ymax, alpha, color, fill, linetype, shape, size

Cartographie

data <- data.frame(meurtre= USArrests$Murder, etat = tolower(rownames(USArrests)))
map <- map_data("etat")
l <- ggplot(data, aes(fill = meurtre))

- l + **geom_map(**aes(map_id = etat), map = map**)** + **expand_limits(**x = map$long, y = map$lat**)** — map_id, alpha, color, fill, linetype, size

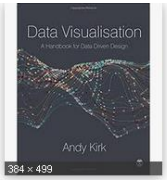
Trois Variables

seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2))
m <- ggplot(seals, aes(long, lat))

- m + **geom_contour(**aes(z = z)**)** — x, y, z, alpha, colour, linetype, size, weight
- m + **geom_raster(**aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE**)** — x, y, alpha, fill (rapide)
- m + **geom_tile(**aes(fill = z)**)** — x, y, alpha, color, fill, linetype, size (lent)

RStudio® is a trademark of RStudio, Inc. • CC BY RStudio • info@rstudio.com • 844 448 1212 • rstudio.com

Et aujourd'hui ?



Amazon.fr - Data Visualisa...
amazon.fr

Datavisualisation : 3 outils pour créer de belles info...
archimag.com

Datavisualisation : les logiciels en vogue - ww...
datavisualization.fr

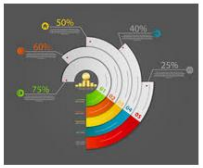
Data visualisation : mieux communiquer ...
http5000.com

12 solutions de datavisualisation
e-marketing.fr

Data visualisation charts ...
vectorstock.com

Data Visualization : Top des sta...
lebigdata.fr

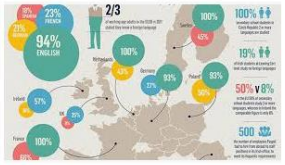
Data visualisation - ParlonsCrypto.fr
parlonscrypto.fr



Introduction aux librairies de data...
blog.equancy.com

Big Data Visualisation Colorée. Infographie Fut...
fr.123rf.com

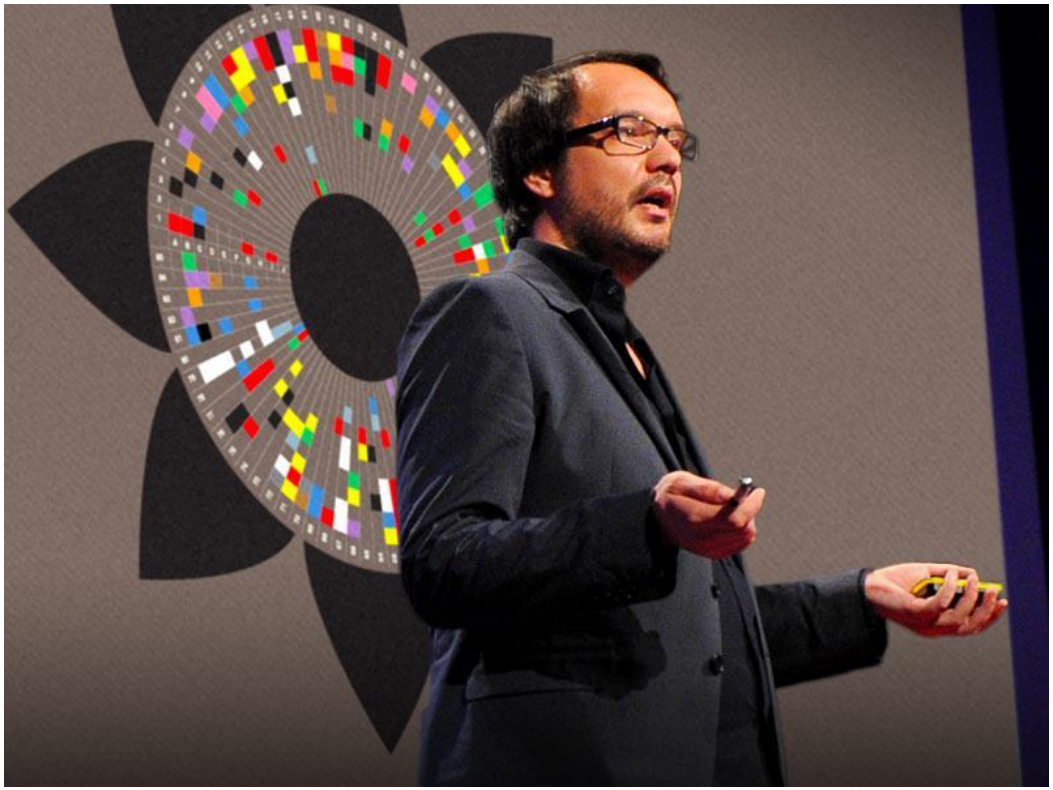
Data Visualization : top des meilleurs outils de visua...
lebigdata.fr

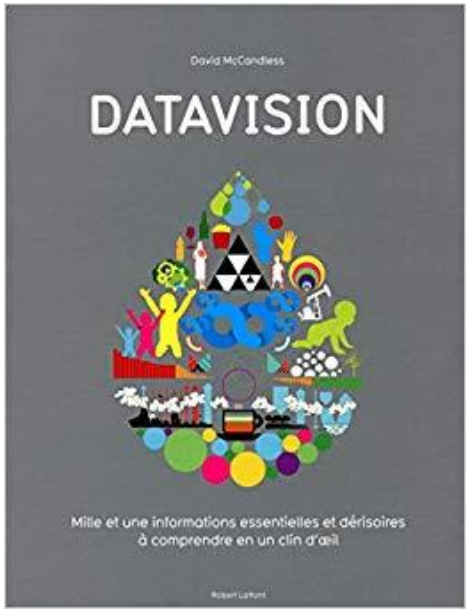


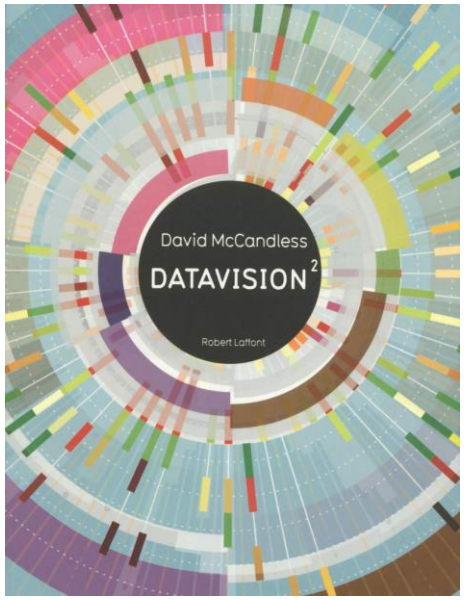
Data Visualisation Design &MK47 | Aieasyspain
aieasyspain.com

Mes astuces pour réussir la data-visualisation | | Hu...
news.humancoders.com

Et aujourd'hui ?

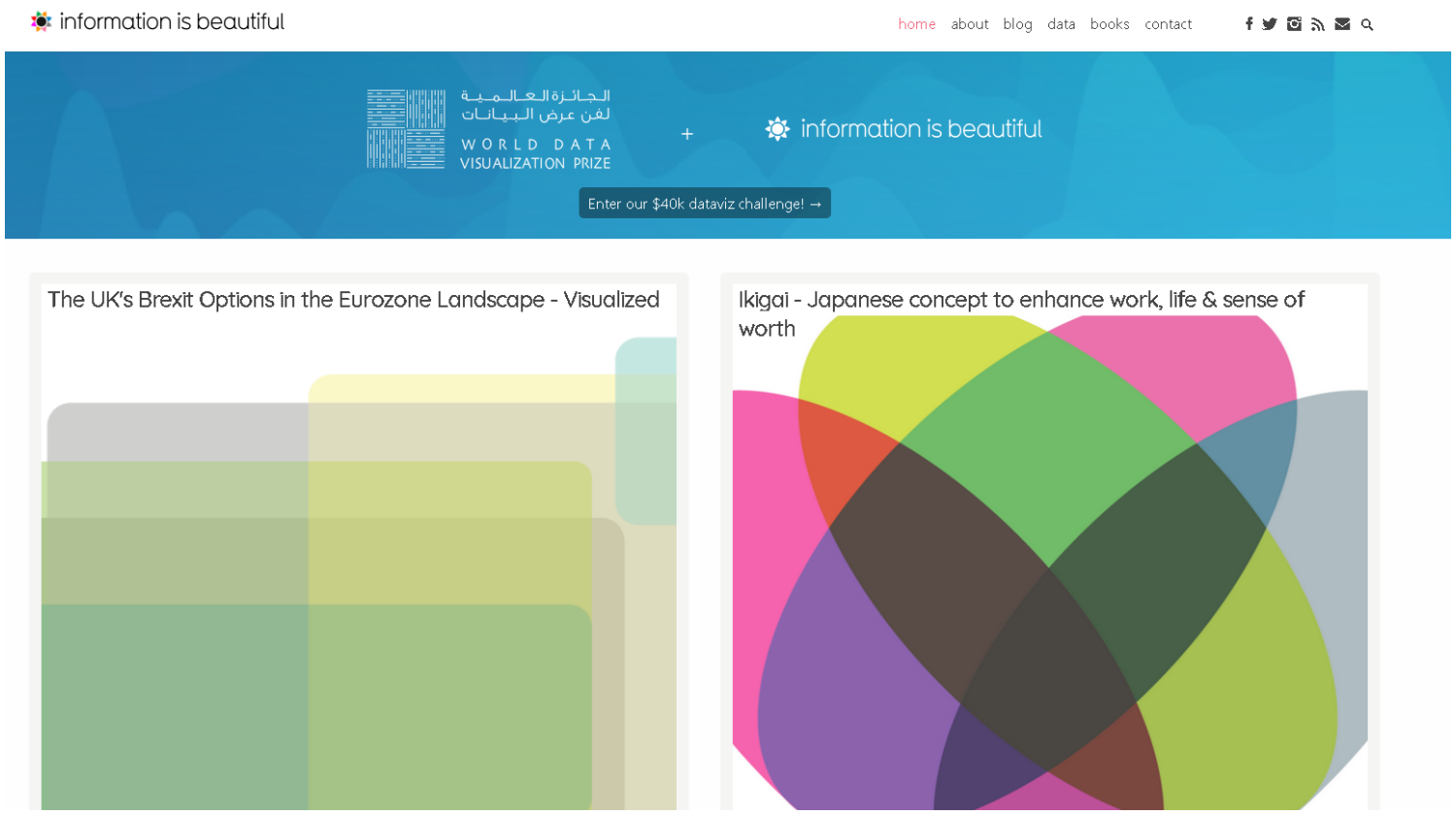






Et aujourd'hui ?

https://informationisbeautiful.net/



Et aujourd'hui ?

David Bihanic, designer-researcher

Design research
Datavis aesthetics
Creative informatics

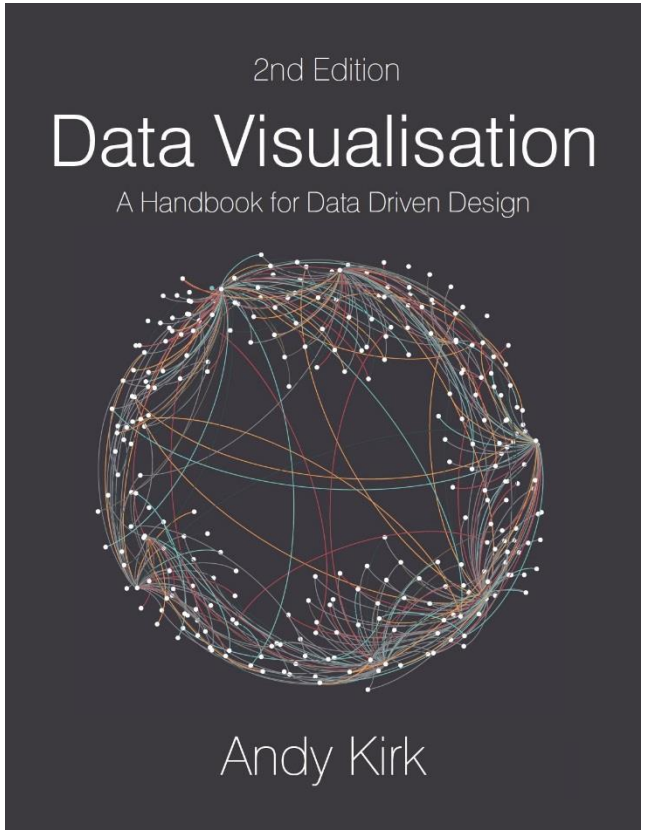
I'm an associate professor at the University of Paris I Pantheon-Sorbonne and the founder of FXDS, a creative agency.

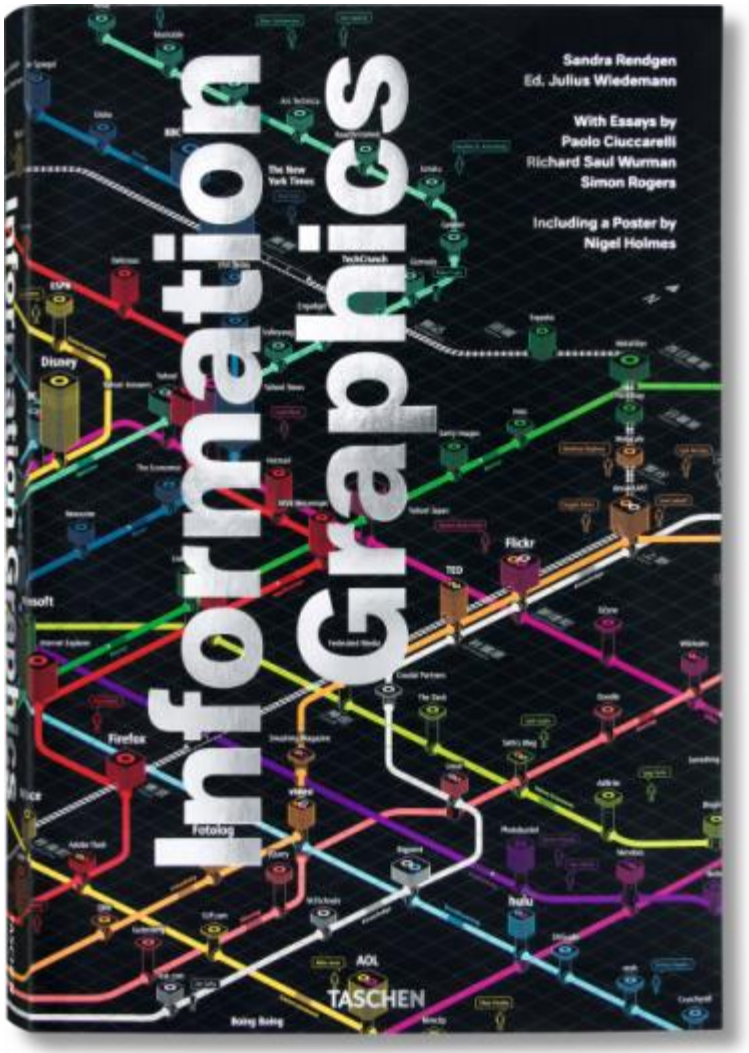
Main res areas

My research addresses theoretical and practical issues of design — most of my work seeks to identify the new challenges of the discipline considering, in particular, the linkages between design and engineering sciences. Specifically, I examine the formal and aesthetic stakes underlying data visualization, data design processes and methods, and groundbreaking software solutions.

http://www.davidbihanic.com/

Et aujourd'hui ?

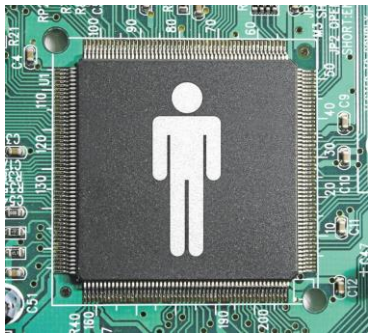




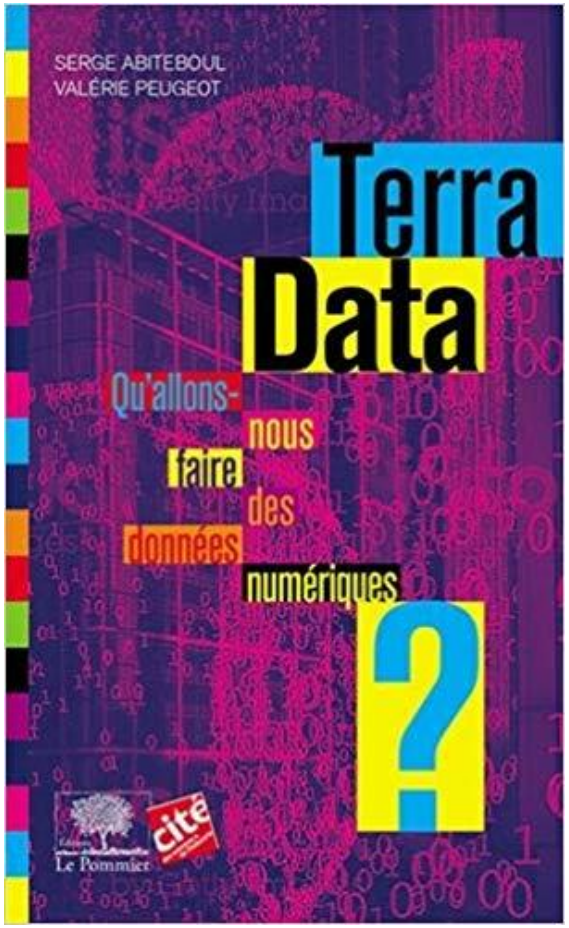
Mise en contexte

Datafication du monde

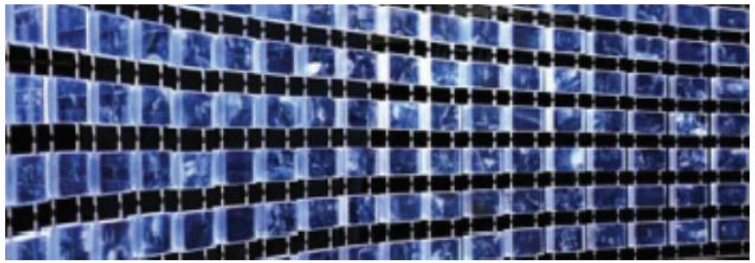
- Amitiés, pensées, échanges, déplacements : la plupart des activités humaines donnent désormais lieu à une production massive de données numérisées
  - Leur collecte et leur analyse ouvrent des perspectives parfois enthousiasmantes qui aiguisent l'appétit des entreprises
  - Mais la mise en données du monde risque aussi de menacer les libertés, comme le montre les programmes de surveillance
- En deux décennies, la donnée s'est imposée comme un des outil majeur dans la gestion des sociétés humaines Travail, mobilité, consommation, loisirs et même citoyenneté.



La mise en données du monde



MISE EN DONNEES DU MONDE :
IMAGINAIRES EN ÉQUATION



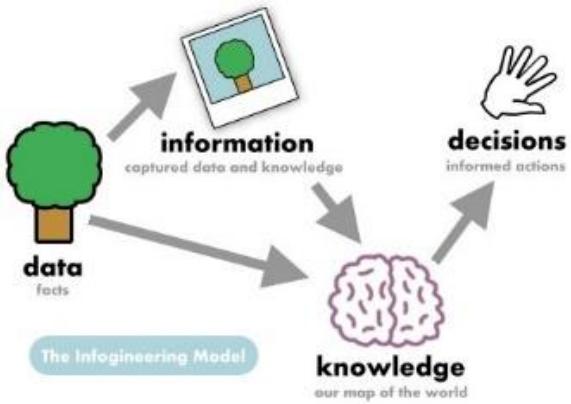
Gouverner par la donnée :
entre transparence et résistance,
quelle prospective ?

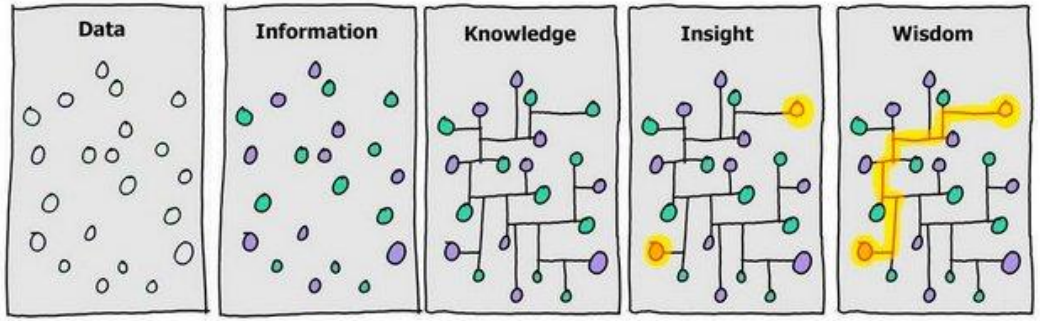
pop-up urbain
Philippe Gargov, Margot Baldassi
et Chloé Rotrou
Avril 2017

GRANDLYON
la métropole

Un nouveau paradigme informationnel

- La **donnée** est une mesure (pas de sens intrasèque)
- **Information** : donnée à laquelle on donne un sens, une interprétation dans un contexte défini
- **Savoir** : ensemble d'informations connues et interprétées dans la durée permettant de dégager un principe, une loi





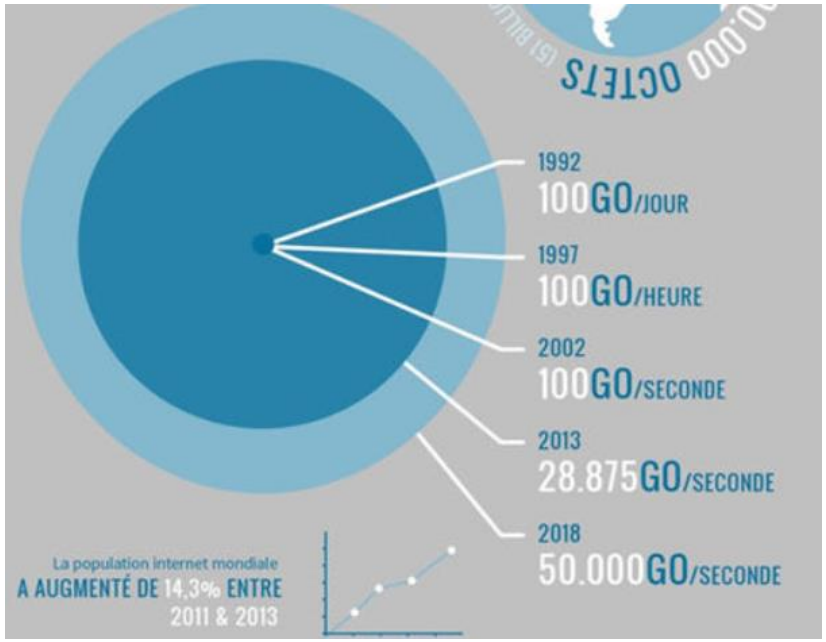
Un nouveau paradigme informationnel

- **L'aide à la décision a pour objectif d'accompagner les décideurs dans les processus de gestion / aménagement**
  - La gestion et la maîtrise du développement territorial impose le recours à une vision globale et objective des caractéristiques et des dynamiques d'un espace
  - La matière première de l'aide à la décision est l'information

    → Connaitre pour agir

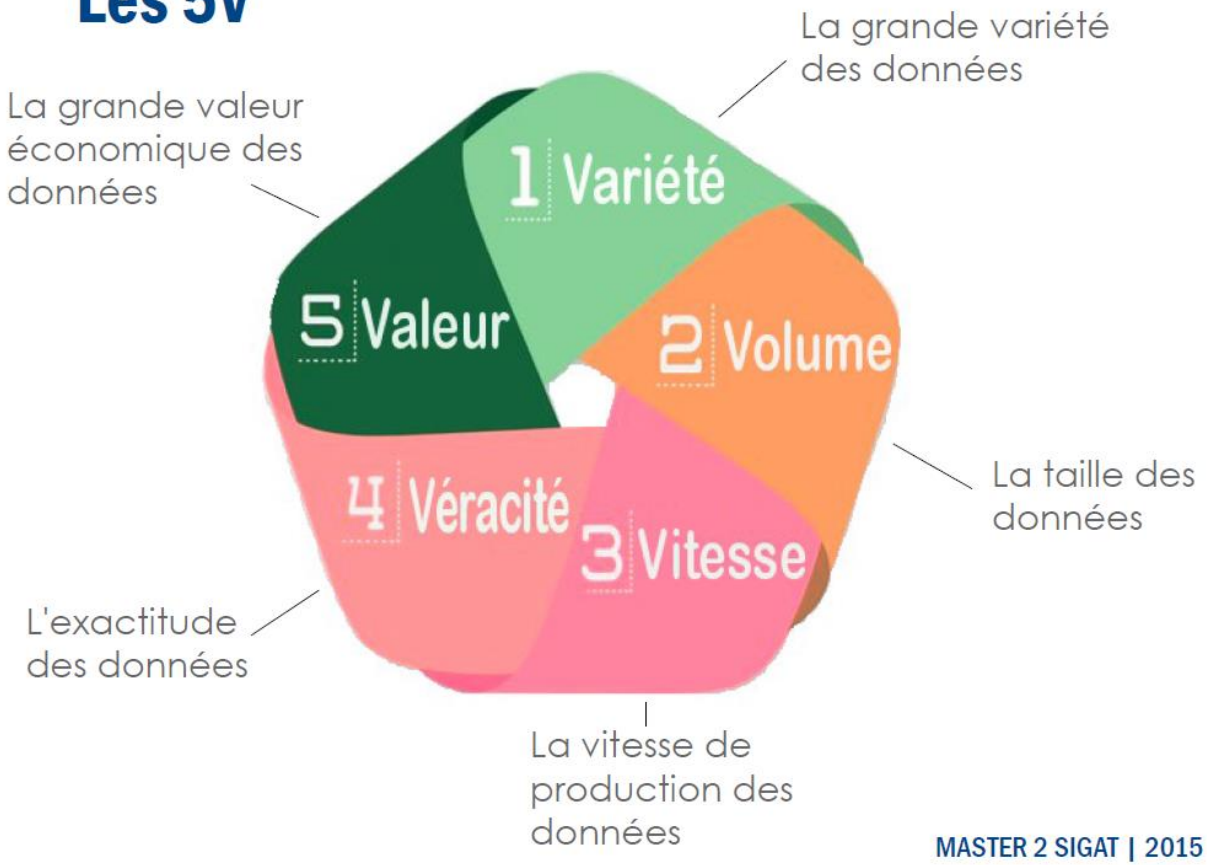
BIG DATA

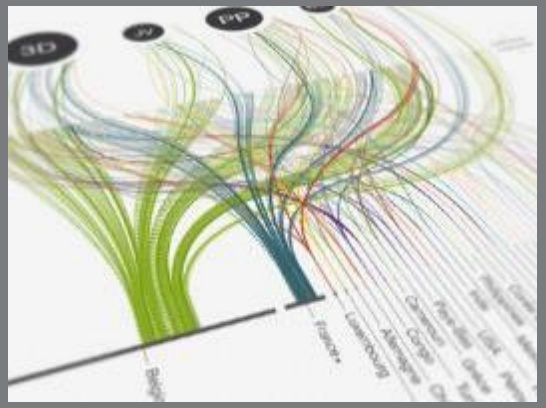
- **Un monde de données**
  - La quantité de données exploitables atteint aujourd'hui un volume inédit, qui continue de croître de jour en jour.



BIG DATA

INTRODUCTION
Les 5V





Des données à la visualisation

Big data

Pour résumer :

- Des données exponentielles :
  - Variées
  - Volumineuses
  - En temps réel
  - De sources très hétérogènes
  - Mais qui encore « indomptées »
  - Encore que peu structurées

Mais que faire avec ces données ?

BIG DATA

- 3 discipline dans l'analyse des données

L'analyse des données pour tous

- Outils dédiés à l'analyse des données
- Facilité d'utilisation pour tous les utilisateurs de la donnée
- Fonctionnalités de Dataviz

Apporter de l'intelligence aux données

- Outils issus du monde du Datamining
- Dédiés à des utilisateurs statisticiens / data-Scientists
- Apport d'une très forte valeur ajoutée sur les données

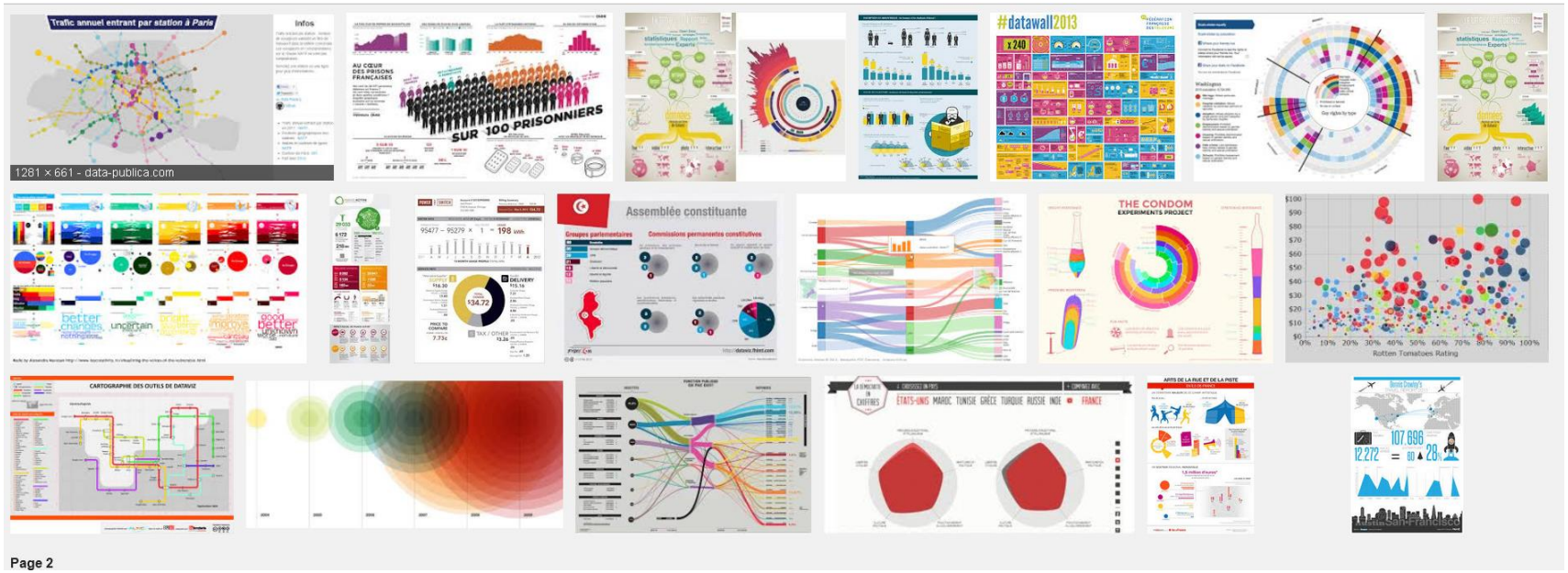
Présenter les données pour mieux les partager

- Capacités graphiques avancées et innovantes
- Support du Data-Storytelling ou des infographies
- Présentation dynamique ou statique

@Gaultier 2015

DataViz

- La dataviz (Data Visualisation) : désigne les techniques permettant de présenter des données sous forme visuelle afin d'en faciliter la compréhension et/ou l'analyse.



L’ère de la visualisation

Validité des données

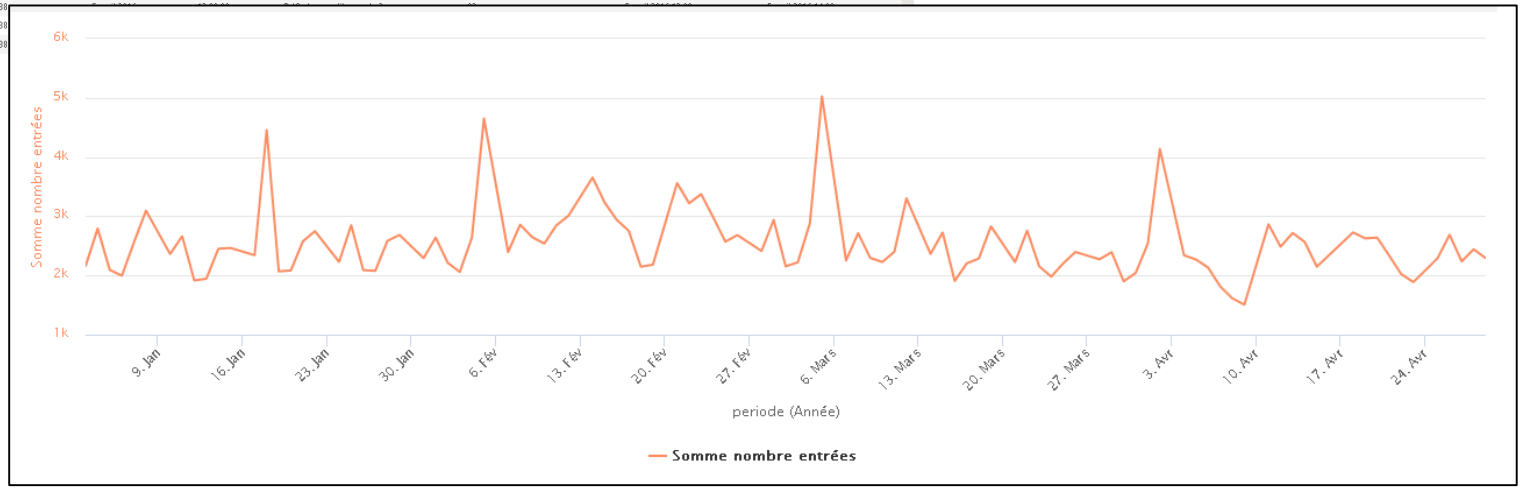
Des données bien représentées

Visualisation de données

Fréquentation par heure des Champs Libres

Informations | Tableau | Analyse | Calendrier | Export | API | Commentaires (0)

	site	code insee	periode	heure	zone	nombre entrées	datetime	datetimeend
1	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	12:00:00	RdC champs libres, rdc 2	1	1 avril 2016 12:00	1 avril 2016 14:00
2	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	12:00:00	RdC champs libres, rdc retour	13	1 avril 2016 12:00	1 avril 2016 14:00
3	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	14:00:00	RdC vie du citoyen	43	1 avril 2016 14:00	1 avril 2016 16:00
4	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	15:00:00	RdC champs libres, rdc retour	15	1 avril 2016 15:00	1 avril 2016 17:00
5	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	16:00:00	RdC champs libres, rdc 4	53	1 avril 2016 16:00	1 avril 2016 18:00
6	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	17:00:00	RdC champs libres, rdc 2	5	1 avril 2016 17:00	1 avril 2016 19:00
7	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	17:00:00	RdC champs libres, rdc 3	178	1 avril 2016 17:00	1 avril 2016 19:00
8	Bibliothèques Rennes les Champs libres	35238	1 avril 2016	17:00:00	RdC champs libres, rdc retour	22	1 avril 2016 17:00	1 avril 2016 19:00
9	Bibliothèques Rennes les Champs libres	35238	2 avril 2016	14:00:00	1er etage vie du citoyen	15	2 avril 2016 14:00	2 avril 2016 16:00
10	Bibliothèques Rennes les Champs libres	35238	2 avril 2016	14:00:00	RdC champs libres, rdc 3	582	2 avril 2016 14:00	2 avril 2016 16:00
11	Bibliothèques Rennes les Champs libres	35238	2 avril 2016	16:00:00	RdC champs libres, rdc 2	27	2 avril 2016 16:00	2 avril 2016 18:00
12	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	15:00:00	1er etage vie du citoyen	18	3 avril 2016 15:00	3 avril 2016 17:00
13	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	15:00:00	RdC champs libres, rdc retour	44	3 avril 2016 15:00	3 avril 2016 17:00
14	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	16:00:00	1er etage vie du citoyen	28	3 avril 2016 16:00	3 avril 2016 18:00
15	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	16:00:00	RdC champs libres, rdc 3	694	3 avril 2016 16:00	3 avril 2016 18:00
16	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	16:00:00	RdC champs libres, rdc retour	37	3 avril 2016 16:00	3 avril 2016 18:00
17	Bibliothèques Rennes les Champs libres	35238	3 avril 2016	16:00:00	RdC vie du citoyen	112	3 avril 2016 16:00	3 avril 2016 18:00
18	Bibliothèques Rennes les Champs libres	35238	5 avril 2016	10:00:00	RdC champs libres, rdc retour	5	5 avril 2016 10:00	5 avril 2016 12:00
19	Bibliothèques Rennes les Champs libres	35238	5 avril 2016	12:00:00	1er etage vie du citoyen	7	5 avril 2016 12:00	5 avril 2016 14:00
20	Bibliothèques Rennes les Champs libres	35238						
21	Bibliothèques Rennes les Champs libres	35238						
22	Bibliothèques Rennes les Champs libres	35238						

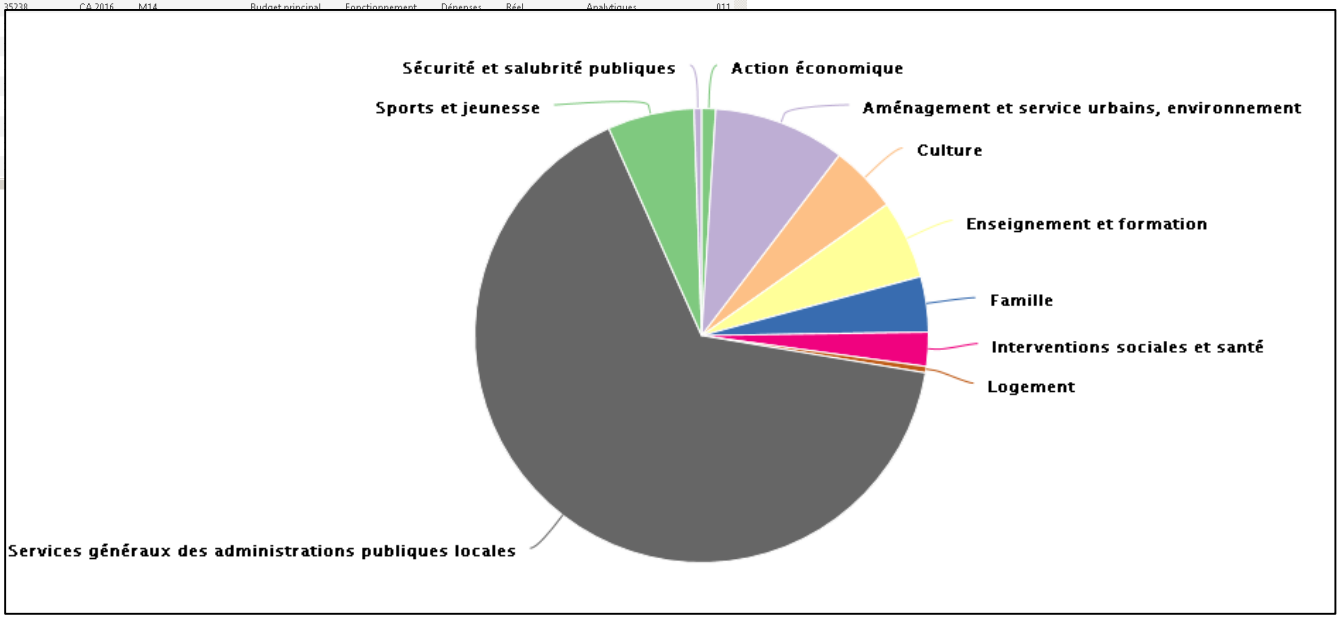


Visualisation de données

CA 2016 - Ville de Rennes - Budget Principal

Informations | Tableau | Analyse | Export | API | Commentaires (0)

	Exercice	Collectivité	Nom Collectivité	Code INSEE	Etape	Type nomenclature	Budget	Section	Sens	Mouvement	Budgétaire / analytique	Chapitr
1	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Ordre	Budgétaires	042
2	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
3	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
4	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
5	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
6	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
7	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
8	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
9	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
10	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
11	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
12	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
13	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
14	2 016	Commune	Ville de Rennes	35238	CA 2016	M14	Budget principal	Fonctionnement	Dépenses	Réel	Analytiques	011
15	2 016	Commune	Ville de Rennes									
16	2 016	Commune	Ville de Rennes									
17	2 016	Commune	Ville de Rennes									
18	2 016	Commune	Ville de Rennes									
19	2 016	Commune	Ville de Rennes									
20	2 016	Commune	Ville de Rennes									
21	2 016	Commune	Ville de Rennes									
22	2 016	Commune	Ville de Rennes									

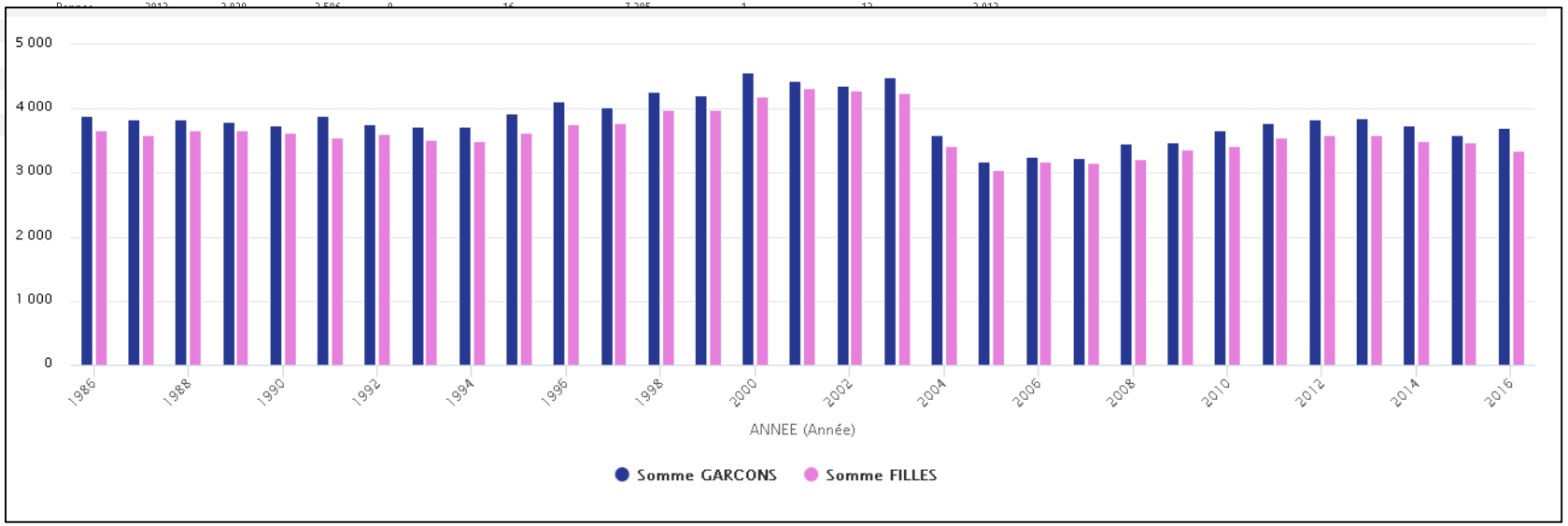


Visualisation de données

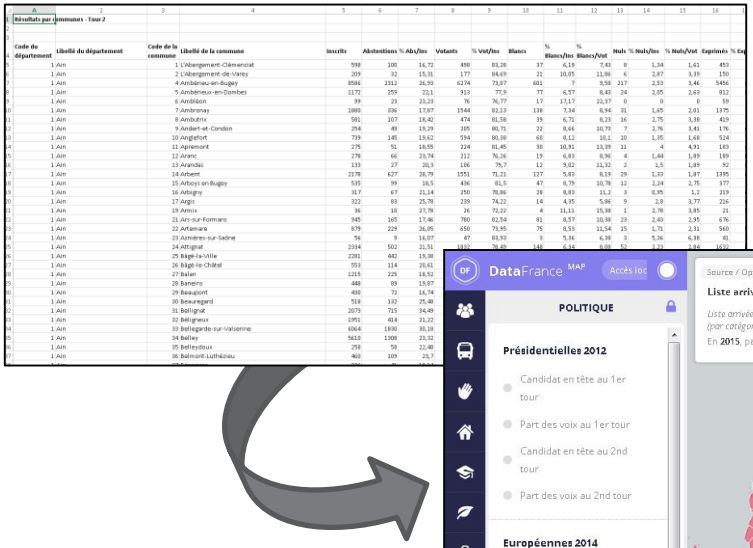
Naissances à Rennes

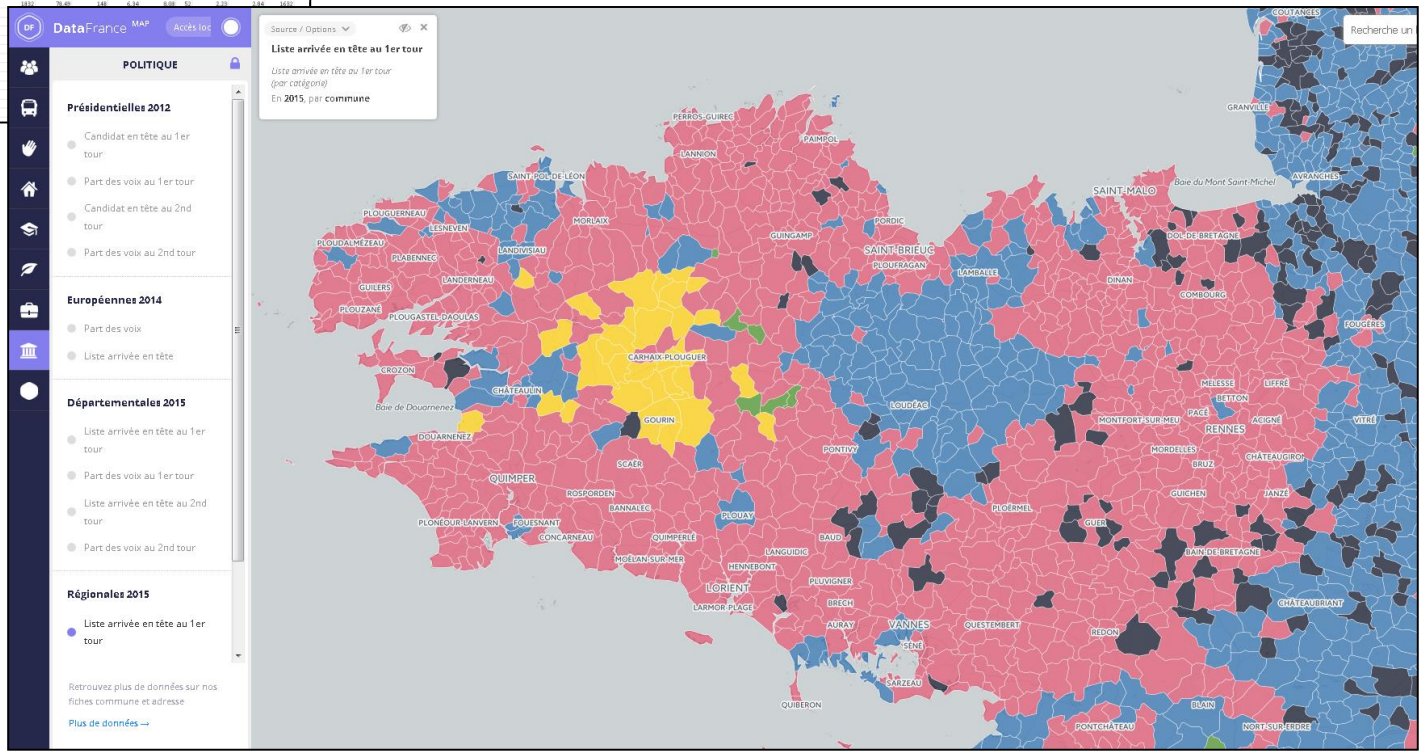
Informations | Tableau | Analyse | Export | API | Commentaires (0)

	CODE_INSEE	COMMUNE	ANNEE	GARCONS	FILLES	INDETERMINE	LIEU_DOMICILE	LIEU_HOPITAL	LIEU_CLINIQUE	LIEU_AUTRES	QJ_PARENTS_MARIES
1	35238	Rennes	1989	3 800	3 666	0	0	508	11	6 947	768
2	35238	Rennes	1992	3 765	3 610	0	0	463	14	6 898	908
3	35238	Rennes	1993	3 721	3 520	0	0	510	4	6 727	1 005
4	35238	Rennes	1995	3 922	3 620	0	0	586	8	6 948	1 162
5	35238	Rennes	1996	4 124	3 754	0	2	7 846	16	14	6 241
6	35238	Rennes	1998	4 263	3 990	0	2	8 236	1	14	6 392
7	35238	Rennes	1999	4 213	3 987	0	3	8 158	21	18	6 094
8	35238	Rennes	2001	4 432	4 325	0	0	8 589	155	13	6 189
9	35238	Rennes	2007	3 237	3 150	1	10	6 355	3	20	3 068
10	35238	Rennes	2009	3 475	3 356	0	12	6 808	1	10	3 059
11	35238										
12	35238										
13	35238										
14	35238										
15	35238										
16	35238										



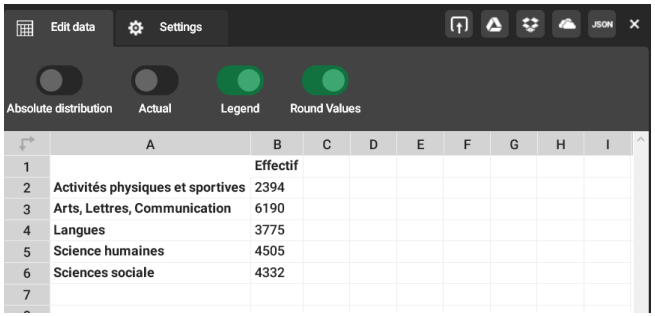
Visualisation de données



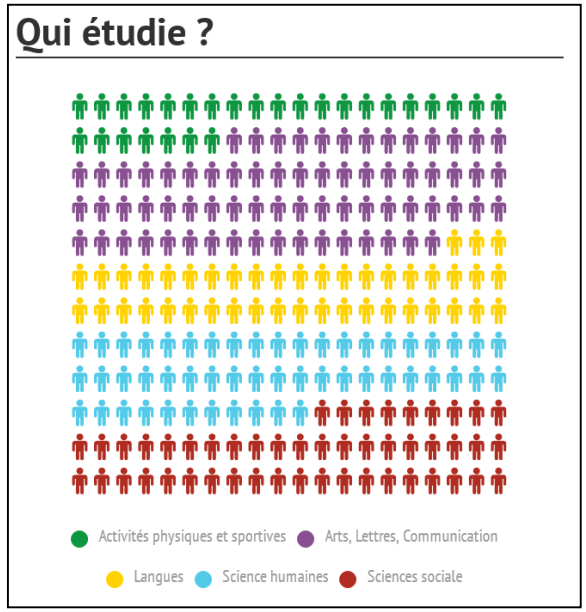


DataViz

- **La visualisation pour l'analyse des données**
  - Elle offre une ligne d'attaque frontale, révèle la structure complexe de données qui ne pourraient être comprises d'aucune autre façon
    - Permet de découvrir des résultats inattendus et de remettre en question les conclusions attendues.
    - Cherche à rendre compréhensibles et accessibles des données brutes en les scénarisant.



	A	B
1		Effectif
2	Activités physiques et sportives	2394
3	Arts, Lettres, Communication	6190
4	Langues	3775
5	Science humaines	4505
6	Sciences sociale	4332
7		

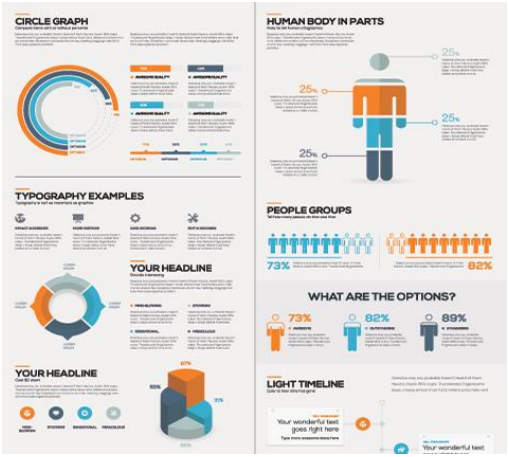


DataViz

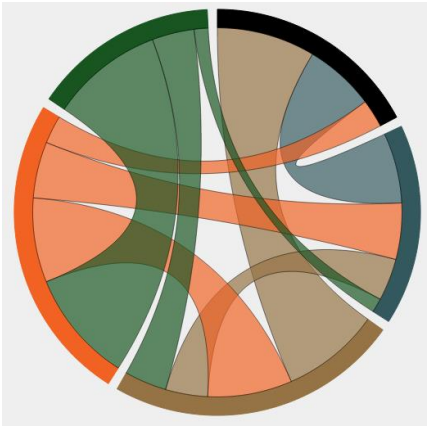
- **L'idée : présenter des données complexes de façon simple, juste et attractive**
  - Une Dataviz sert à rendre l'information lisible par tous pour faciliter la compréhension en s'adressant au côté sensoriel et rationnel
  - Pour répondre à ces objectifs, la DataViz doit fournir une information :
    - Interprétable, c'est-à-dire claire, quelque soit le volume, la nature ou la provenance des données
    - Pertinente, c'est à dire qui réponde à un objectif métier dans un contexte défini
    - Novatrice, en fournissant une perspective nouvelle d'appréhension de données

Infographie / Dataviz

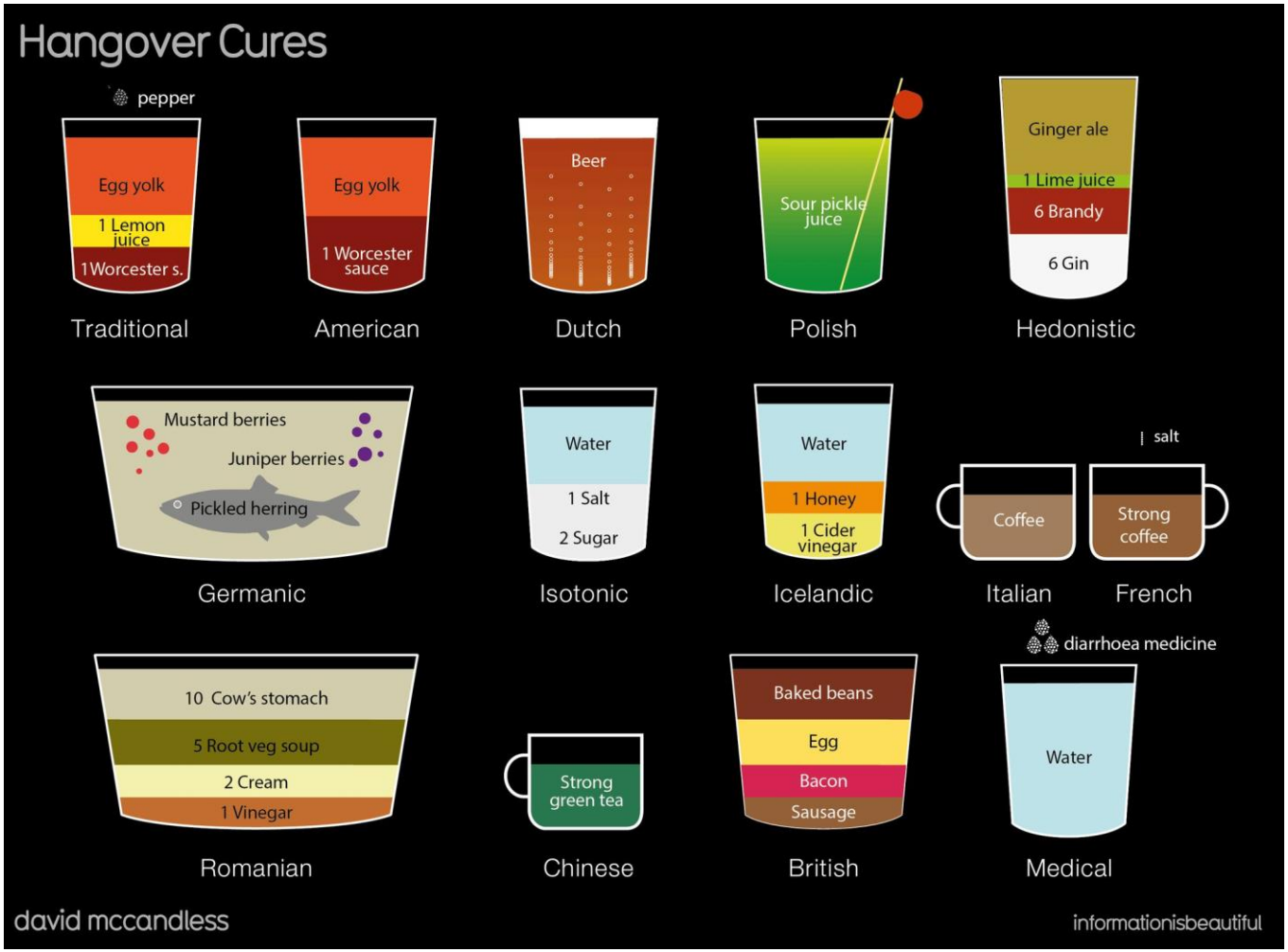
- Les infographies sont créées dans le but de raconter ou d'expliquer une histoire pour une audience spécifique
  - Mobilise des Dataviz, du texte, des images, des pictogrammes,…



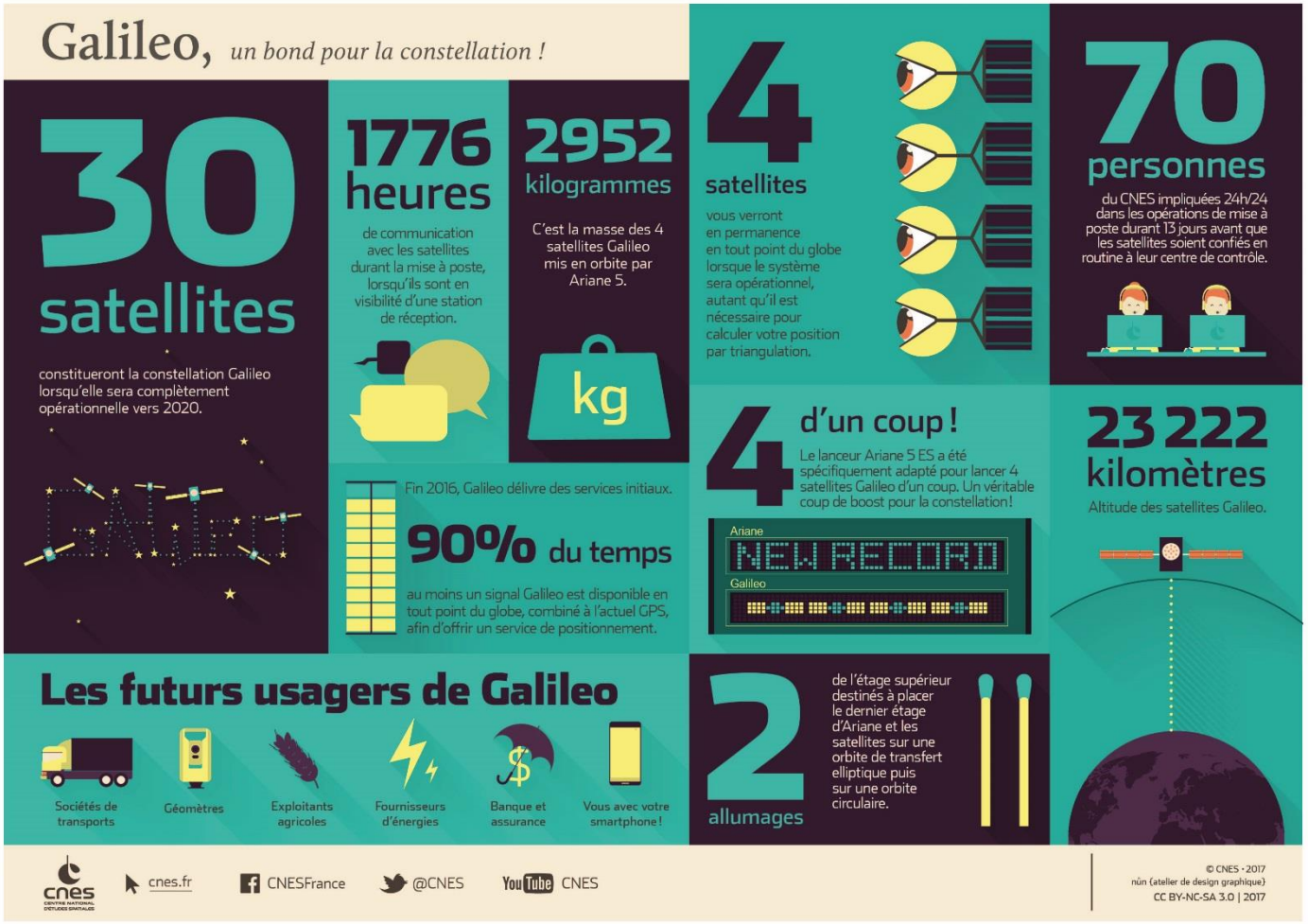
- La Dataviz représente une seule idée (un jeu de données).
  - La visualisation de données permet de comprendre facilement des informations très compliquées par la représentation graphique.



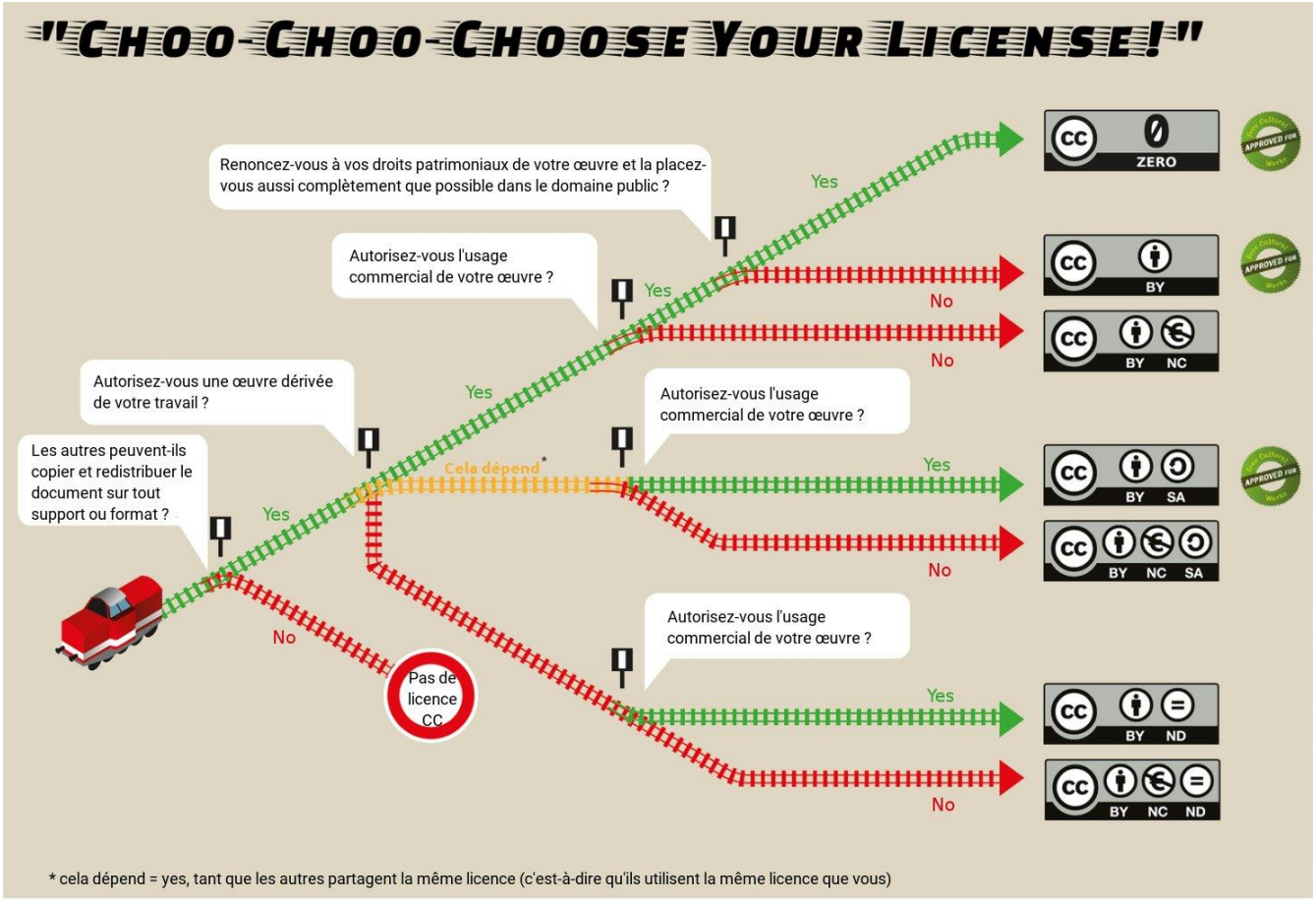
Infographie



Infographie



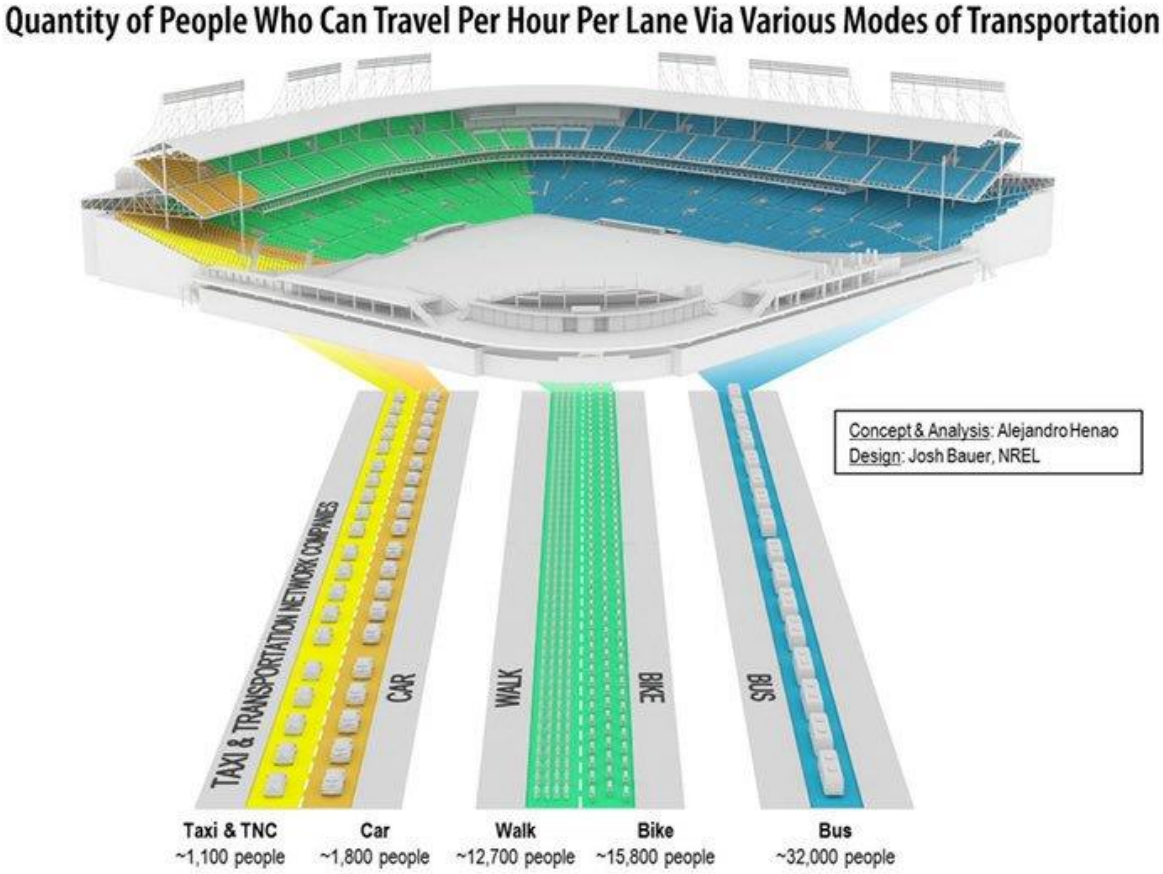
Infographie



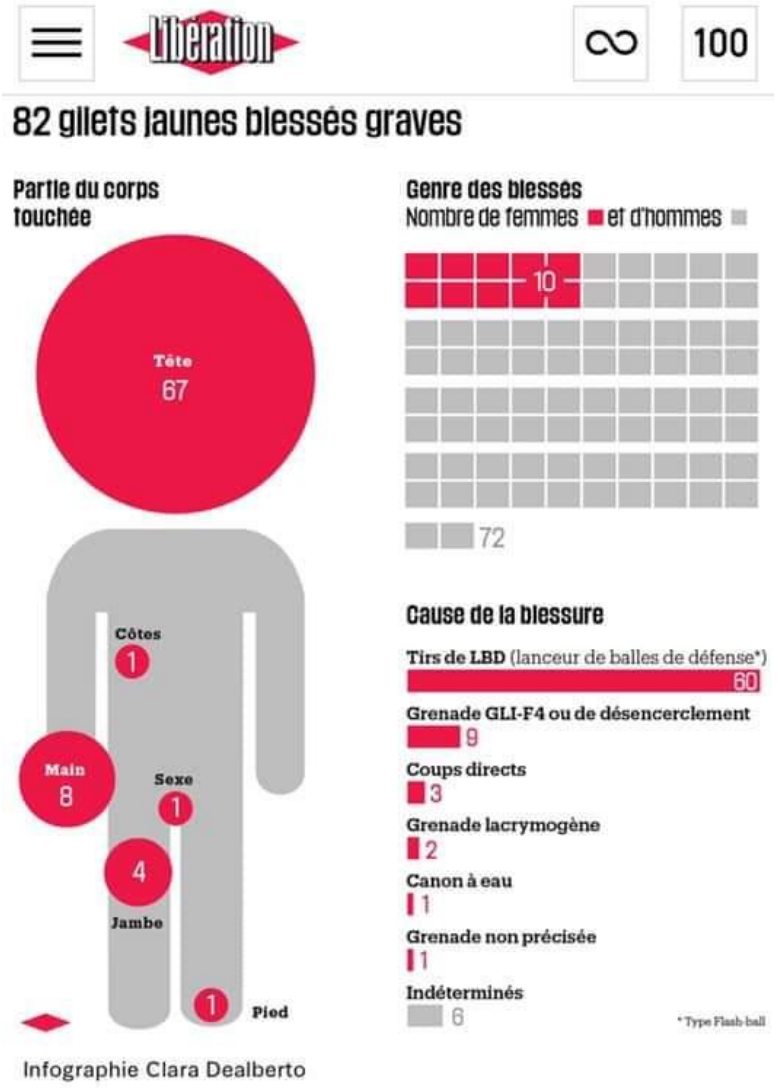
CC BY SA

Ce graphique « Choo-Choo-Choose your license » est basé sur le travail «Welche CC-Lizenz ist die richtige für mich ? », réalisé par Barbara Klute et Jöran Muuß-Merholz pour http://www.wb-web.de sous la licence CC BY-SA 3.0.
La version française est une traduction de Cédric Frayssinet sous la même licence.

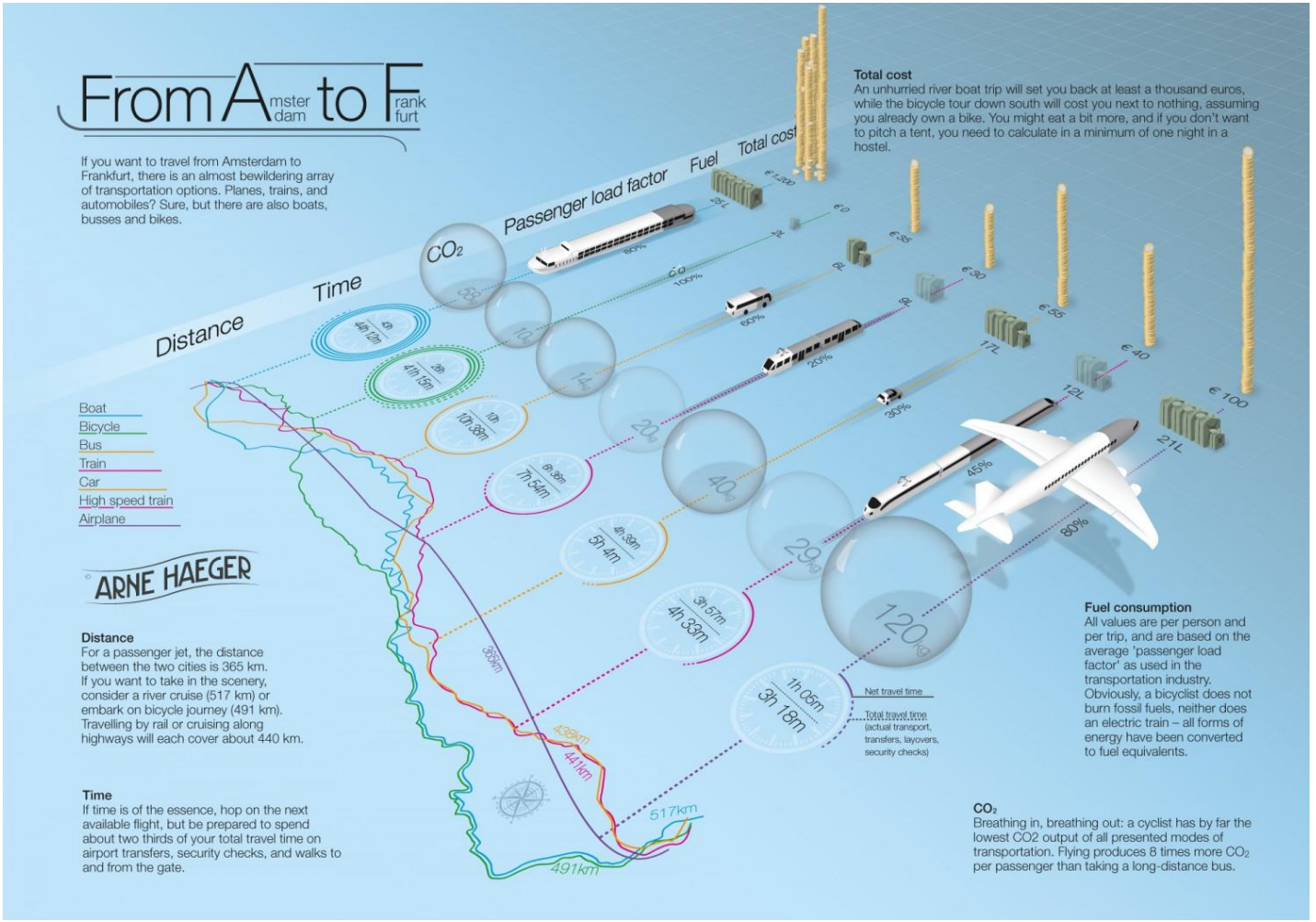
Infographie



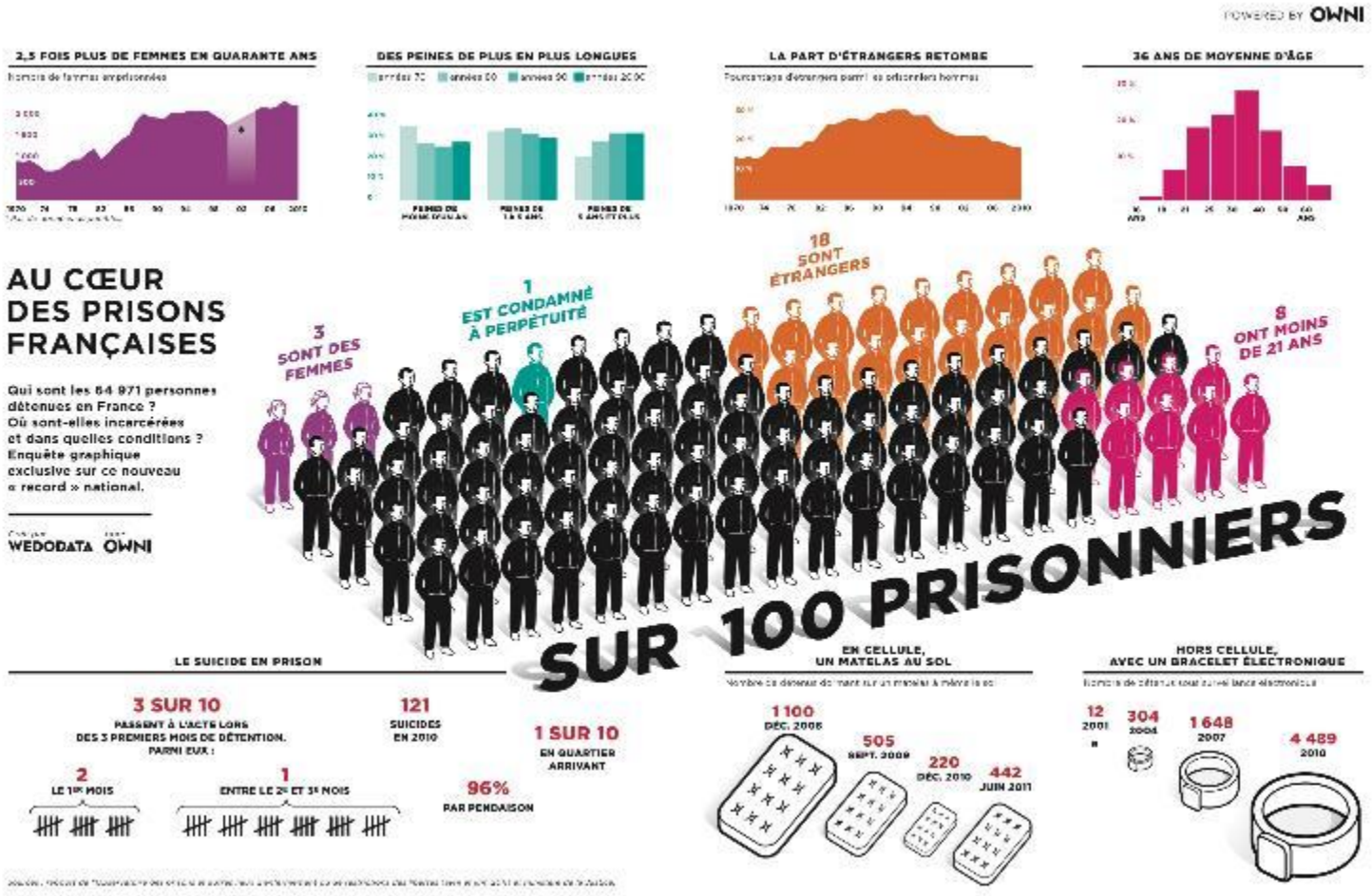
Infographie



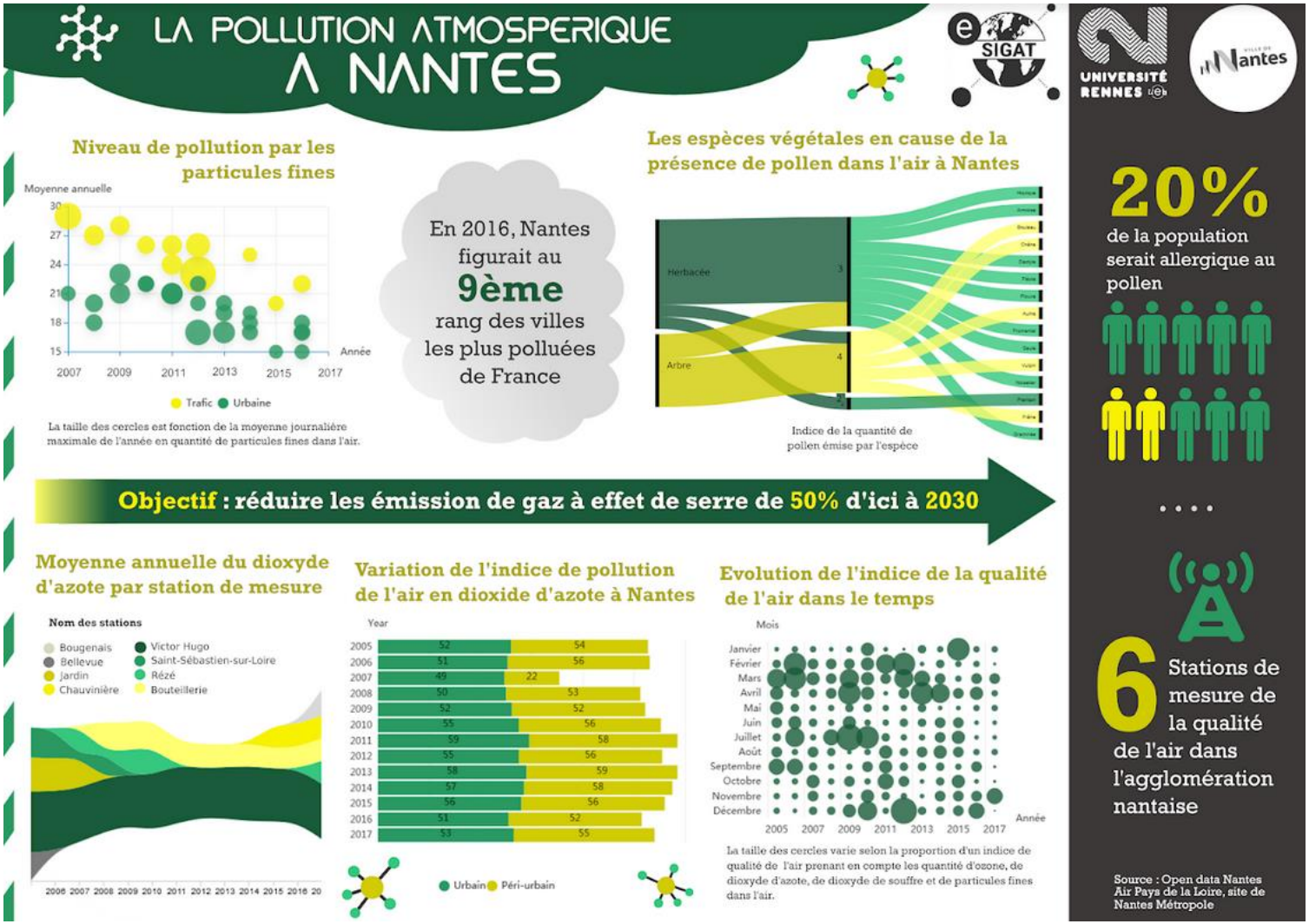
Infographie



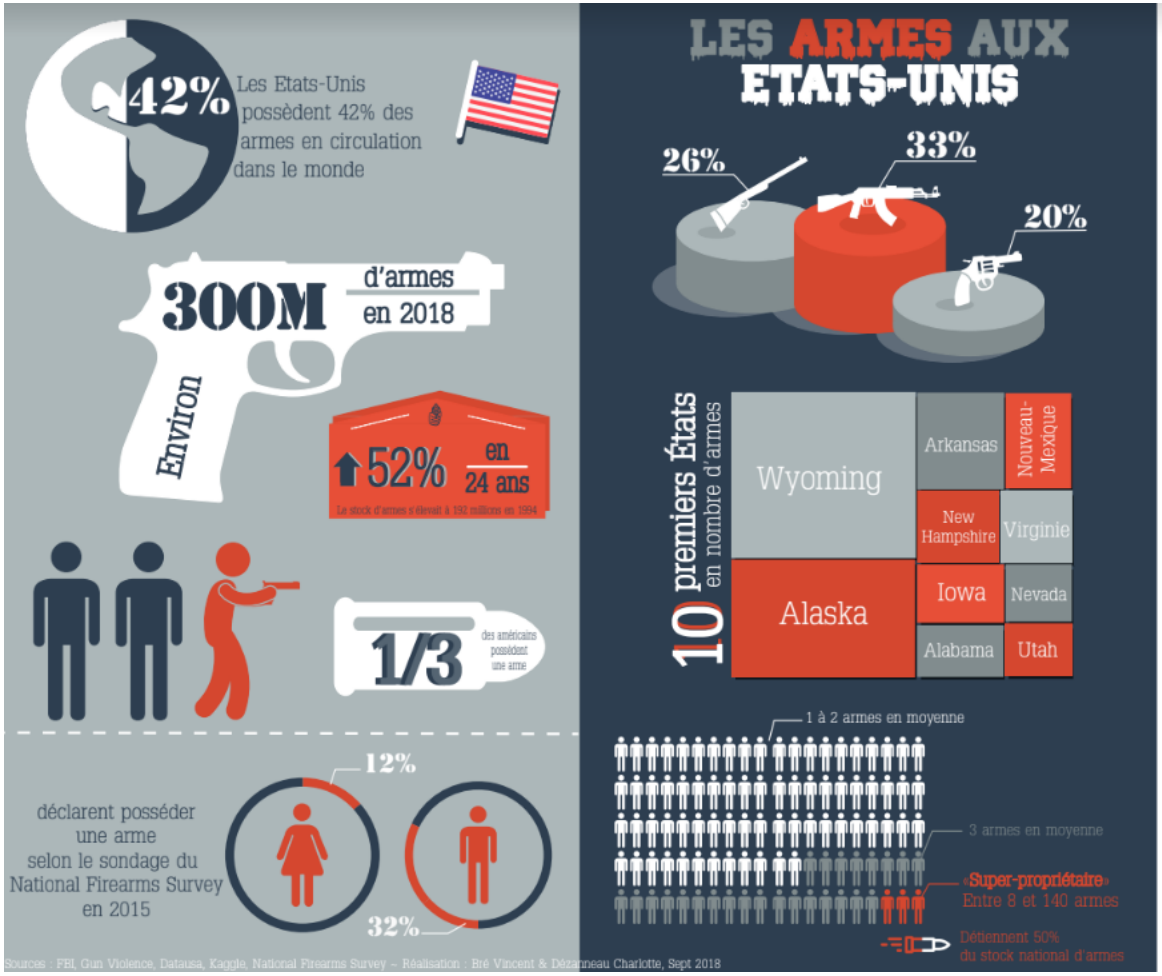
Infographie

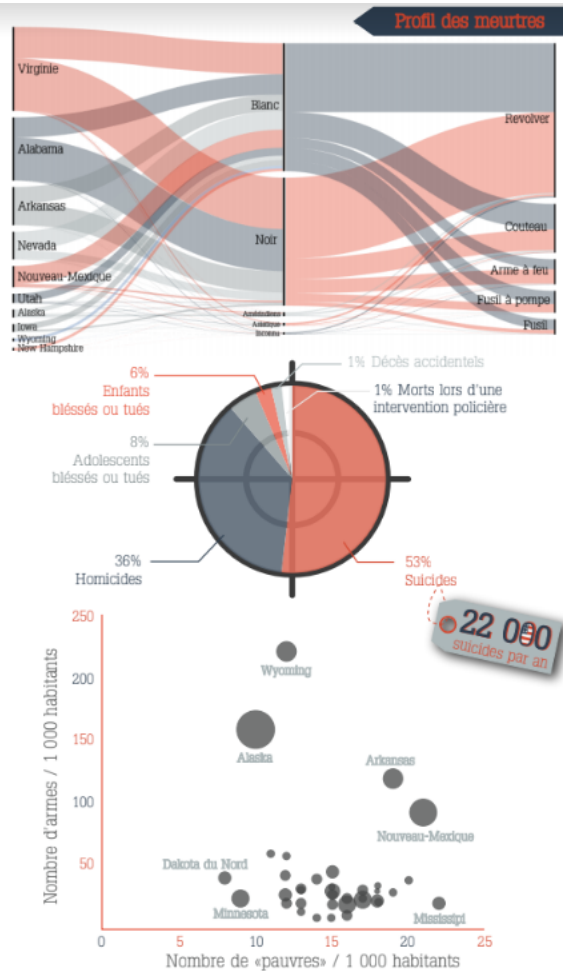


Infographie / SIGAT

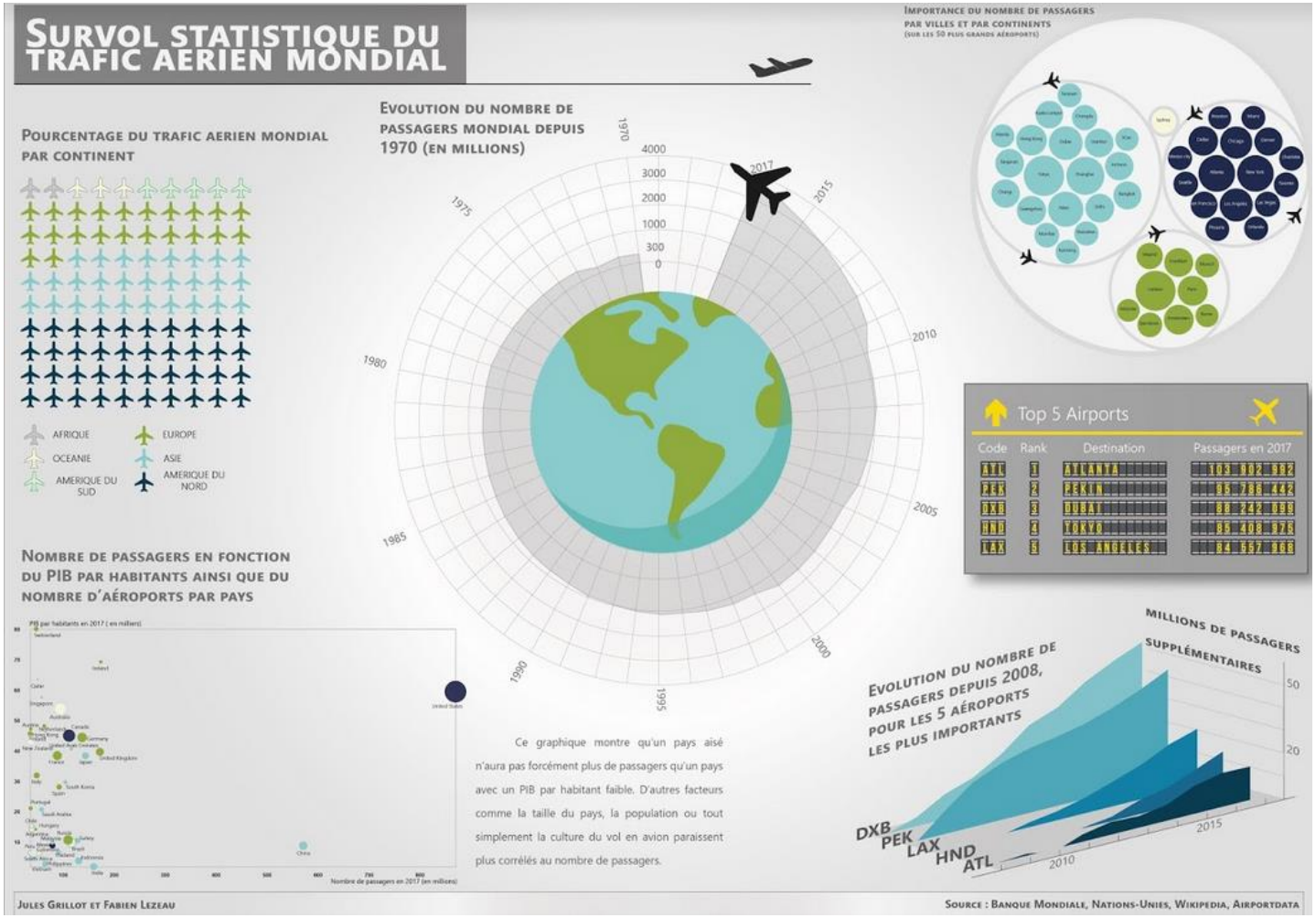


Infographie / SIGAT



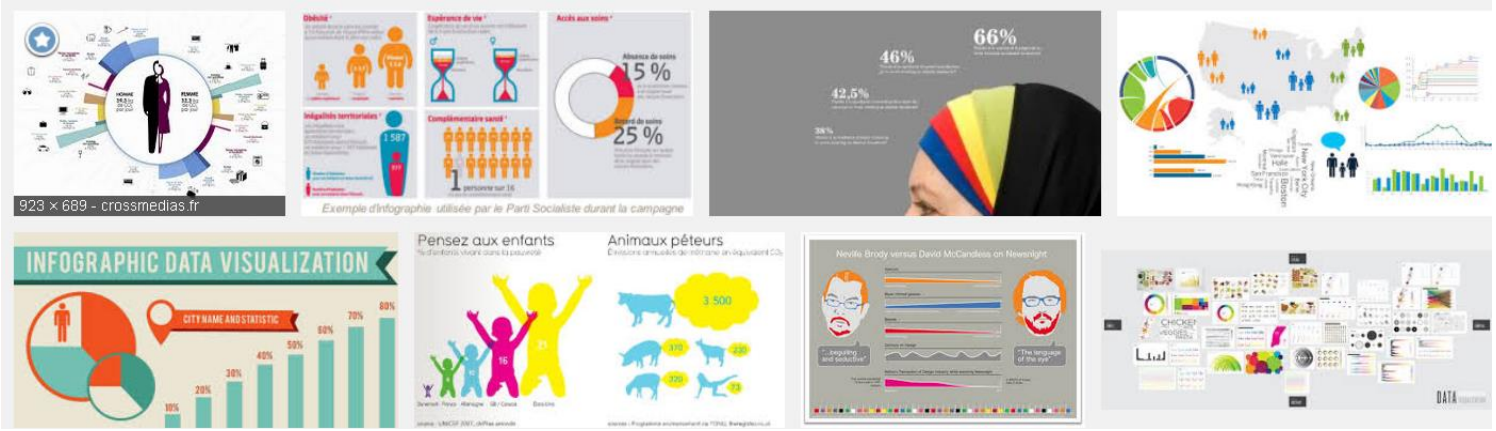


Infographie / SIGAT



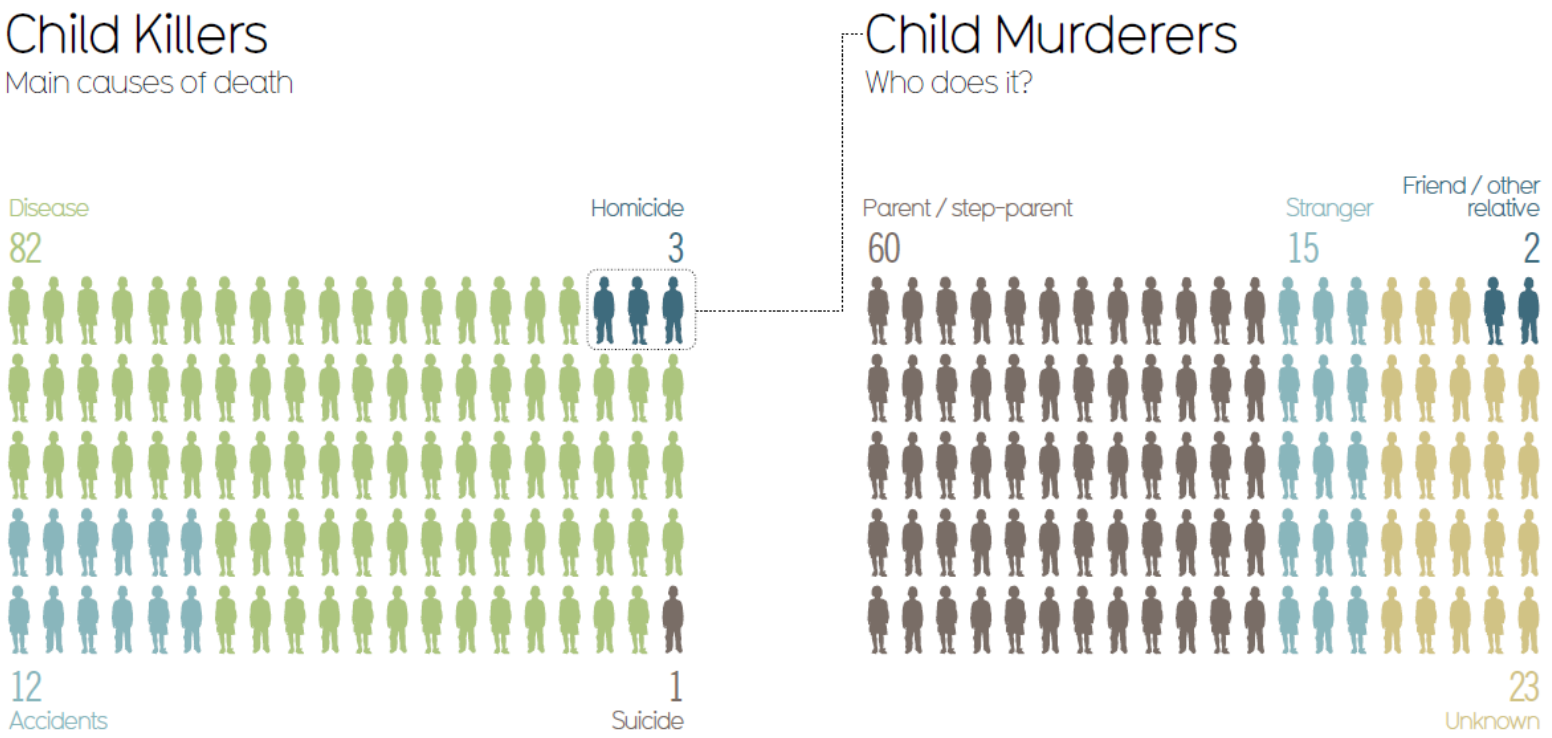
DataViz « statiques »

- **Data Viz « statiques »**
  - Des visualisation de données figées, représentant les données ciblées
  - Cette catégorie a l'avantage d'être complètement adaptée aux contextes d'utilisations des réseaux sociaux mais aussi aux rapports et aux présentations



DataViz « statiques »

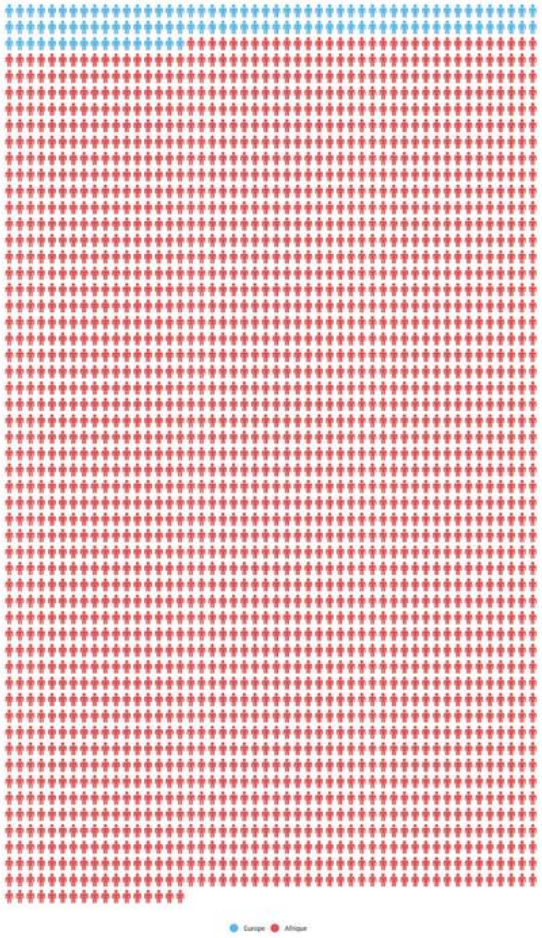
#Composition (répartition)



DataViz « statiques »

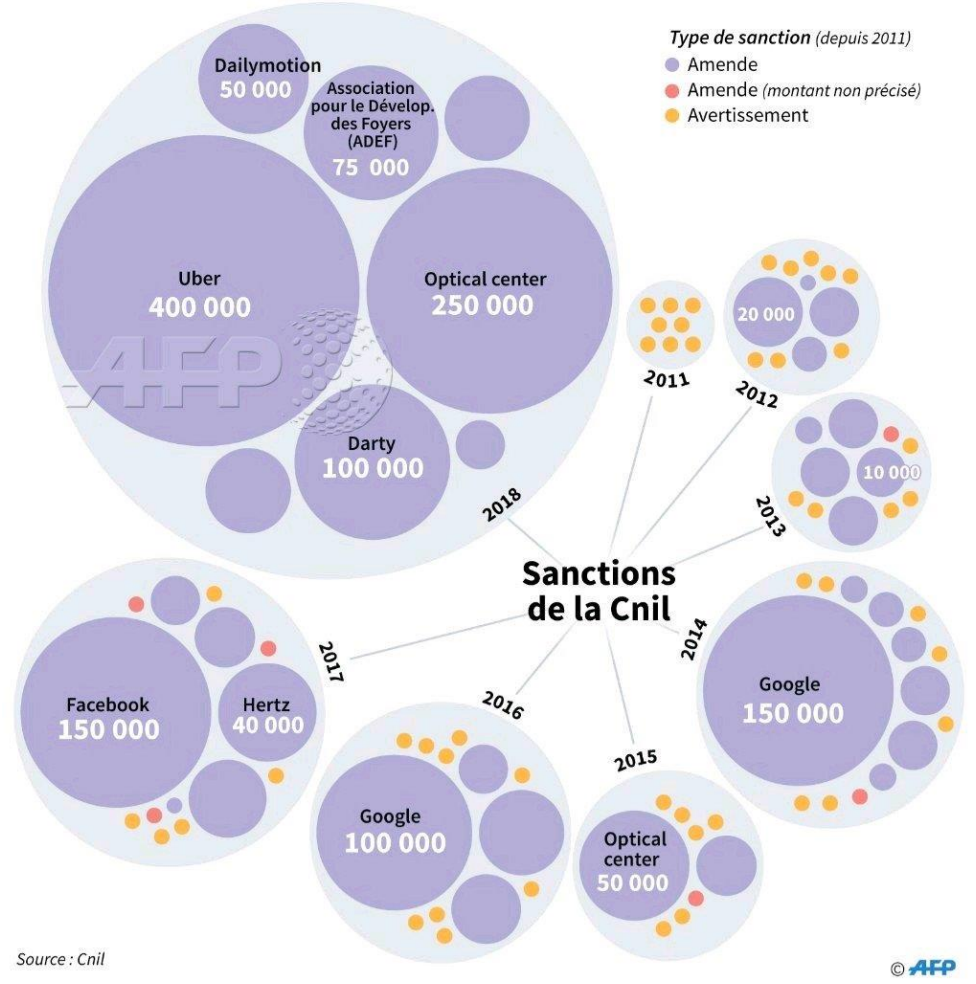
#Composition (répartition)

People killed in terrorist attacks in 2017: Europe (Blue) vs Africa (Red) [OC]



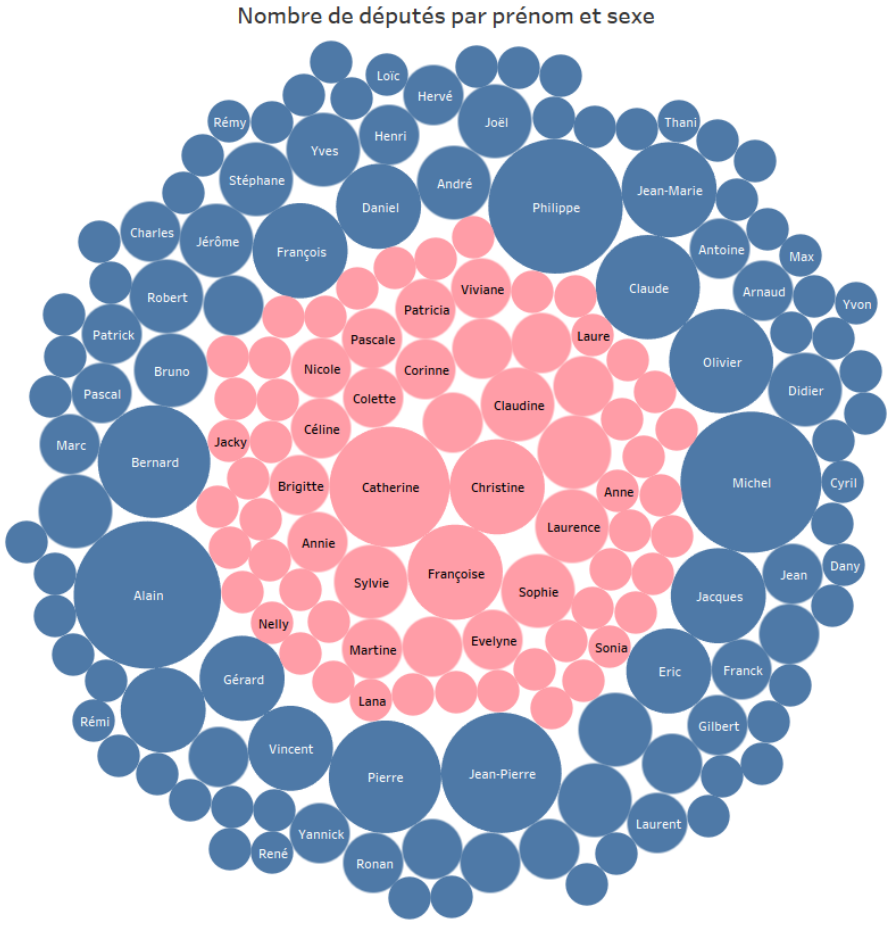
DataViz « statiques »

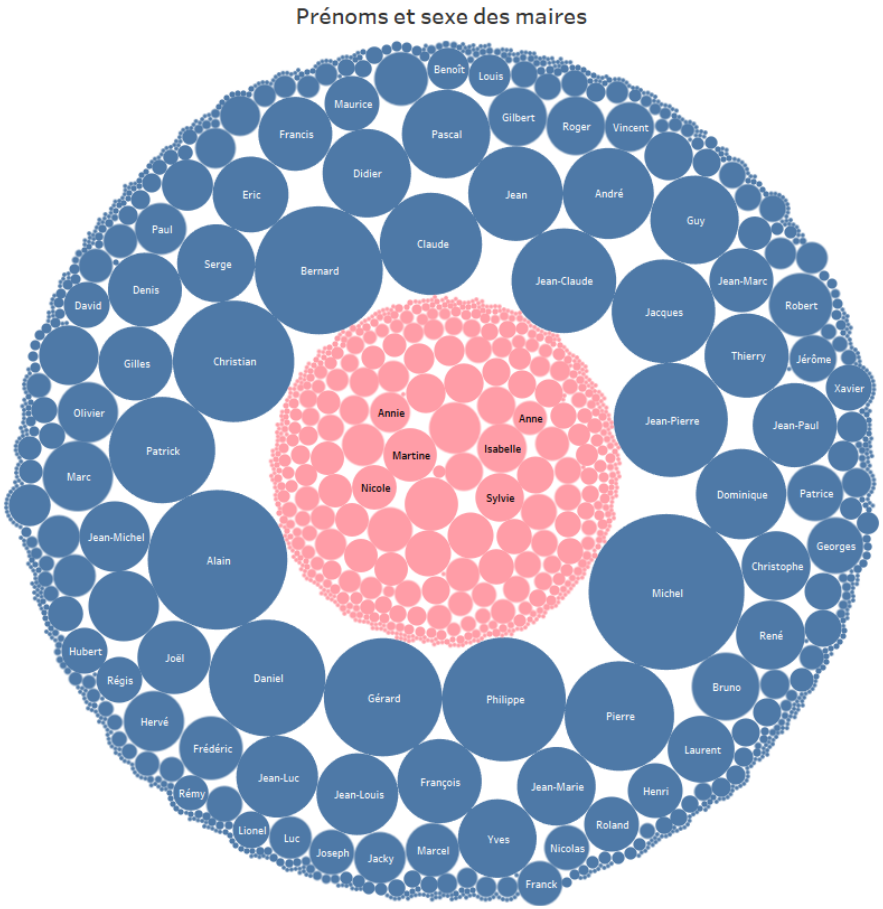
#Composition (répartition)



DataViz « statiques »

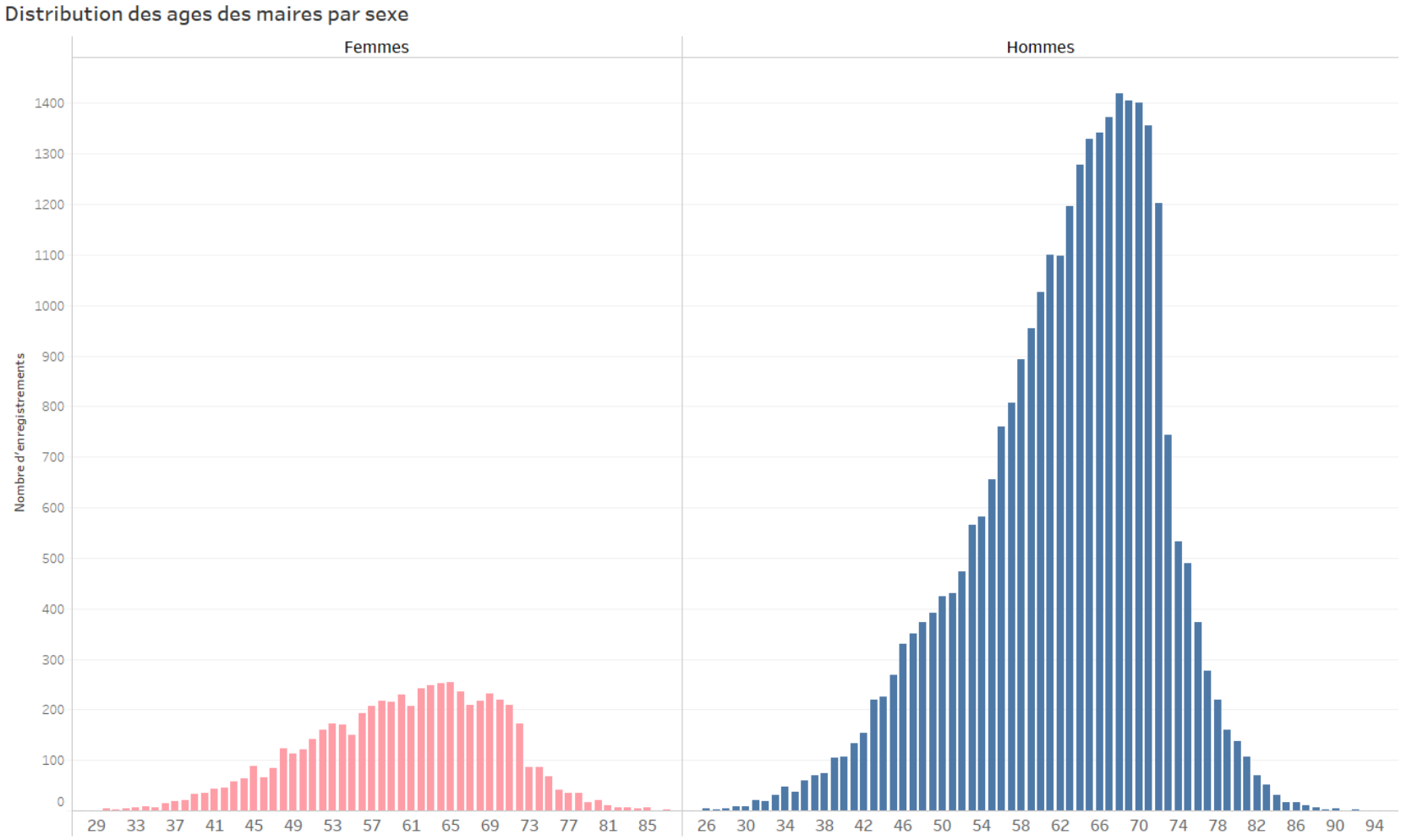
#Composition (répartition)





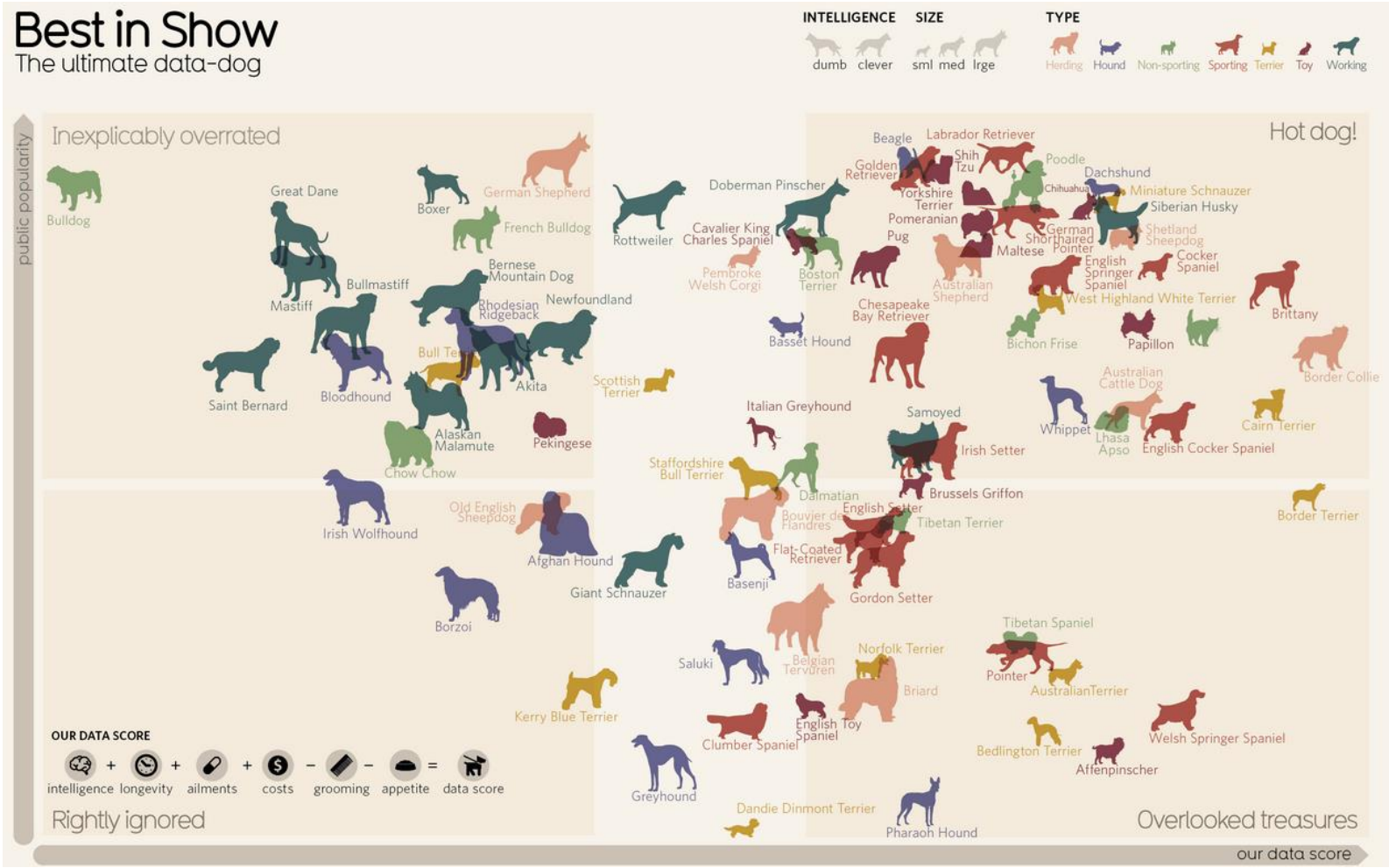
DataViz « statiques »

#Comparaison



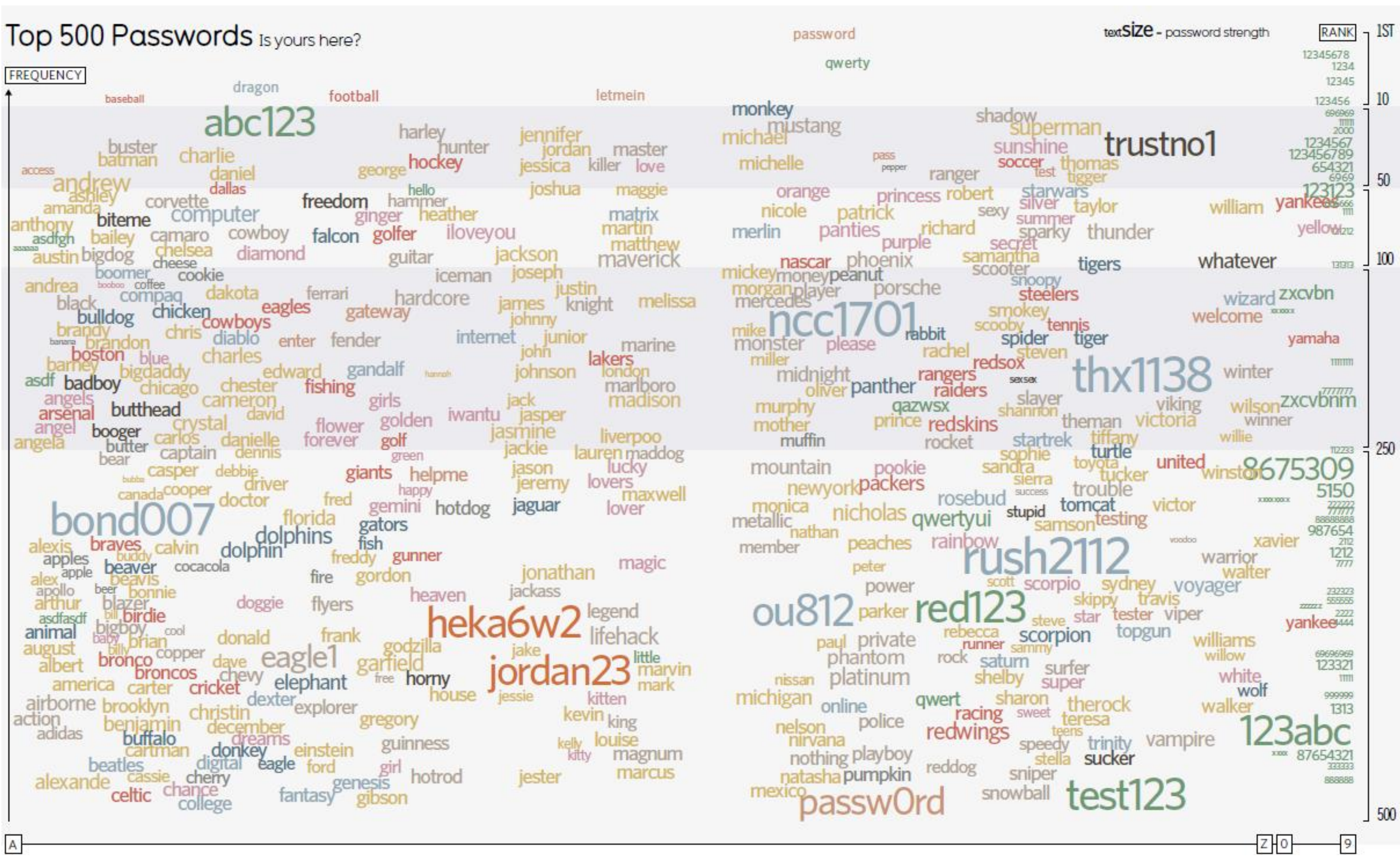
DataViz « statiques »

#Comparaison



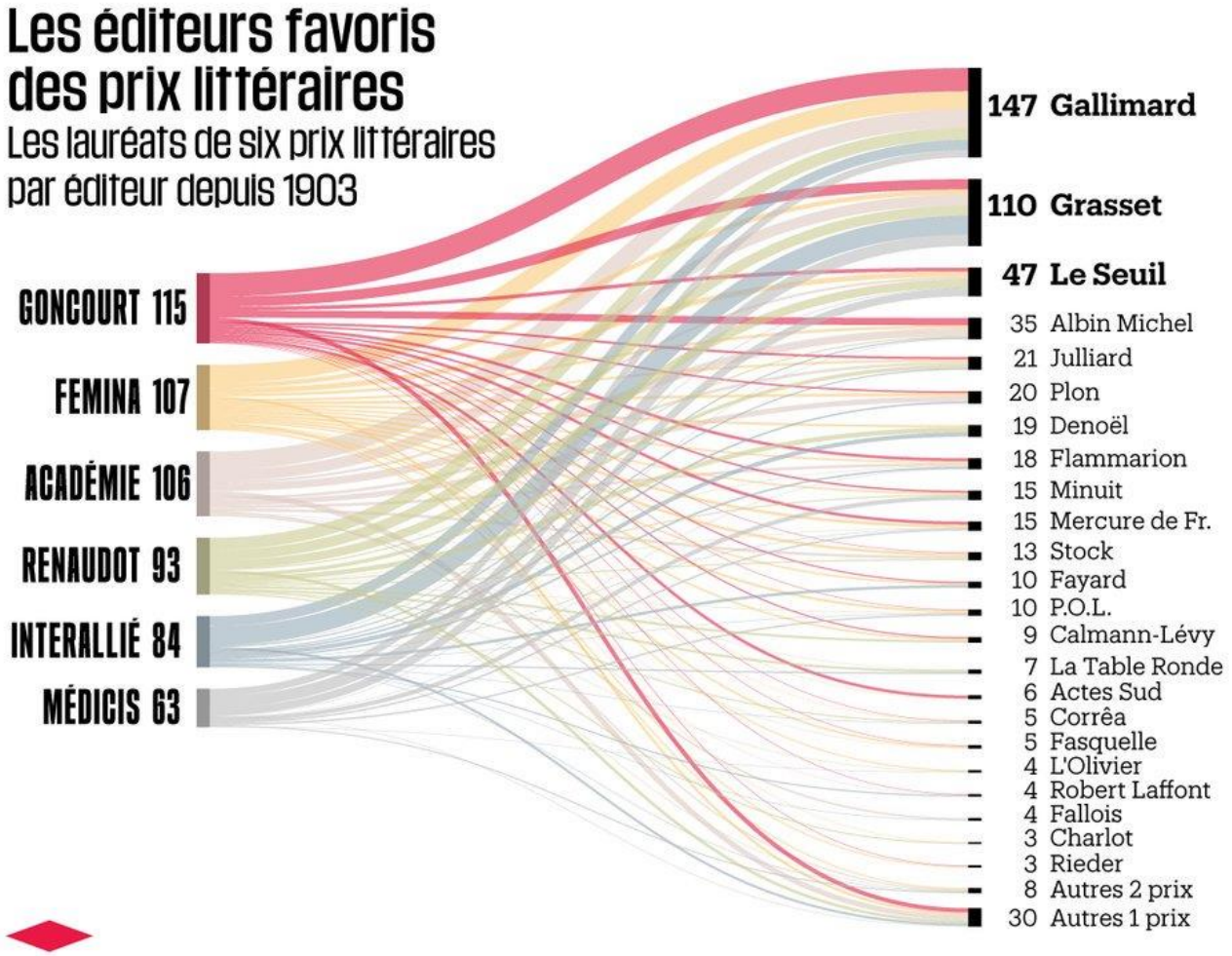
DataViz « statiques »

#Comparaison



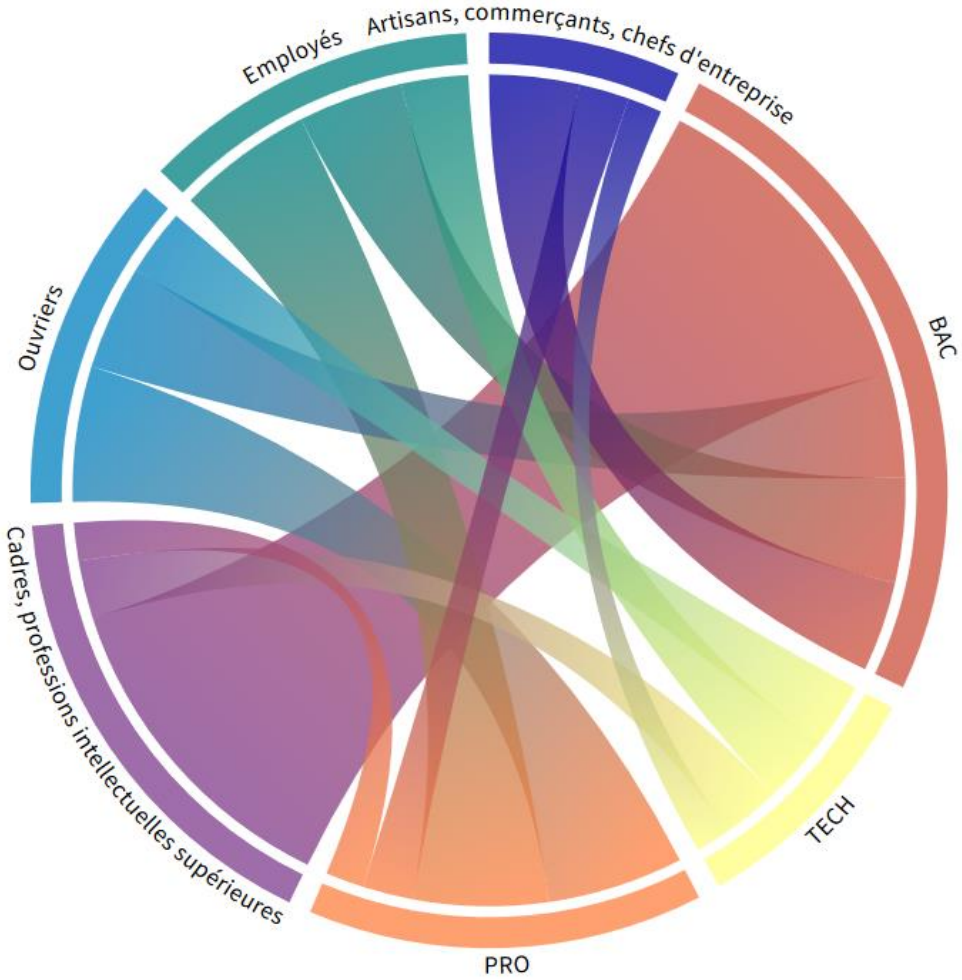
DataViz « statiques »

#Relation / répartition



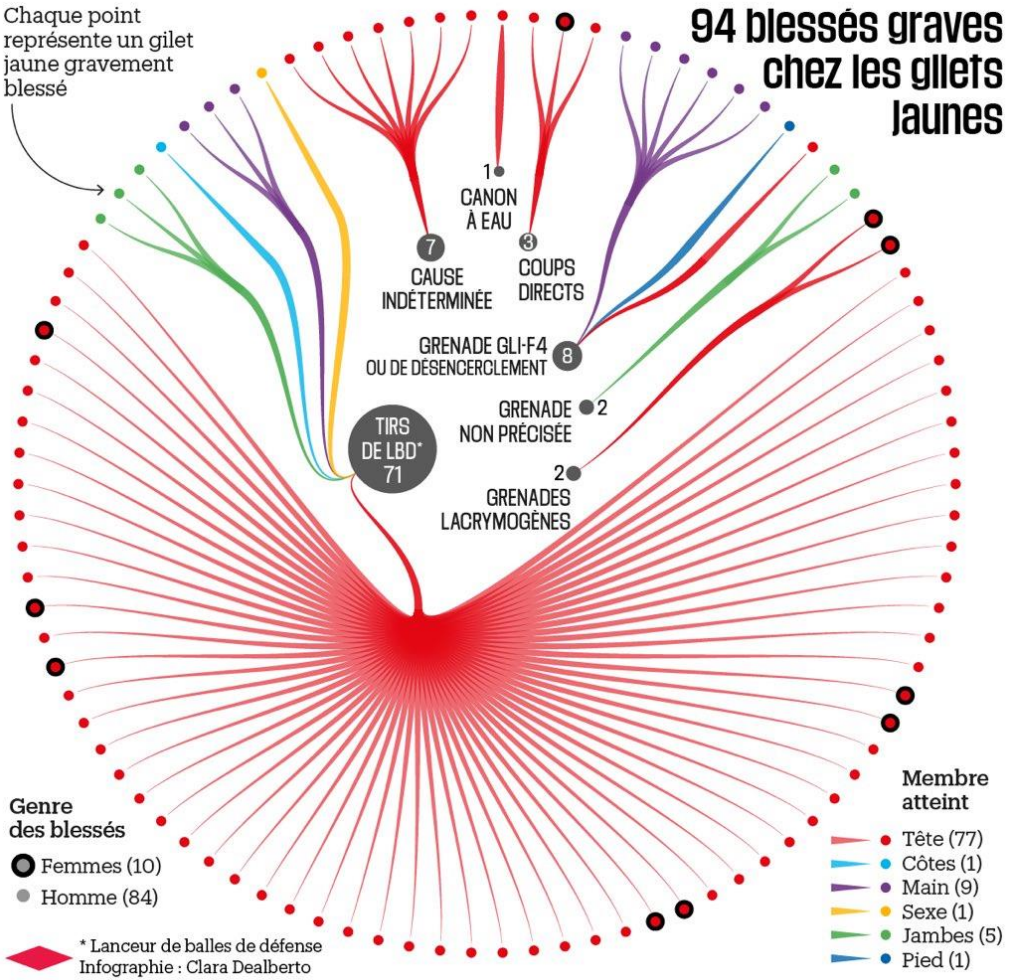
DataViz « statiques »

#Relation / répartition



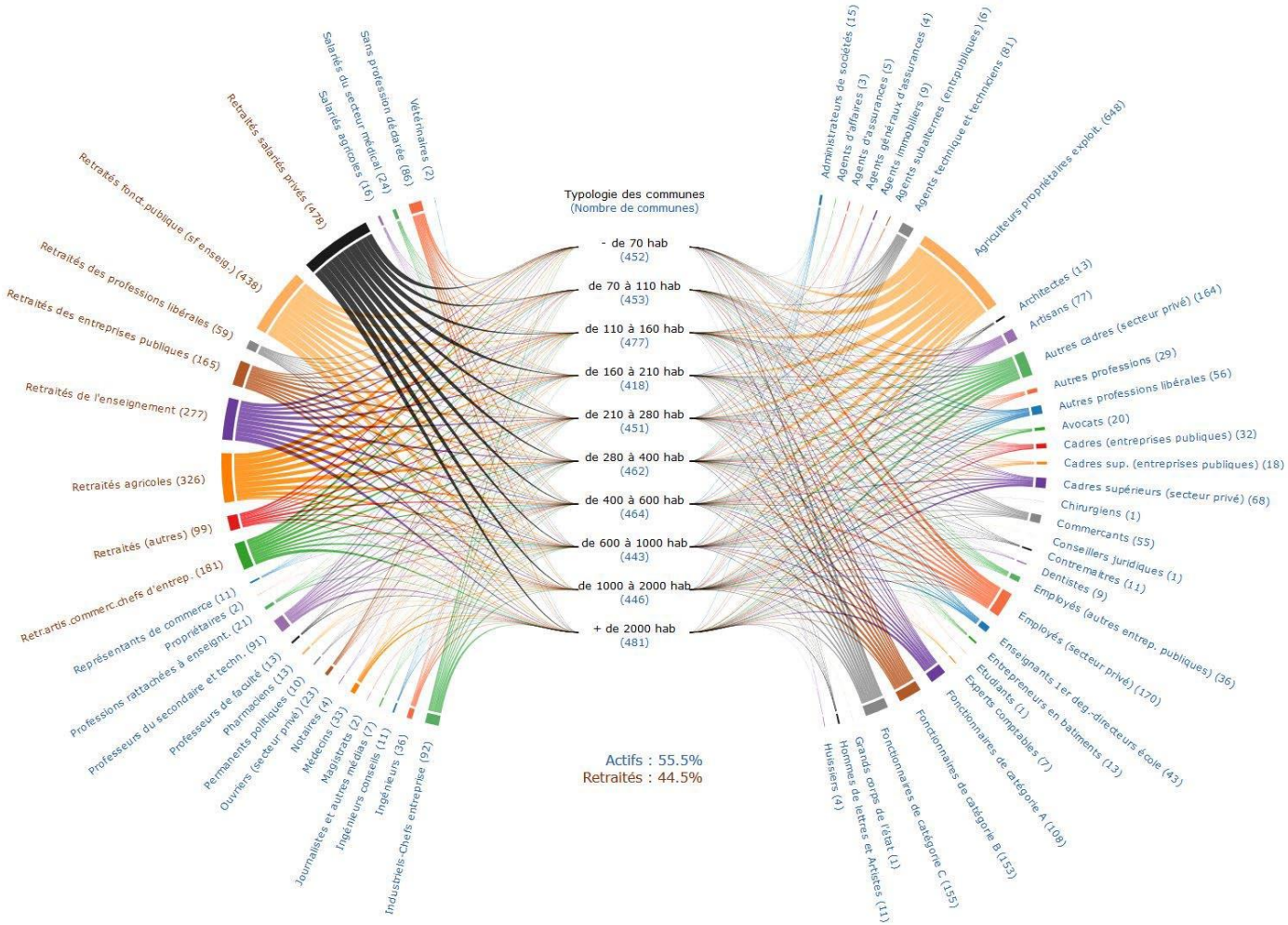
DataViz « statiques »

#Relation / répartition



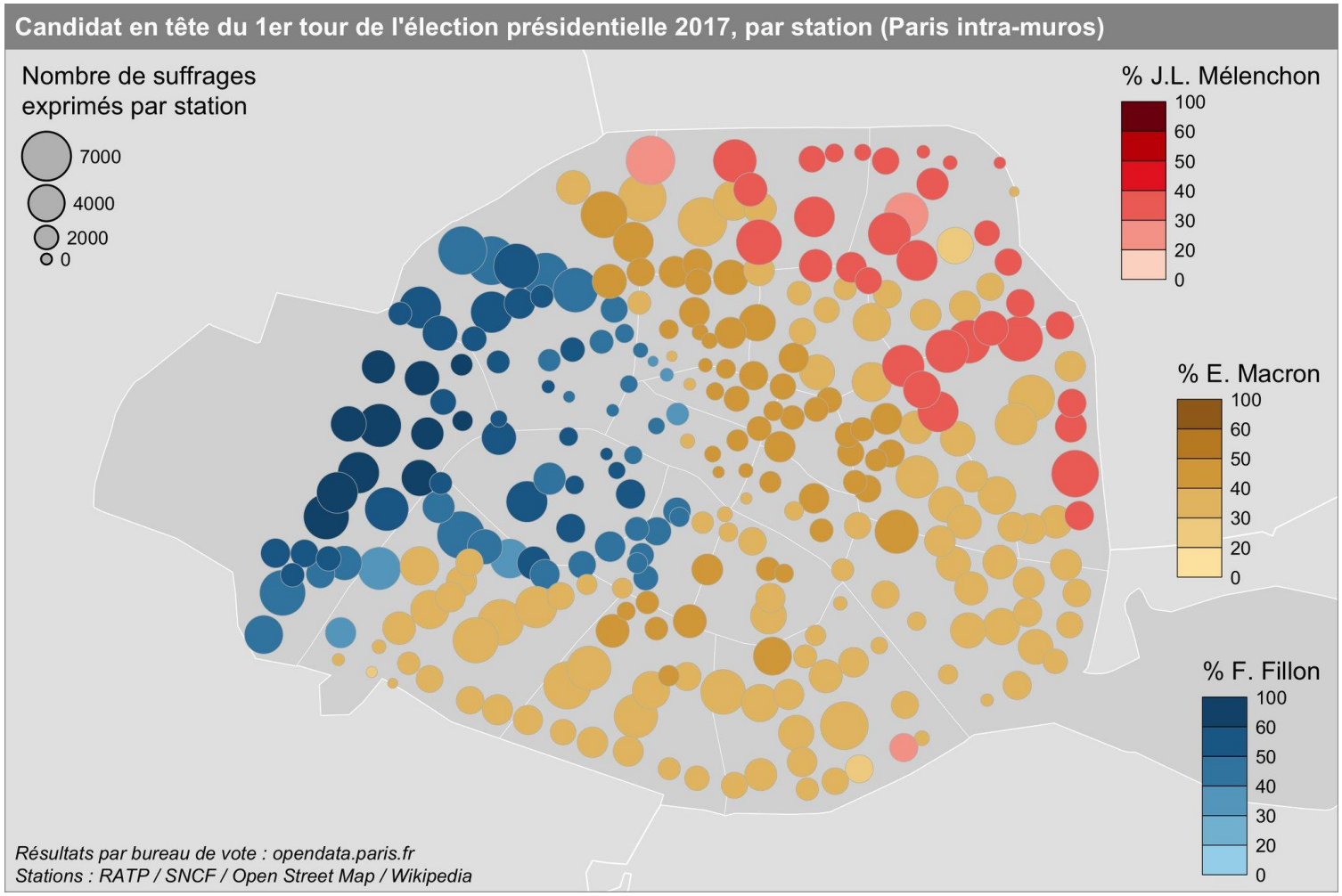
DataViz « statiques »

#Relation / répartition



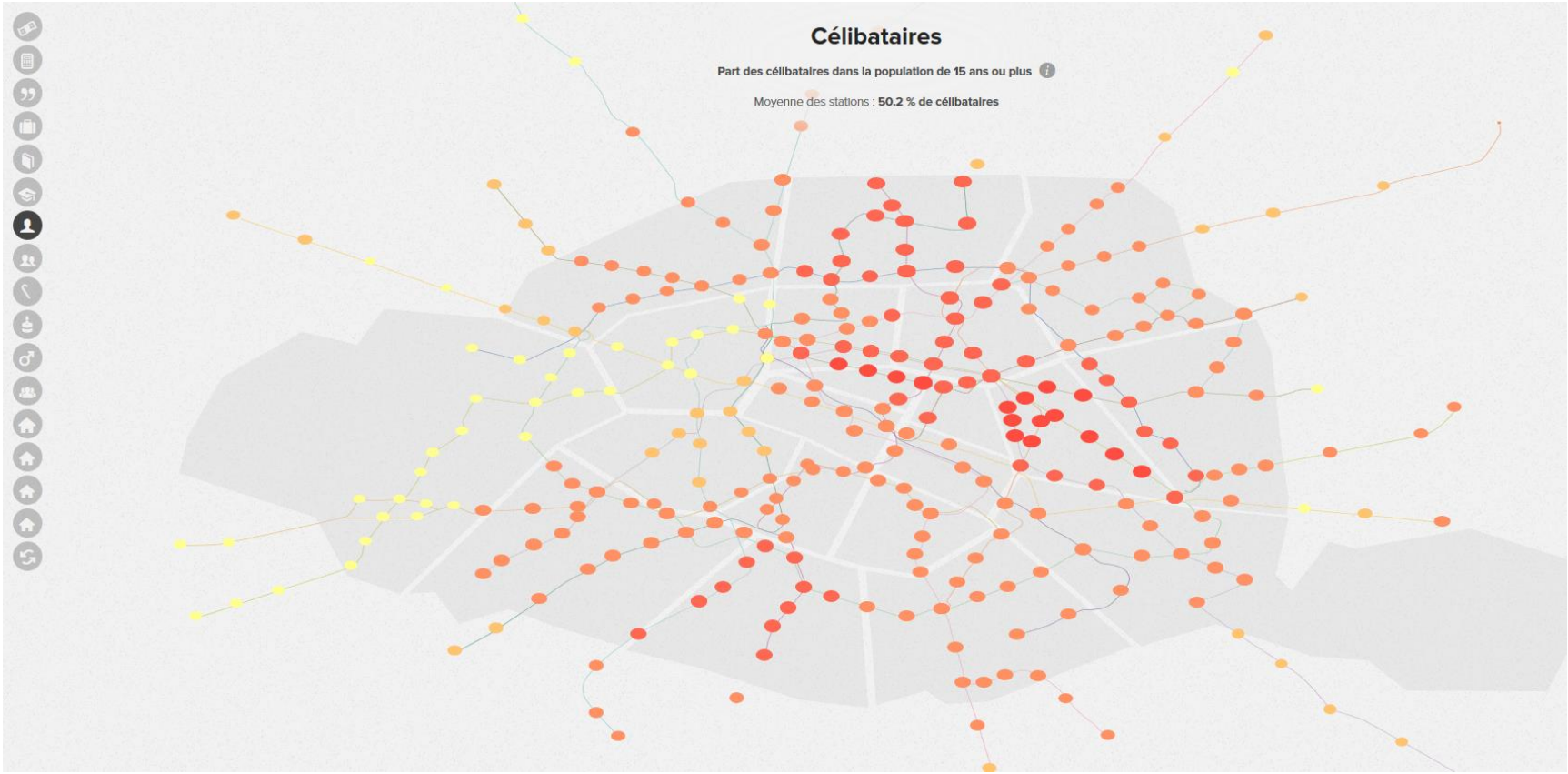
DataViz « statiques »

#Géovisualisation



DataViz « statiques »

#Géovisualisation



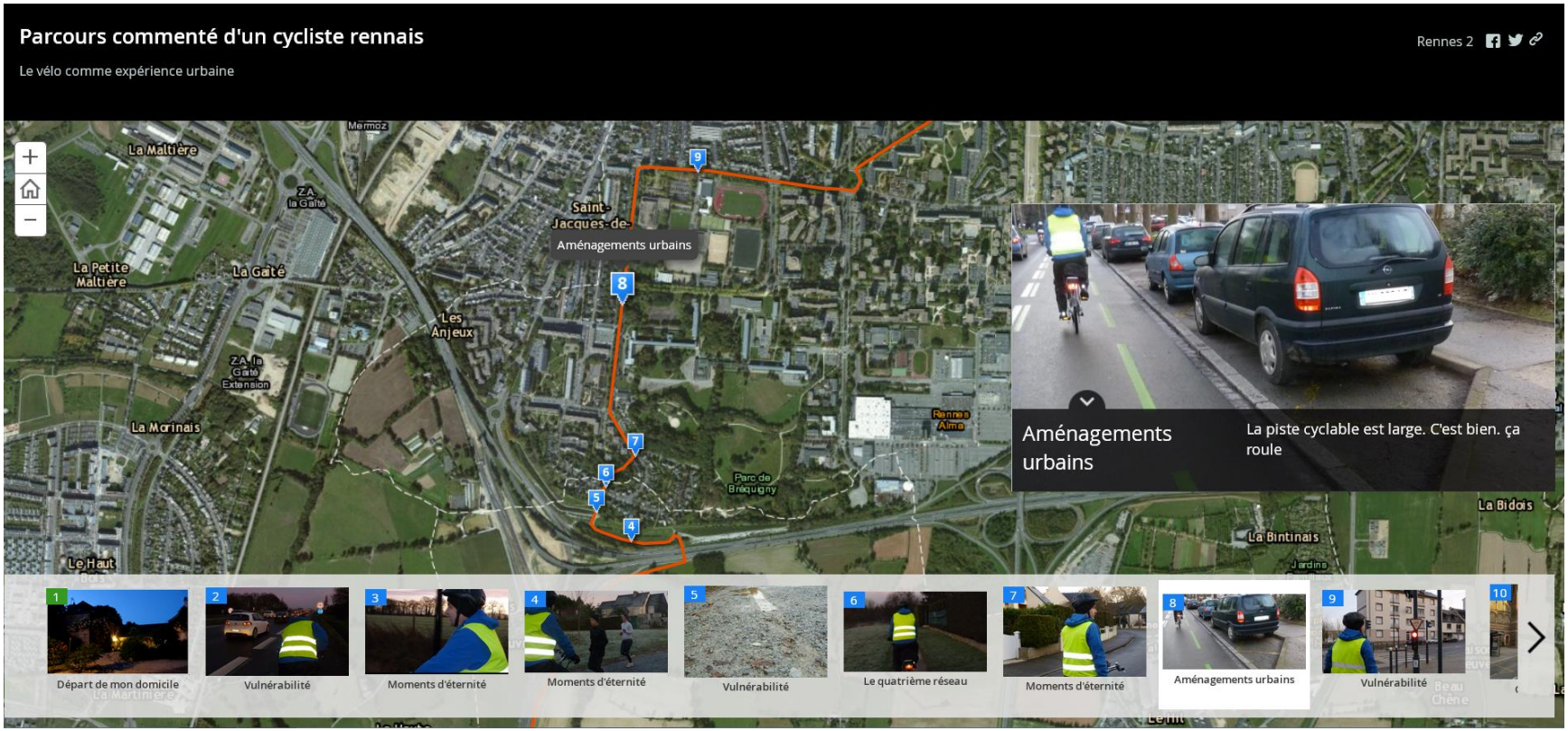
http://vincentgarreau.com/dataparis/#

Data viz « animées »

- La scénarisation, le story-telling

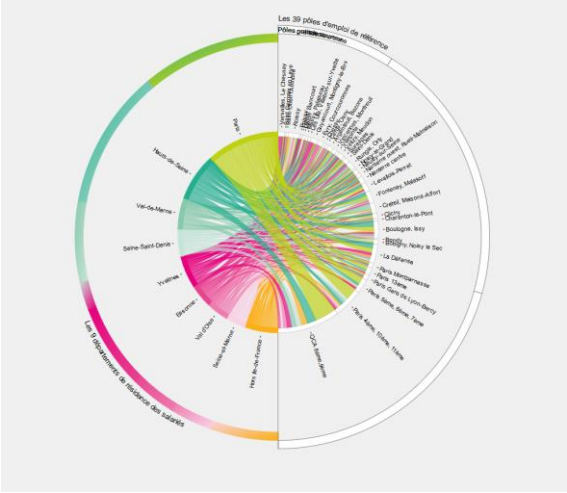
Data viz « animées »

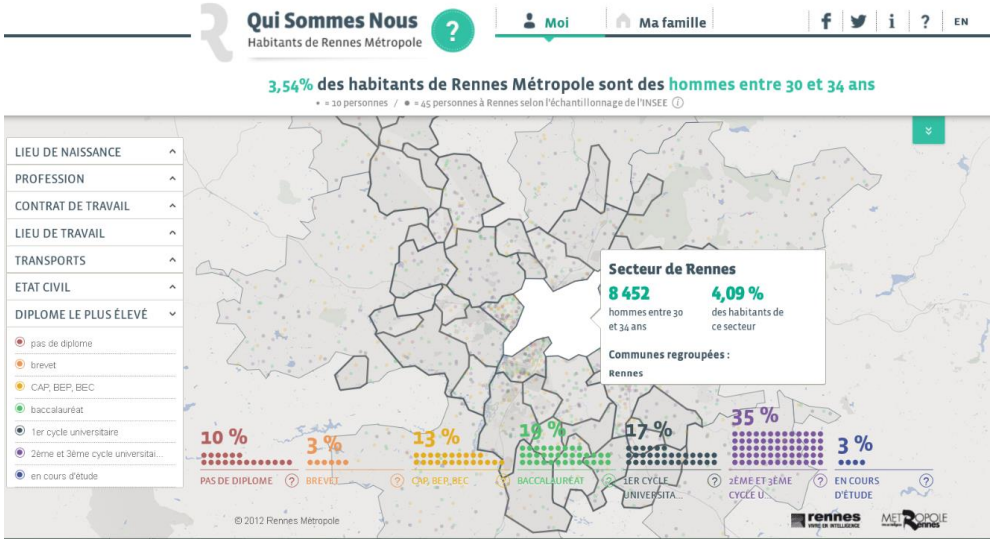
- La scénarisation, le story-mapping



Dataviz « interactives »

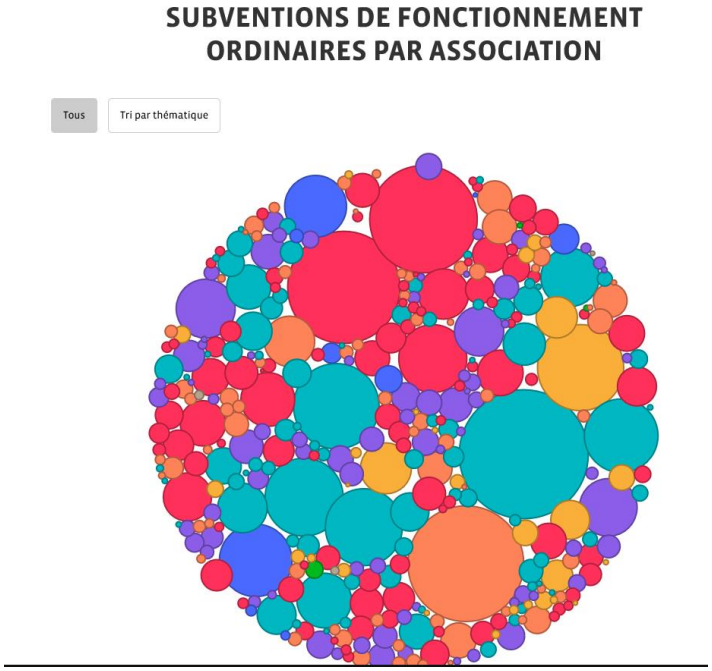
- **Les dataviz « interactives »**
  - Elles offrent la possibilité à l'utilisateur de « jouer », voire de construire sa propre visualisation
  - Si l'objectif à remplir est une prise de décision, on utilisera plutôt une dataviz dite « interactive » permettant d'afficher une quantité importante d'informations en disposant d'une certaine granularité de chiffres ou de nuances





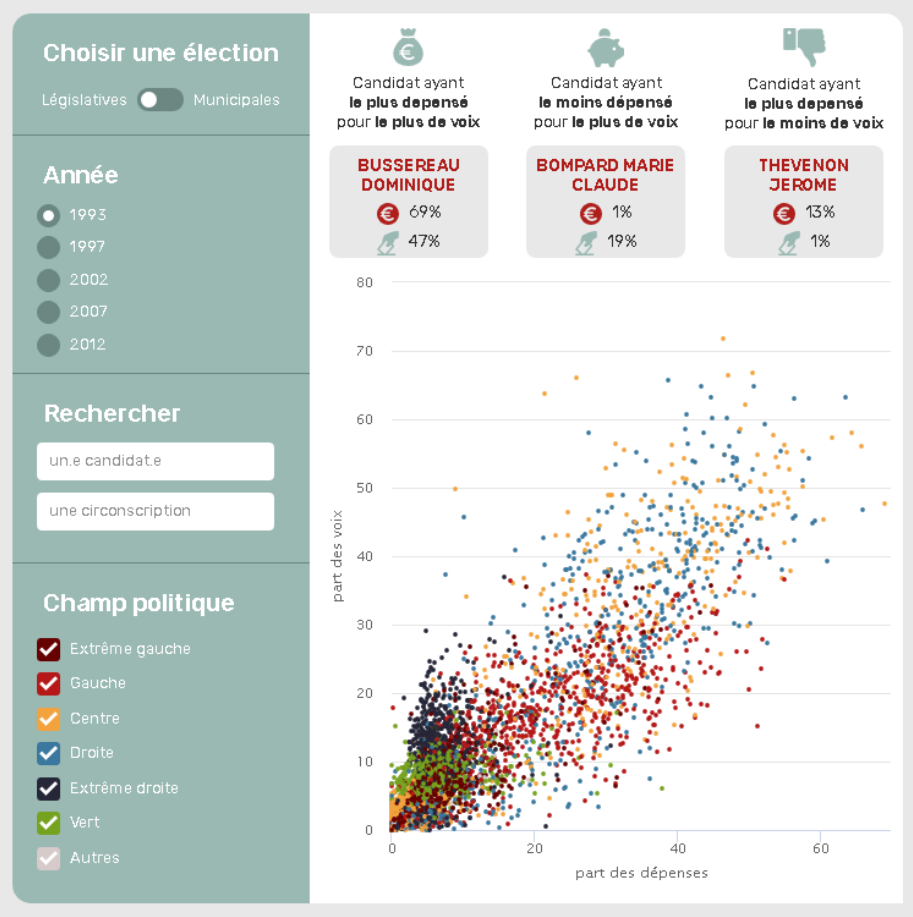
Dataviz « interactives »

- http://dataviz.rennesmetropole.fr/budget/



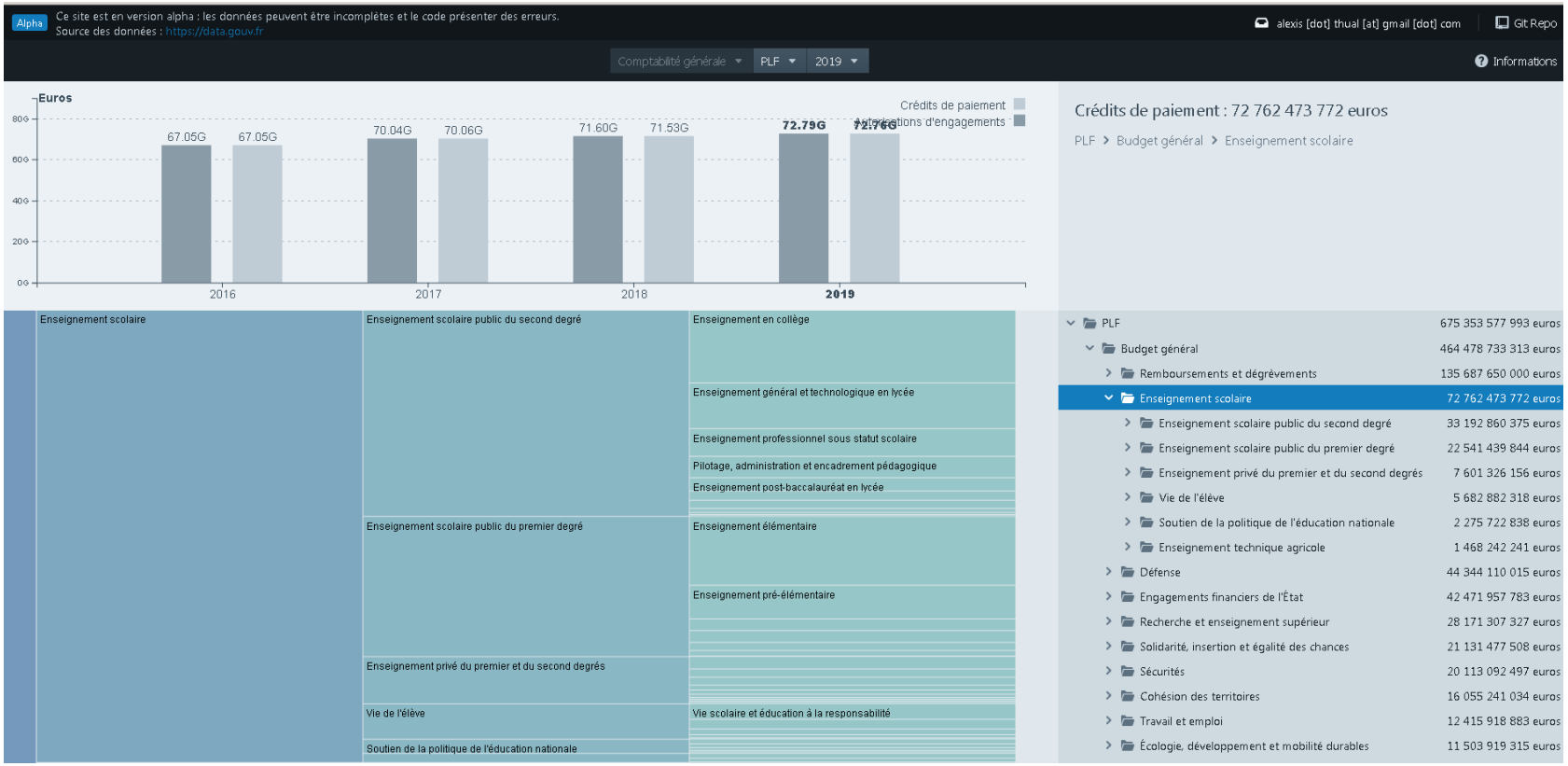
Dataviz « interactives »

- http://www.leprixdelademocratie.fr/data.php



Dataviz « interactives »

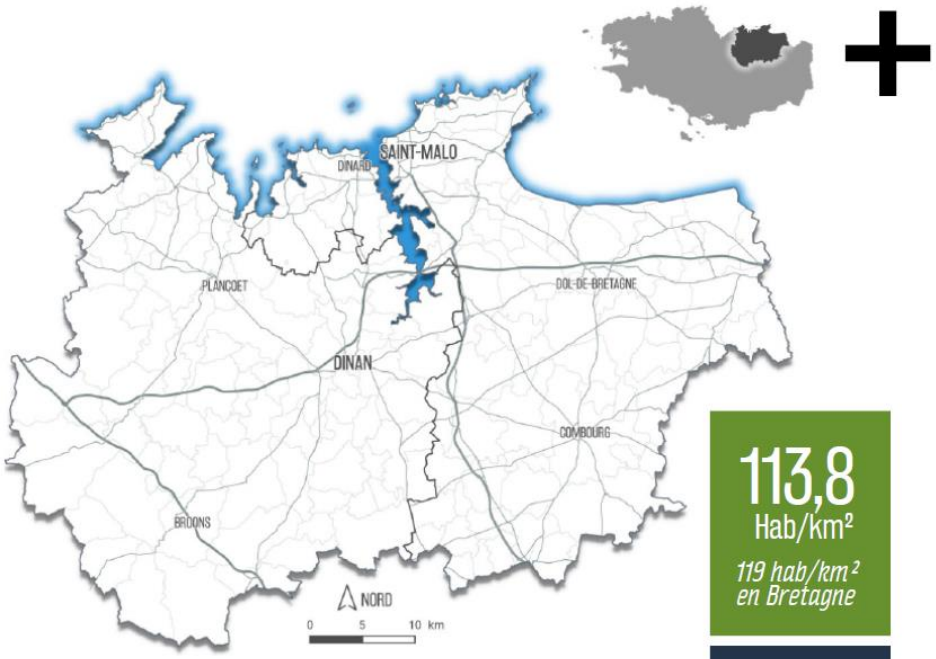
- https://budget.parlement-ouvert.fr/visualisation



De l’intérêt de combiner cartographie et visualisation de données

Exemples

TERRITOIRE
PAYS DE RANCE



113,8 Hab/km²
119 hab/km² en Bretagne

153 communes
12 EPCI

2412 km²
9% de la surface de la Bretagne

274 597 habitants
8% de la population bretonne

Sources : INSEE, IGN | Réalisation M2 SIGAT 2015 -2016

Exemples

UN MILLE-FEUILLLE ADMINISTRATIF

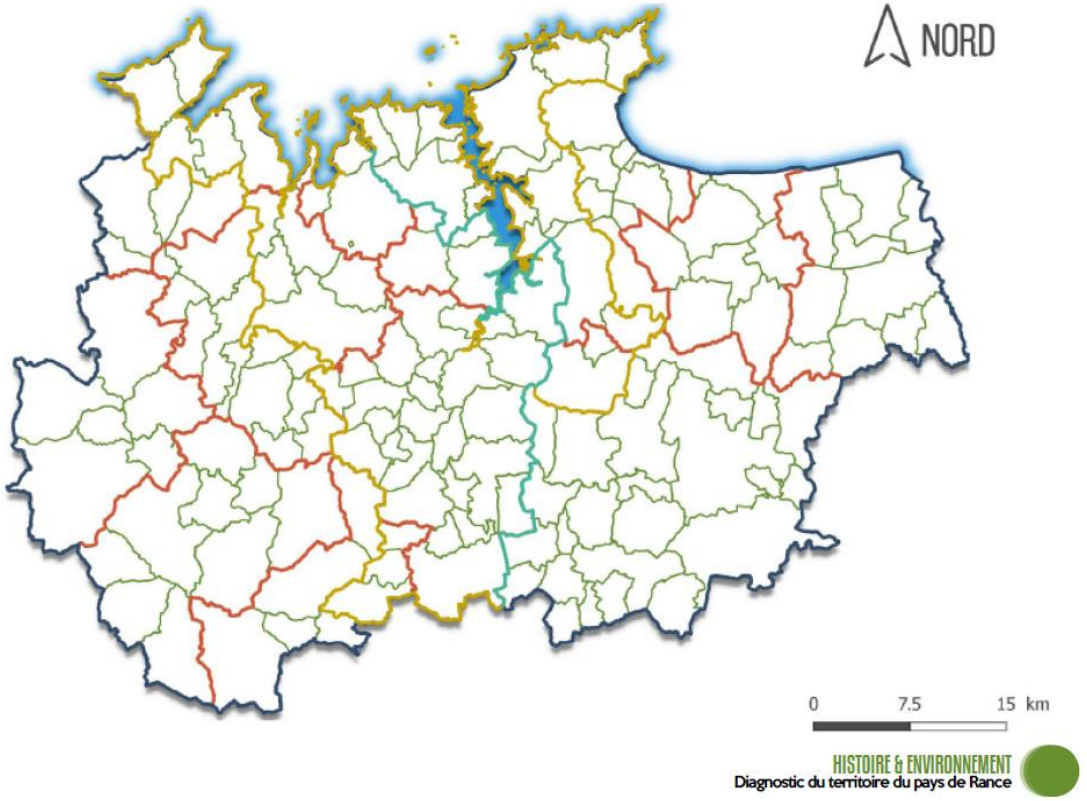
2 Départements

2 Pays

12 EPCI

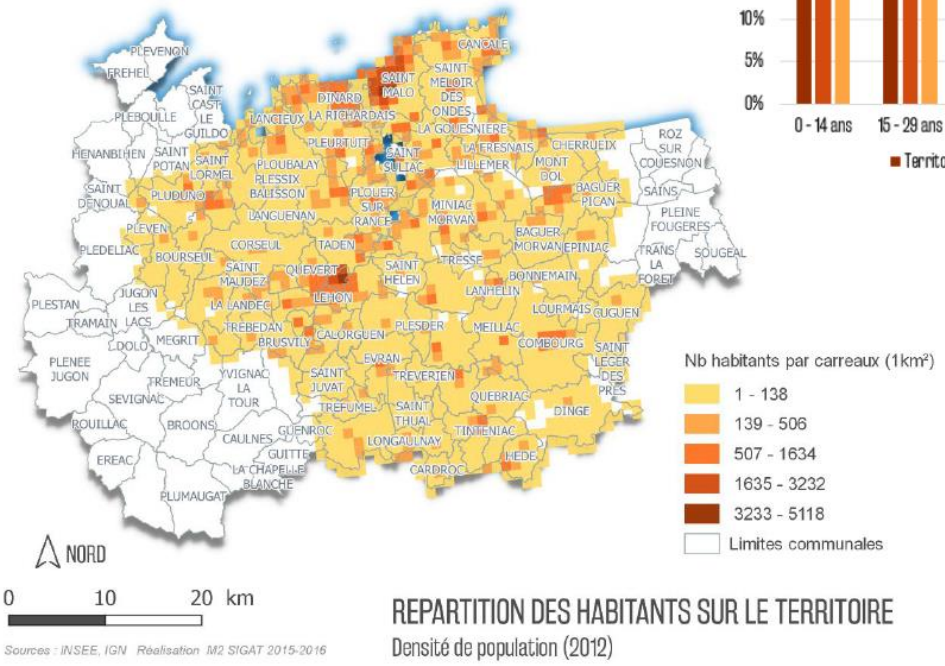
153 Communes

1 Parc naturel



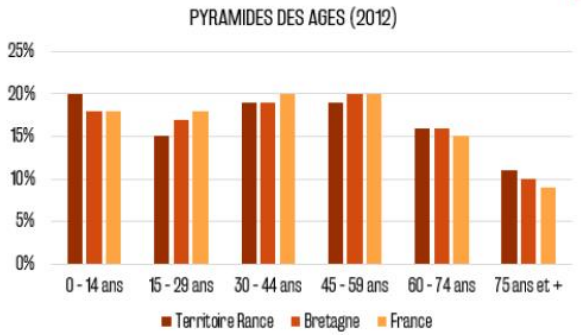
Exemples

LA DYNAMIQUE DÉMOGRAPHIQUE



Sources : INSEE, IGN Réalisation M2 SIGAT 2015-2016

REPARTITION DES HABITANTS SUR LE TERRITOIRE
Densité de population (2012)

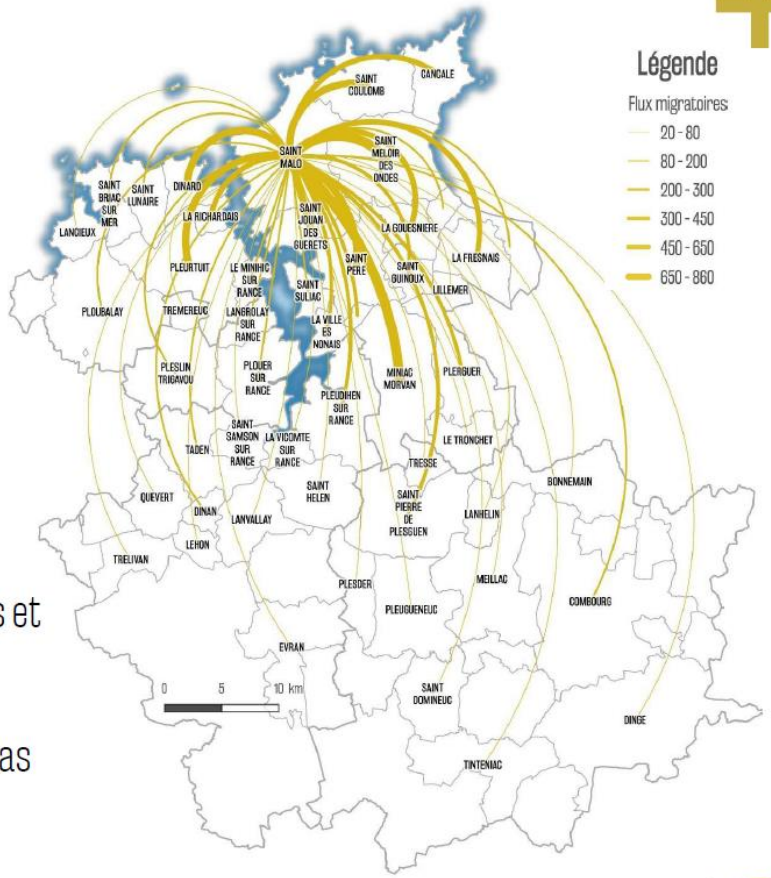


239 000 habitants

+1,3% par an

154 hab/km²

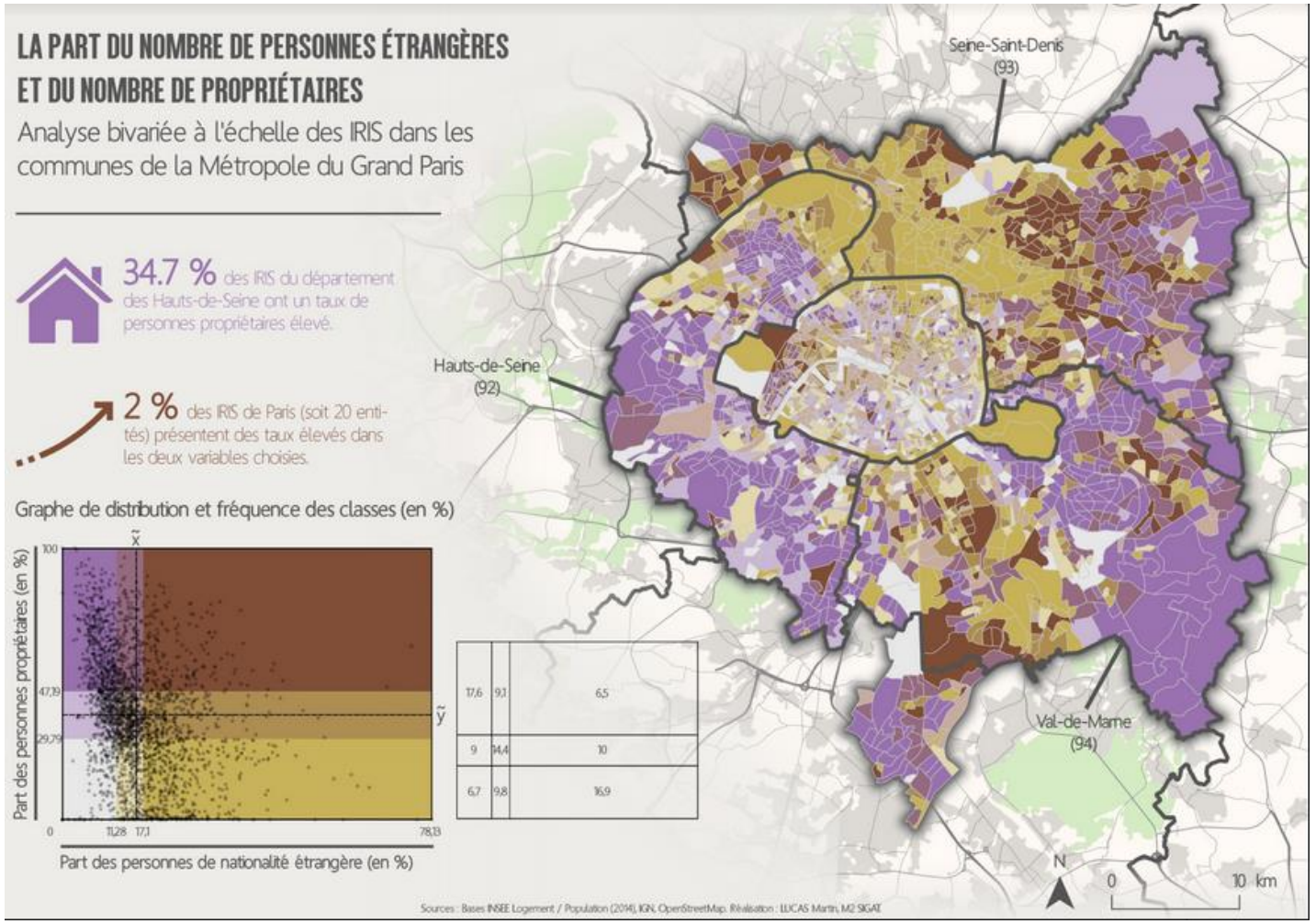
Exemples



> Incohérence entre l'offre de transports et les déplacements domicile-travail

> Les limites administratives ne bridents pas les mobilités de la population

Exemples


LA PART DU NOMBRE DE PERSONNES ÉTRANGÈRES
ET DU NOMBRE DE PROPRIÉTAIRES
Analyse bivariée à l'échelle des IRIS dans les communes de la Métropole du Grand Paris
34.7 % des IRIS du département des Hauts-de-Seine ont un taux de personnes propriétaires élevé.
2 % des IRIS de Paris (soit 20 entités) présentent des taux élevés dans les deux variables choisies.
Graphe de distribution et fréquence des classes (en %)
Part des personnes propriétaires (en %)
Part des personnes de nationalité étrangère (en %)
Seine-Saint-Denis (93)
Hauts-de-Seine (92)
Val-de-Marne (94)
N
0
10 km

Exemples

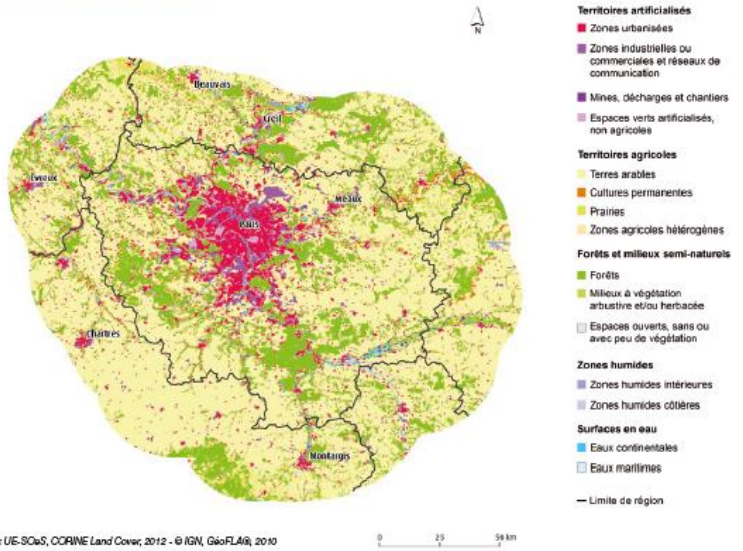
D A
TERRITOIRE
T A
L A B
Atlas régional
de l'occupation des sols
en France
OCTOBRE 2016

Exemples

Panorama général

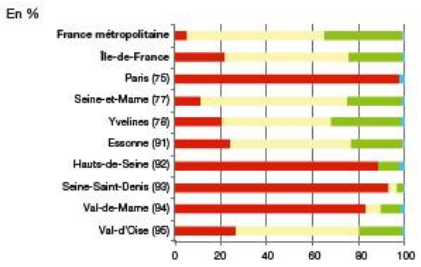
La région Île-de-France est la 1re région pour son taux d'artificialisation, la 9e région pour ses territoires agricoles, la 8e région pour ses forêts et milieux semi-naturels, la 2e région pour ses surfaces en eau et la 12e région pour ses zones humides.

OCCUPATION DES SOLS EN 2012



Sources : UE-SOeS, CORINE Land Cover, 2012 - © IGN, GéoFLA®, 2010

RÉPARTITION DES GRANDS TYPES D'OCCUPATION DES SOLS PAR DÉPARTEMENT EN 2012



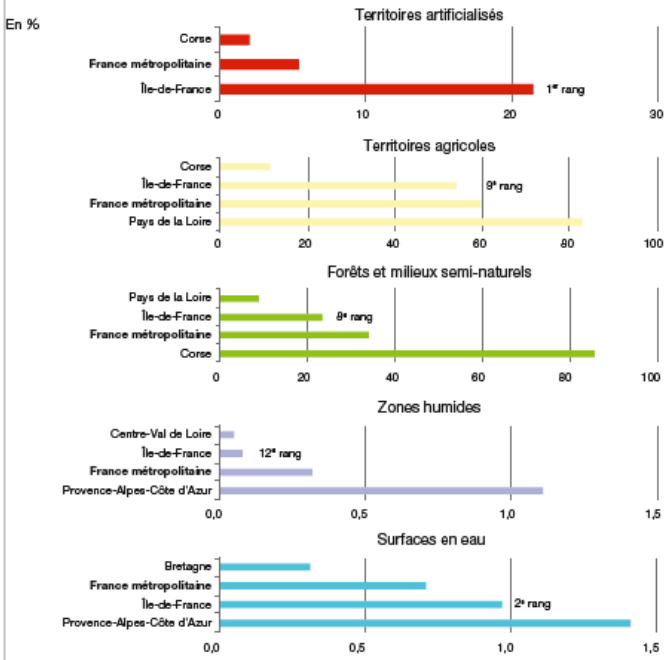
Source : UE-SOeS, CORINE Land Cover, 2012

SURFACES DANS LA RÉGION SELON LE TYPE D'OCCUPATION DES SOLS EN 2012

En ha

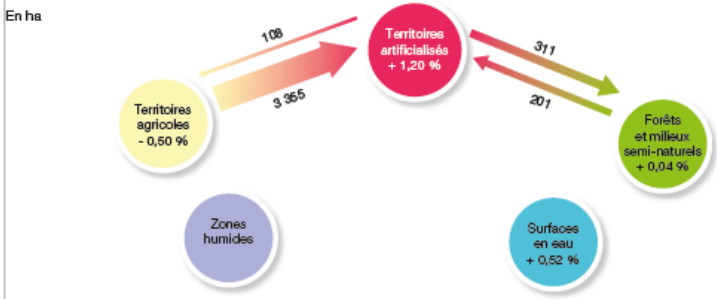
Territoires artificialisés	260 178
Territoires agricoles	650 455
Forêts et milieux semi-naturels	283 160
Zones humides	925
Surfaces en eau	11 698
Total	1 206 415

PLACE DE LA RÉGION VIS-À-VIS DES AUTRES RÉGIONS ET DE LA MÉTROPOLE EN TERMES DE RÉPARTITION DES TYPES D'OCCUPATION DES SOLS EN 2012



Note de lecture : rang décroissant avec la fraction de la superficie régionale occupée dans le thème.
Source : UE-SOeS, CORINE Land Cover, 2012

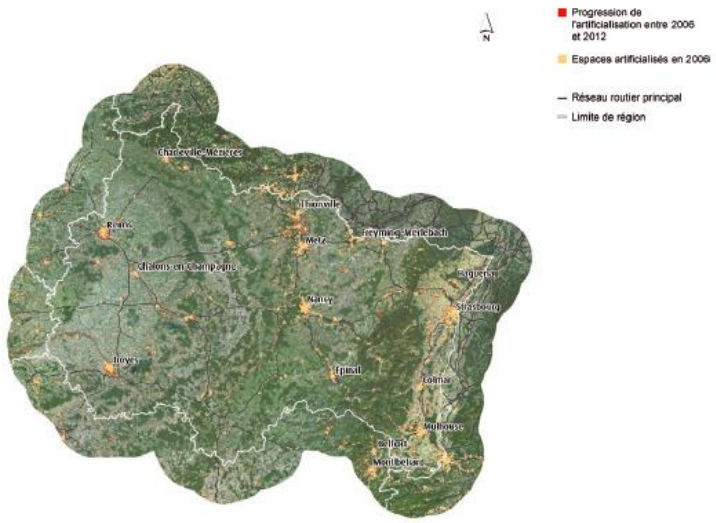
PRINCIPAUX ÉCHANGES DE SURFACES DANS LA RÉGION ENTRE 2006 ET 2012 (FLUX SUPÉRIEURS À 100 HA)



Exemples

Artificialisation des sols

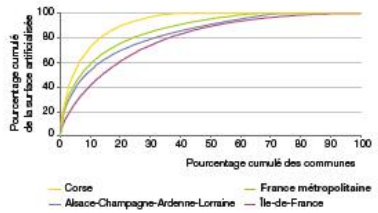
PROGRESSION DE L'ARTIFICIALISATION ENTRE 2006 ET 2012



Sources : UE-SOeS, CORINE Land Cover 2006 révisée & base des changements 2006-2012 - © IGN, GéoFLA®, 2010 - © IGN, Route 500®, 2015 et © IGN, photographies aériennes, 2015

CONCENTRATION DE L'ARTIFICIALISATION PAR COMMUNE EN 2012

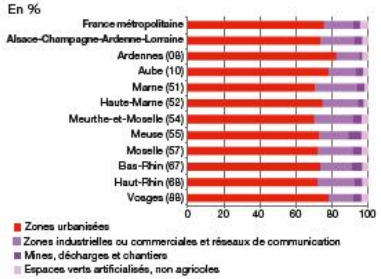
En %



Note de lecture : en Alsace-Champagne-Ardenne-Lorraine, la moitié de la surface artificialisée est concentrée dans 8 % des communes. La moitié des communes de métropole concentre 95 % de la surface artificialisée.

RÉPARTITION DES SURFACES ARTIFICIALISÉES PAR DÉPARTEMENT EN 2012

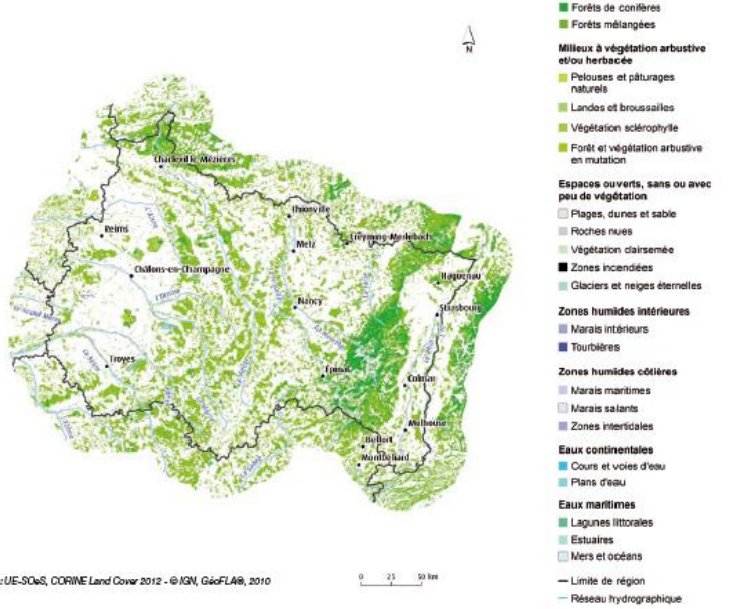
En %



Source : UE-SOeS, CORINE Land Cover, 2012

Milieux naturels, forestiers et aquatiques

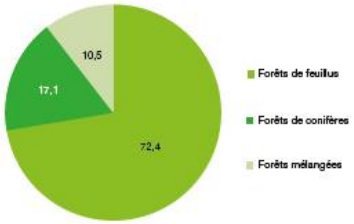
DIVERSITÉ DES MILIEUX NATURELS ET SEMI-NATURELS EN 2012



Sources : UE-SOeS, CORINE Land Cover 2012 - © IGN, GéoFLA®, 2010

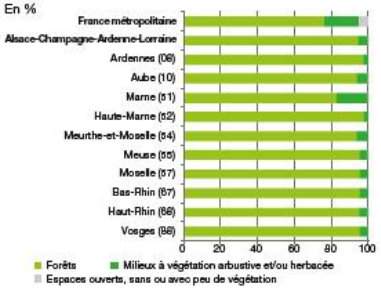
RÉPARTITION DES TYPES DE FORÊTS EN 2012

En %



RÉPARTITION DES FORÊTS ET MILIEUX SEMI-NATURELS PAR DÉPARTEMENT EN 2012

En %

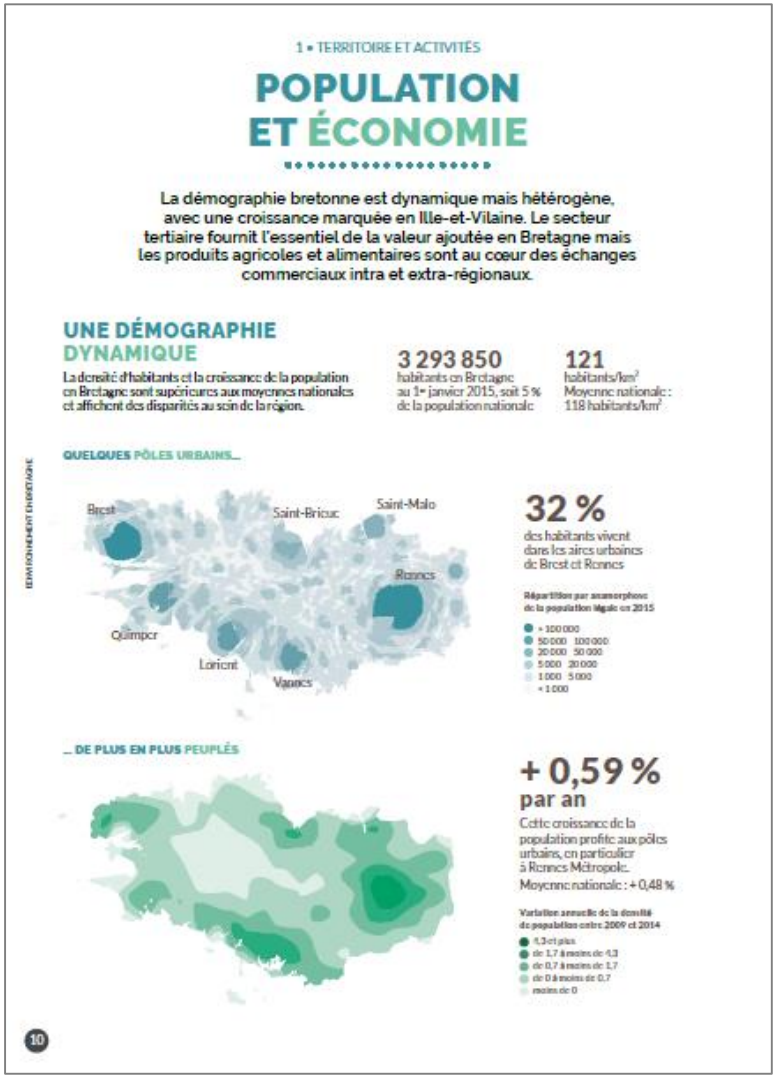


Source : UE-SOeS, CORINE Land Cover, 2012

Exemples

L'ENVIRONNEMENT
EN BRETAGNE
Cartes et chiffres clés 2018
OEB
OBSERVATOIRE DE L'ENVIRONNEMENT EN BRETAGNE

Exemples



1 • TERRITOIRE ET ACTIVITÉS

POPULATION ET ÉCONOMIE

La démographie bretonne est dynamique mais hétérogène, avec une croissance marquée en Ille-et-Vilaine. Le secteur tertiaire fournit l'essentiel de la valeur ajoutée en Bretagne mais les produits agricoles et alimentaires sont au cœur des échanges commerciaux intra et extra-régionaux.

UNE DÉMOGRAPHIE DYNAMIQUE

La densité d'habitants et la croissance de la population en Bretagne sont supérieures aux moyennes nationales et affichent des disparités au sein de la région.

3 293 850
habitants en Bretagne au 1er janvier 2015, soit 5 % de la population nationale

121
habitants/km²
Moyenne nationale : 118 habitants/km²

QUELQUES PÔLES URBAINS...

32 %
des habitants vivent dans les aires urbaines de Brest et Rennes

Répartition par anamorphose de la population légale en 2015

- + 100 000
- 50 000 100 000
- 20 000 50 000
- 5 000 20 000
- 1 000 5 000
- < 1 000

... DE PLUS EN PLUS PEUPLÉS

+ 0,59 % par an

Cette croissance de la population profite aux pôles urbains, en particulier à Rennes Métropole.

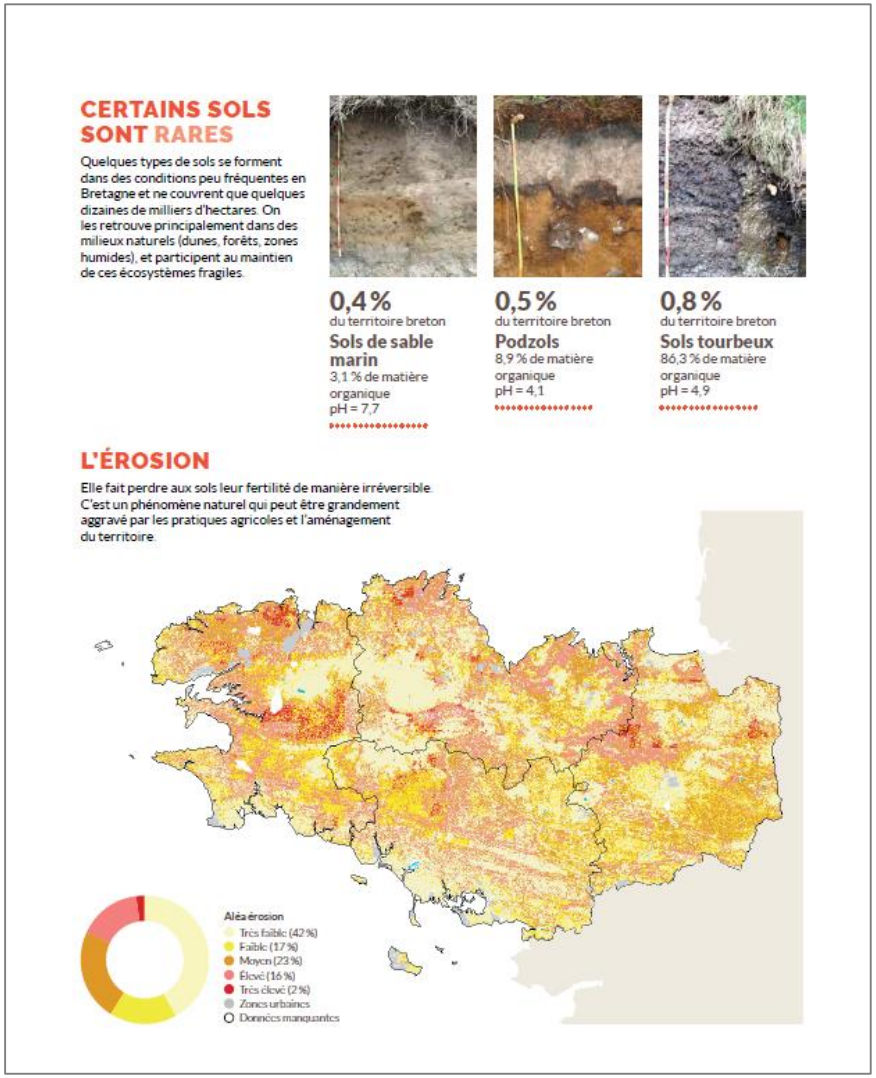
Moyenne nationale : + 0,48 %

Variation annuelle de la densité de population entre 2009 et 2014

- 1,3 et plus
- de 1,7 à moins de 1,3
- de 0,7 à moins de 1,7
- de 0 à moins de 0,7
- moins de 0

ENVIRONNEMENT ÉNERGIE

10



CERTAINS SOLS SONT RARES

Quelques types de sols se forment dans des conditions peu fréquentes en Bretagne et ne couvrent que quelques dizaines de milliers d'hectares. On les retrouve principalement dans des milieux naturels (dunes, forêts, zones humides), et participent au maintien de ces écosystèmes fragiles.

0,4 %
du territoire breton
Sols de sable marin
3,1 % de matière organique
pH = 7,7

0,5 %
du territoire breton
Podzols
8,9 % de matière organique
pH = 4,1

0,8 %
du territoire breton
Sols tourbeux
86,3 % de matière organique
pH = 4,9

L'ÉROSION

Elle fait perdre aux sols leur fertilité de manière irréversible. C'est un phénomène naturel qui peut être grandement aggravé par les pratiques agricoles et l'aménagement du territoire.

Exemples

1 • TERRITOIRE ET ACTIVITES

RESSOURCES ET ACTIVITÉS

Alors que certaines ressources naturelles comme l'eau et les matières minérales sont utilisées par toute la population et des secteurs d'activité variés, certaines ressources sont plus spécifiques des secteurs agricoles, sylvicoles et de la pêche.

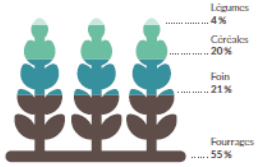
LES RESSOURCES AGRICOLES

60 % du territoire breton est consacré à l'agriculture et produit près de 24 millions de tonnes de produits agricoles.

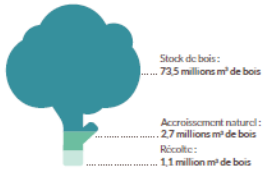
LES RESSOURCES FORESTIÈRES

Avec 14 % de forêt, la Bretagne est peu boisée au regard du reste de la France (moyenne nationale : 30 %). Néanmoins, elle voit son stock de bois croître chaque année car la récolte annuelle est inférieure à l'accroissement naturel.

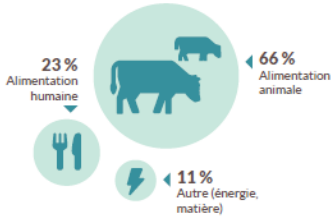
LES PRODUCTIONS VÉGÉTALES AGRICOLES EN BRETAGNE



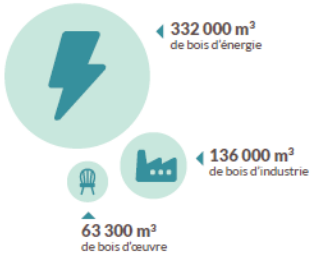
LA PRODUCTION DE BOIS EN BRETAGNE



LES USAGES DES CULTURES EN BRETAGNE



LES USAGES DU BOIS EN BRETAGNE



4 • L'EAU

LE CYCLE DE L'EAU DOMESTIQUE

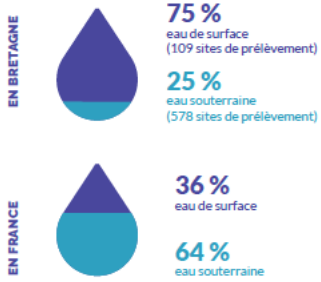
Le cycle de l'eau domestique en Bretagne a plusieurs spécificités. L'eau est prélevée en surface plutôt que dans des nappes profondes. L'assainissement autonome est très utilisé. Et le prix de l'eau est plus élevé qu'en moyenne en France.

LA RESSOURCE EN EAU

Pour ses besoins domestiques, l'homme interfère avec le grand cycle de l'eau en prélevant de l'eau en surface ou en profondeur. Avec cette eau brute, il produit de l'eau potable qui est distribuée jusqu'au robinet. Une fois utilisée, les eaux usées sont assainies avant de retourner dans le milieu naturel.

Les zones de captage d'eau brute (périmètres de protection) font l'objet d'actions renforcées pour garantir leur conformité sanitaire, notamment en nitrates et pesticides.

NIVEAU DE CONFORMITÉ PAR RAPPORT AUX LIMITES RÉGLEMENTAIRES « EAU BRUTE » EN BRETAGNE



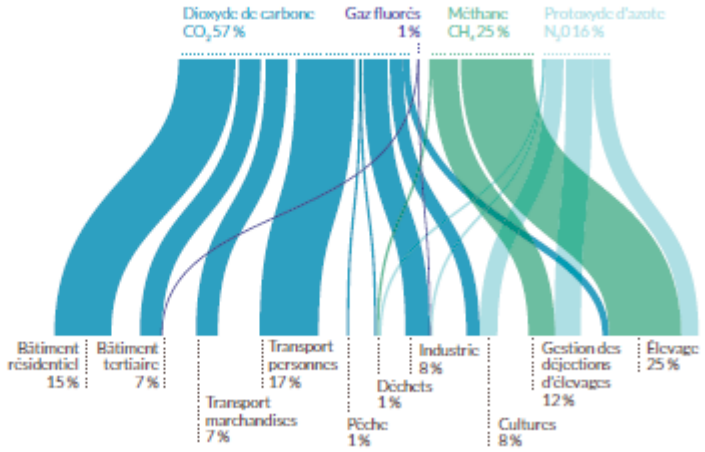
687 sites de captage d'eau brute (c'est-à-dire avant potabilisation) sont reliés à un maillage d'interconnexions de réseaux garantissant la sécurité de l'alimentation en eau potable de la région.

5 sites de captage d'eau brute alimentent 20 % de la population

Exemples

LES ÉMISSIONS RÉGIONALES

En excluant les émissions indirectes, générées par la production des biens de consommation importés, les émissions de gaz à effet de serre en Bretagne viennent surtout de la consommation d'énergie et des cheptels.

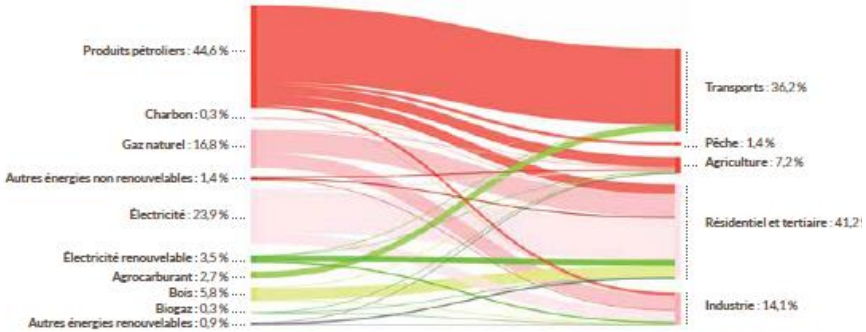


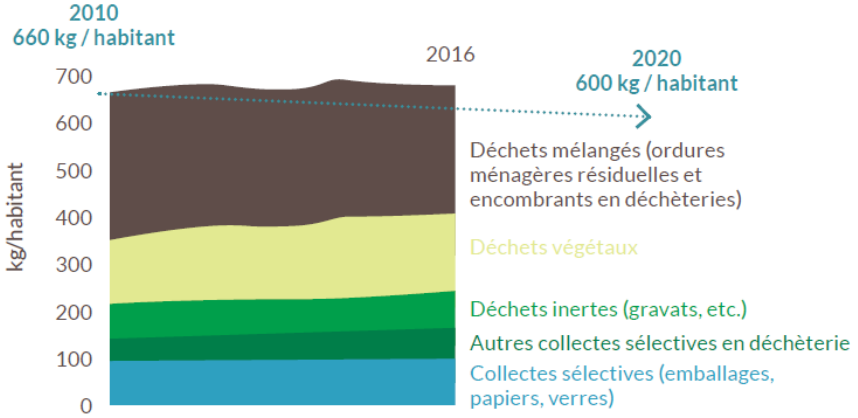
QUI CONSOMME QUOI ?

La Bretagne a recourt à de nombreuses sources d'énergie, en majorité d'origine fossile ou nucléaire, et de ce fait importées. L'importance de la consommation s'explique notamment par un bâti thermiquement peu performant, la prédominance des transports routiers et l'étalement urbain.

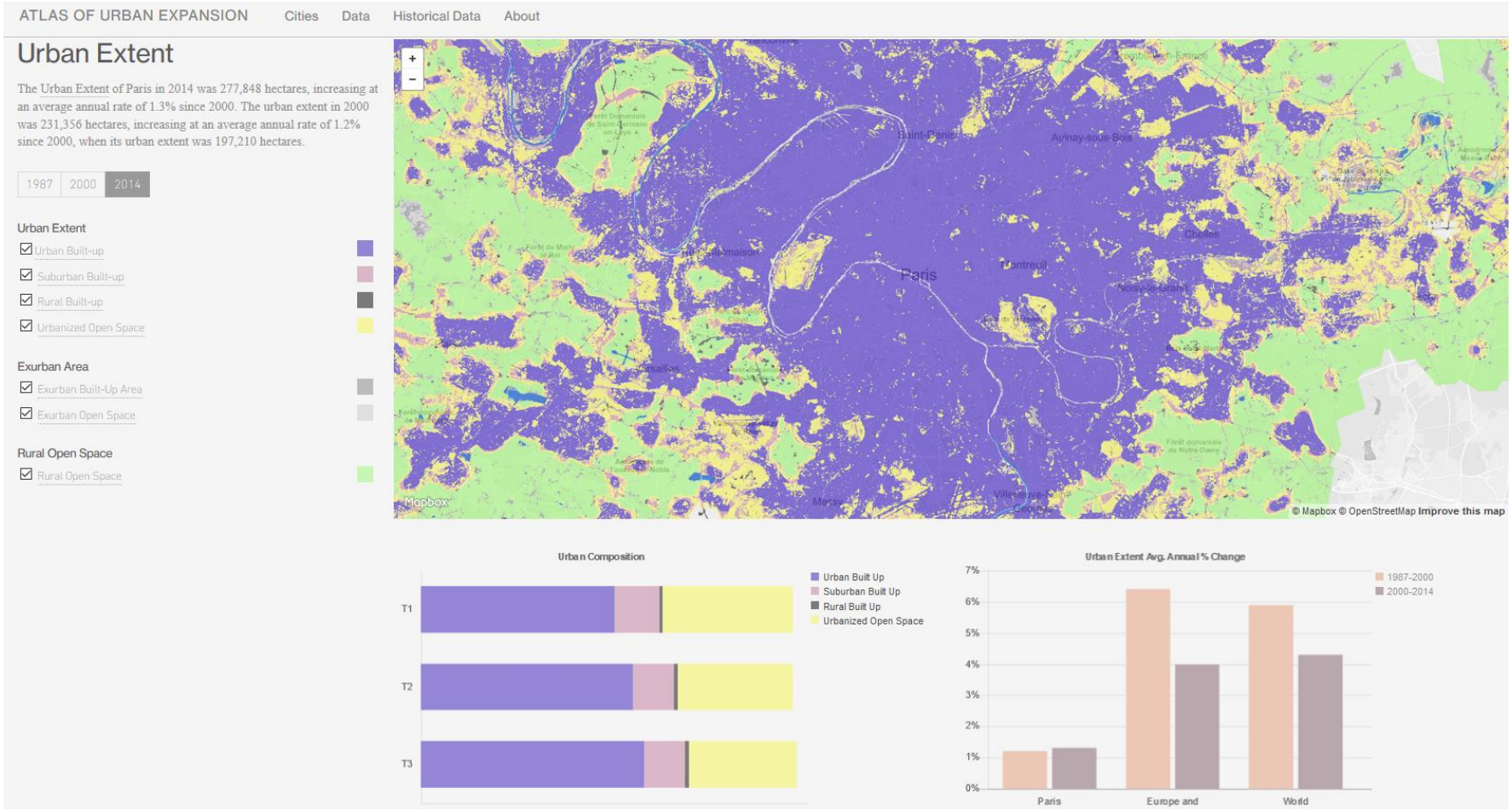
x 2 en 15 ans

La consommation d'énergie renouvelable a doublé entre 2000 et 2015 pour atteindre environ 13 %.



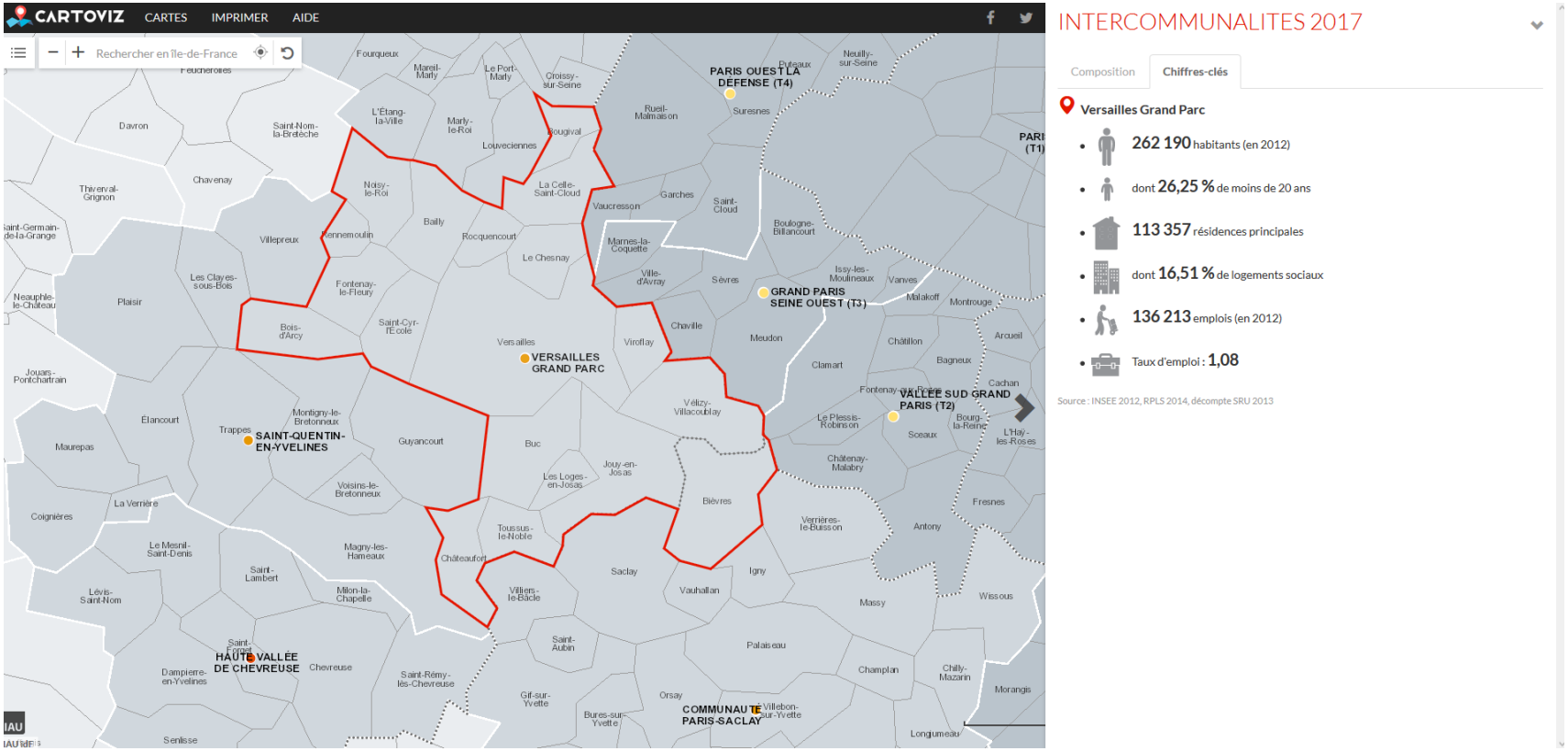


Cartographie en ligne



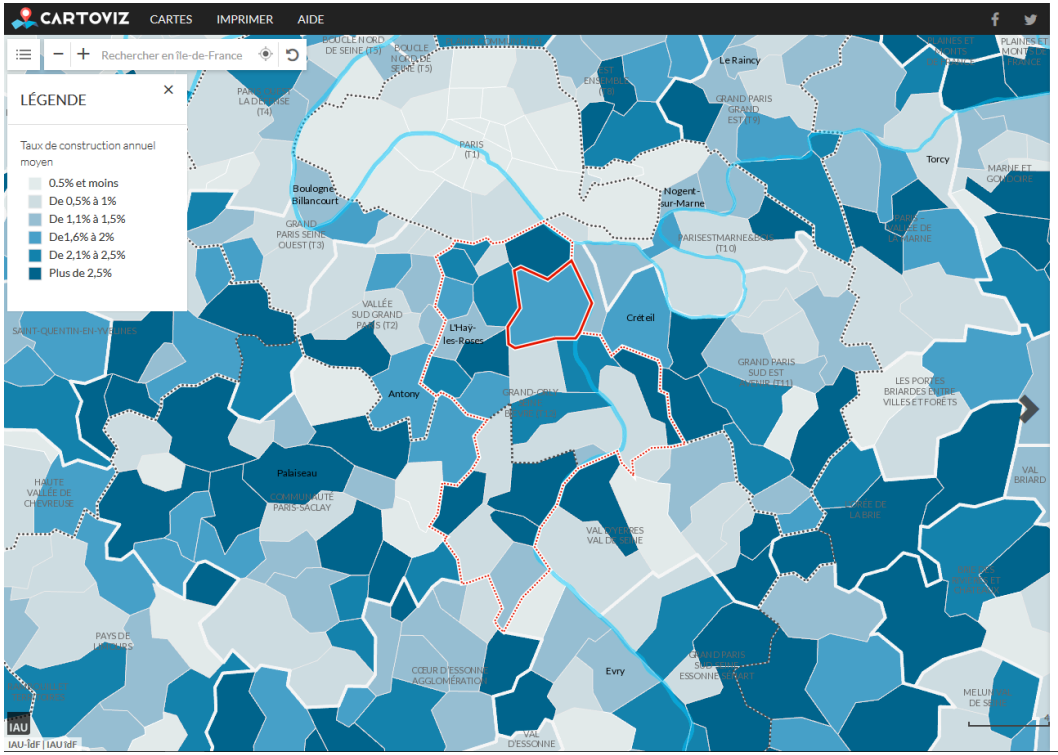
http://atlasofurbanexpansion.org/cities/view/Le_Mans

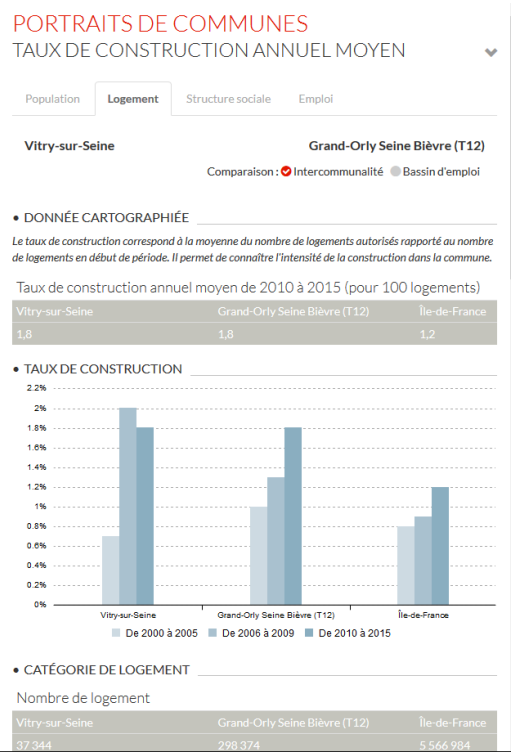
Cartographie en ligne



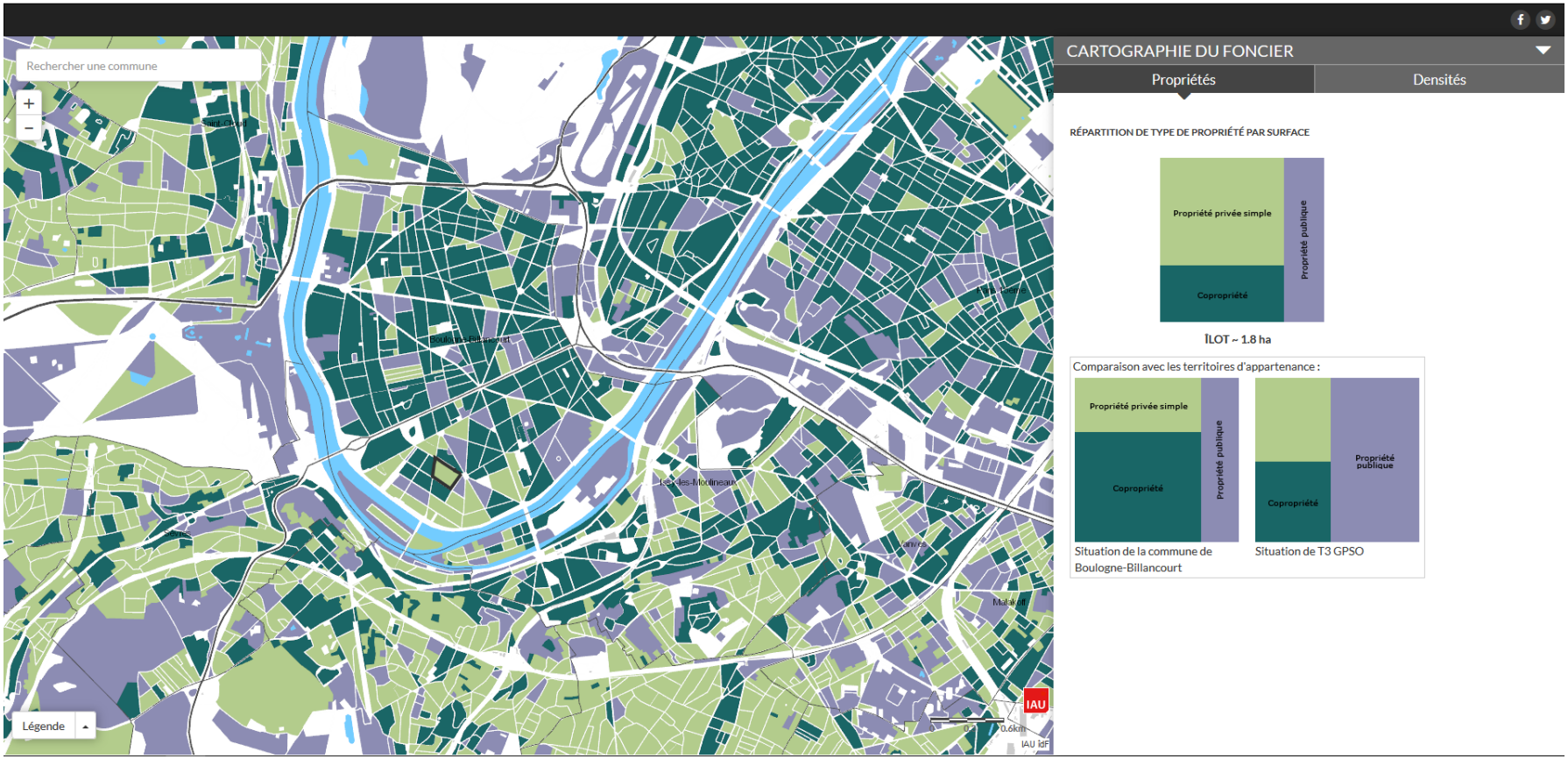
https://cartoviz.iau-idf.fr/

Cartographie en ligne



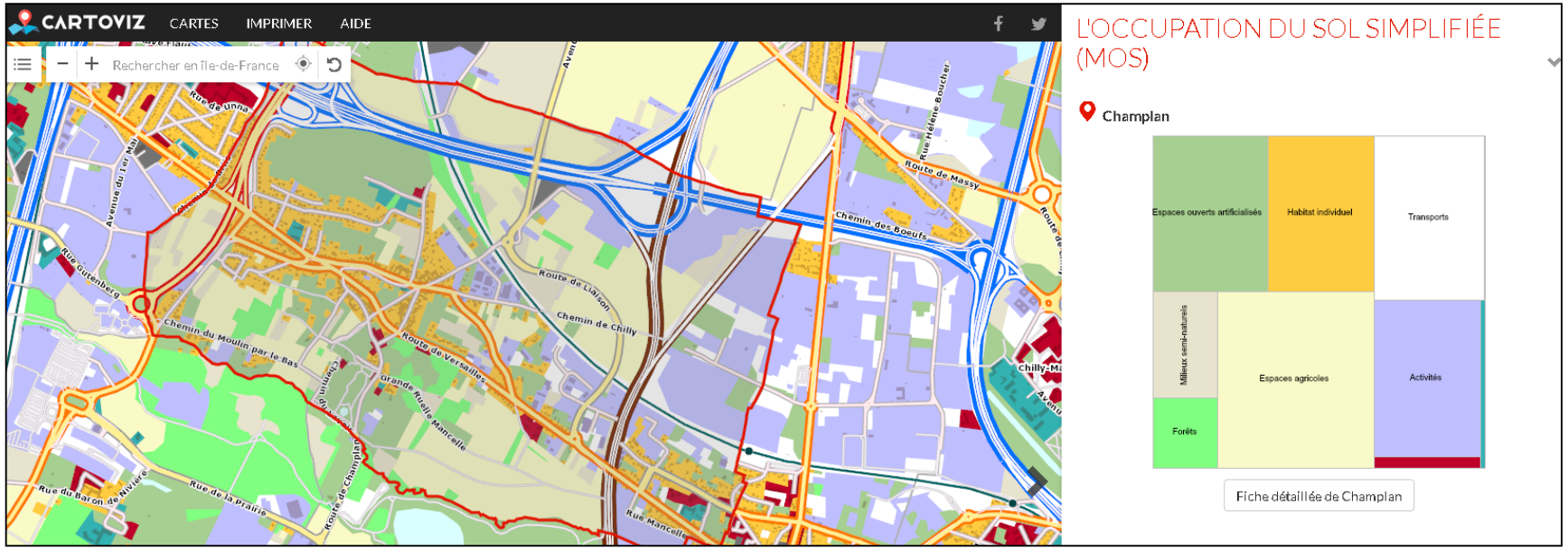


Cartographie en ligne

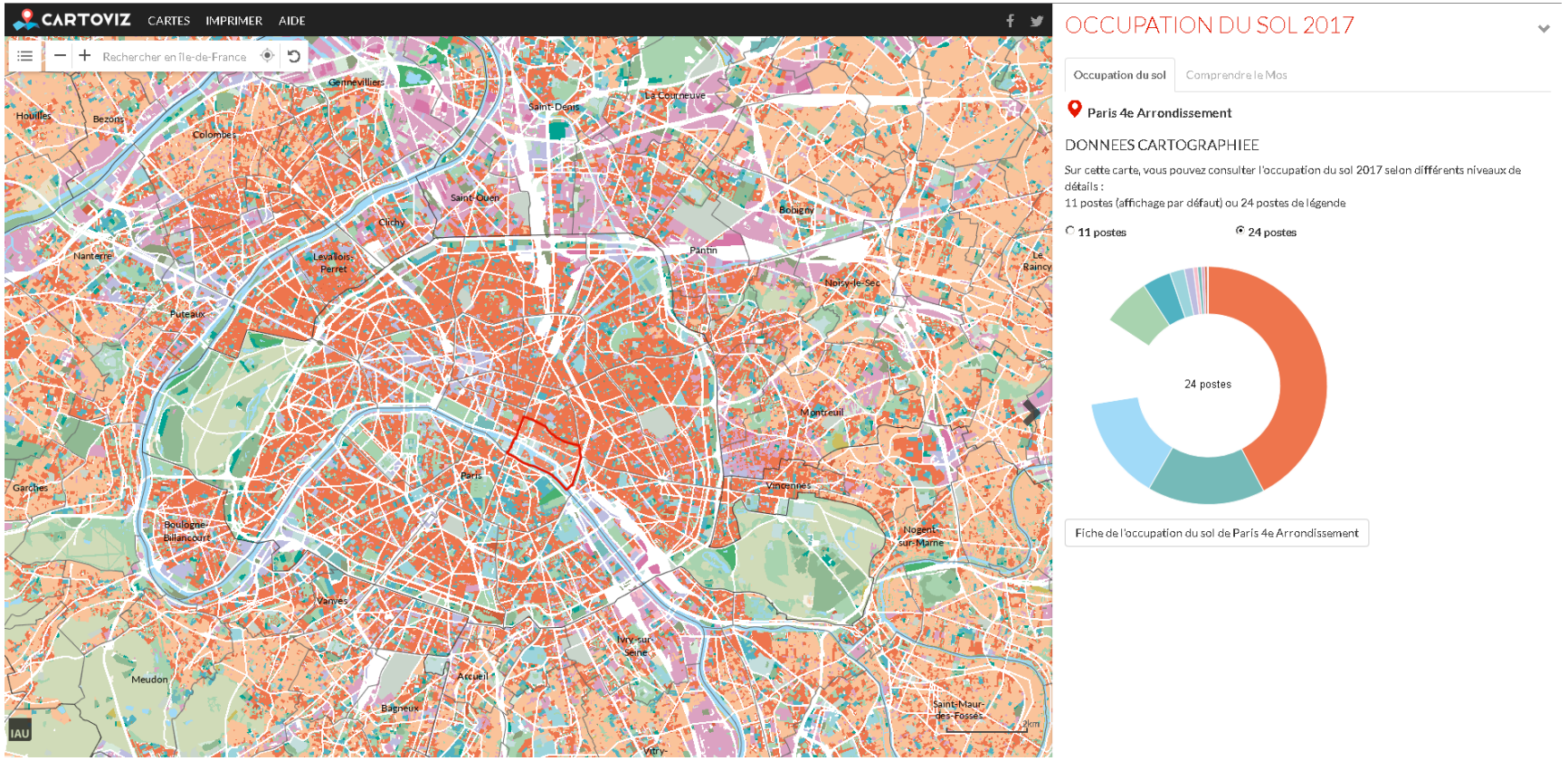


http://carto.iau-idf.fr/webapps/foncier/

Cartographie en ligne

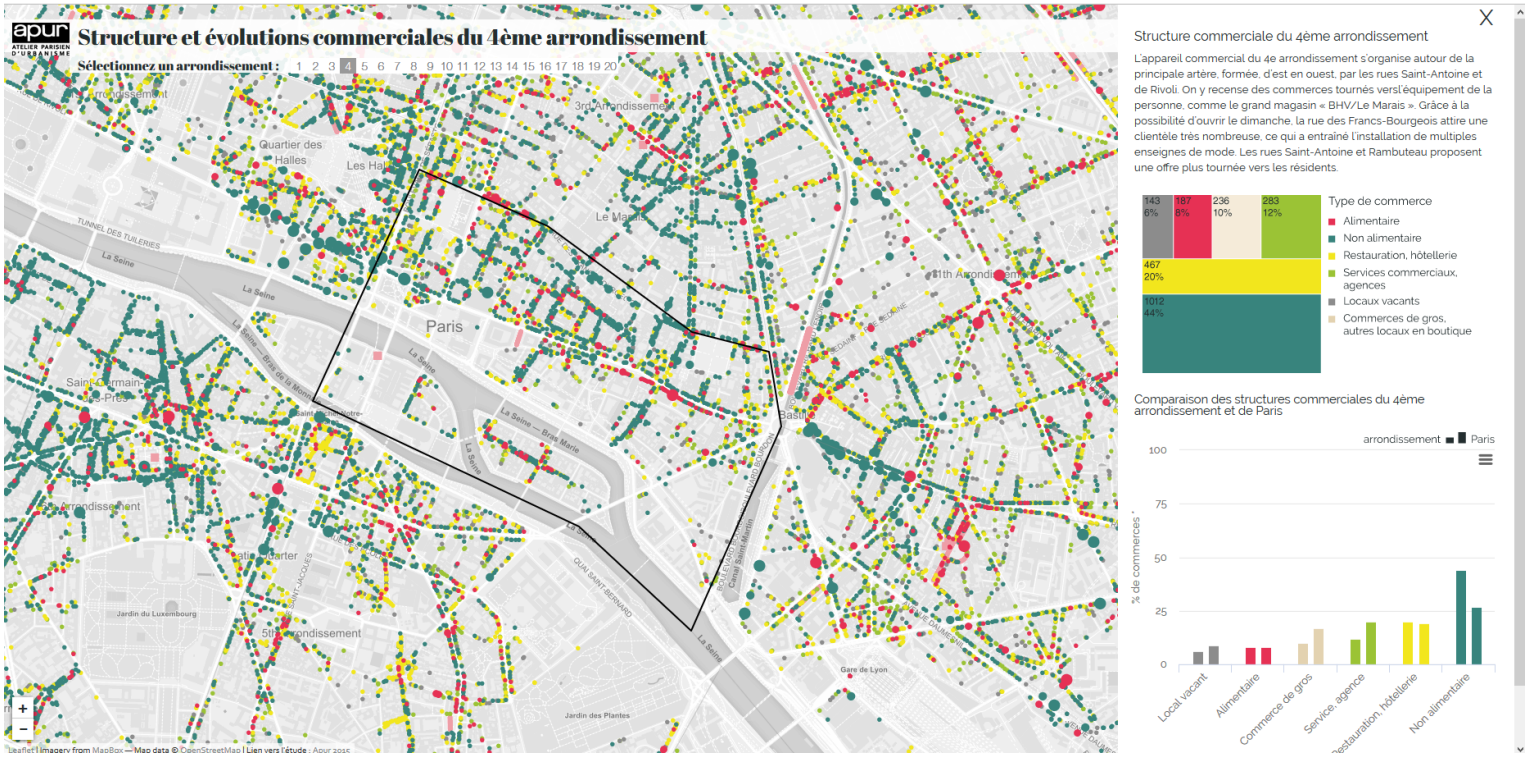


Cartographie en ligne



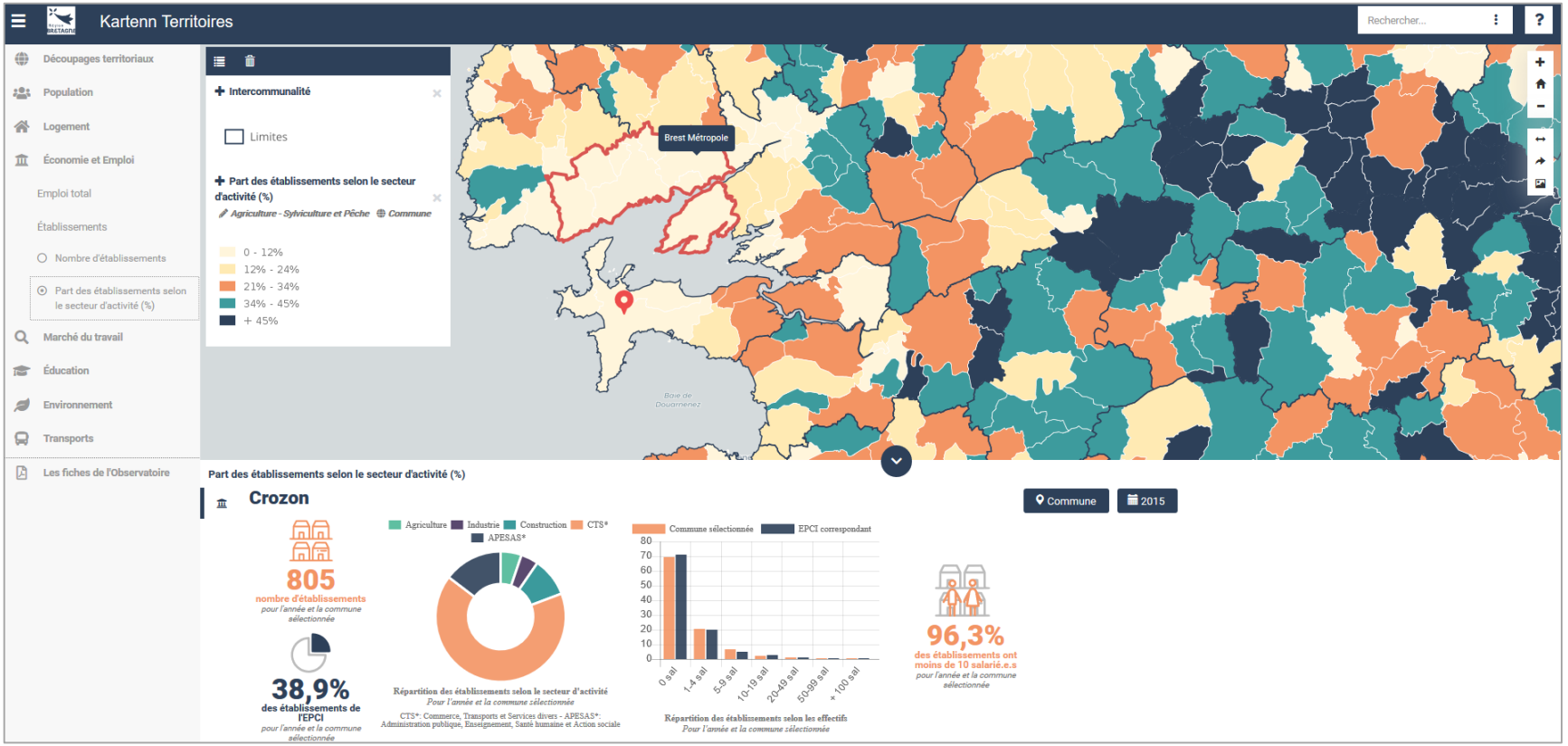
https://cartoviz.iau-idf.fr/?id_appli=Mos2017&x=652436.1421903754&y=6863112.9568629265&zoom=8

Cartographie en ligne



http://www.apur.org/dataviz/fiches_commerciales/index.html#

Cartographie en ligne

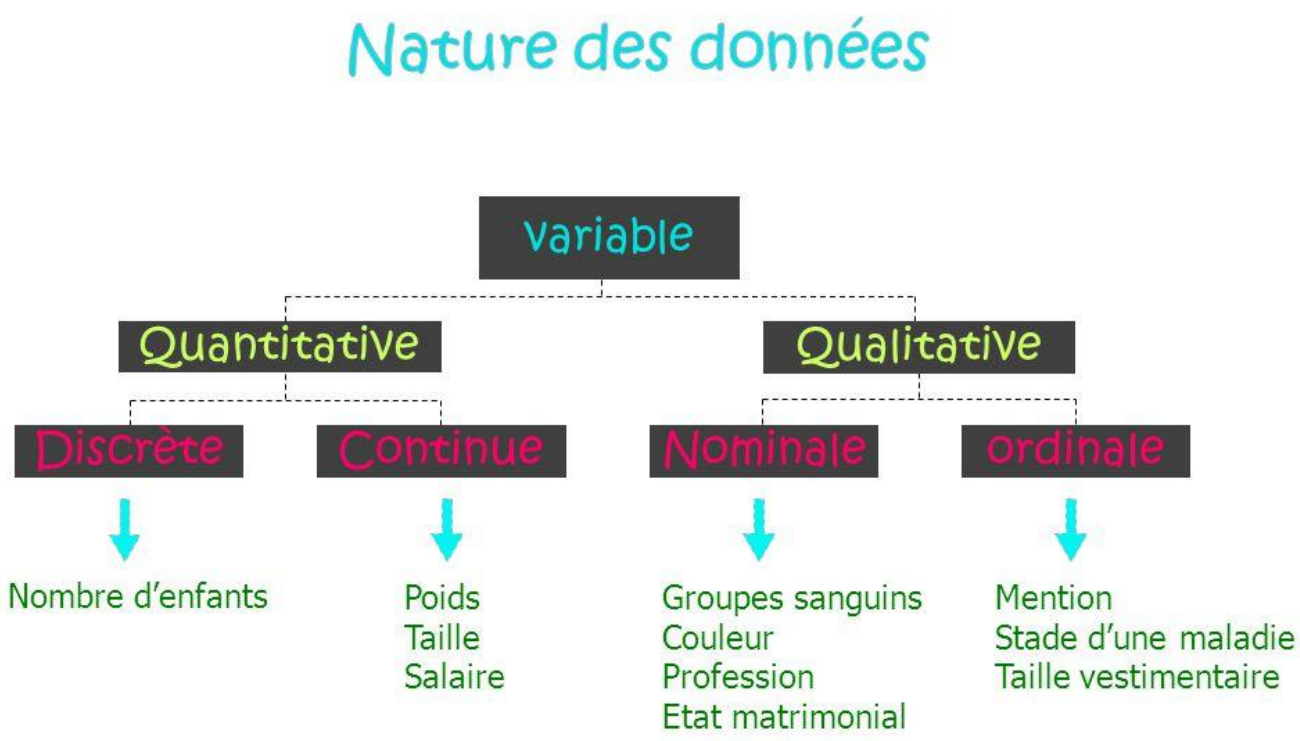


http://kartenn.region-bretagne.fr/territoires/

+

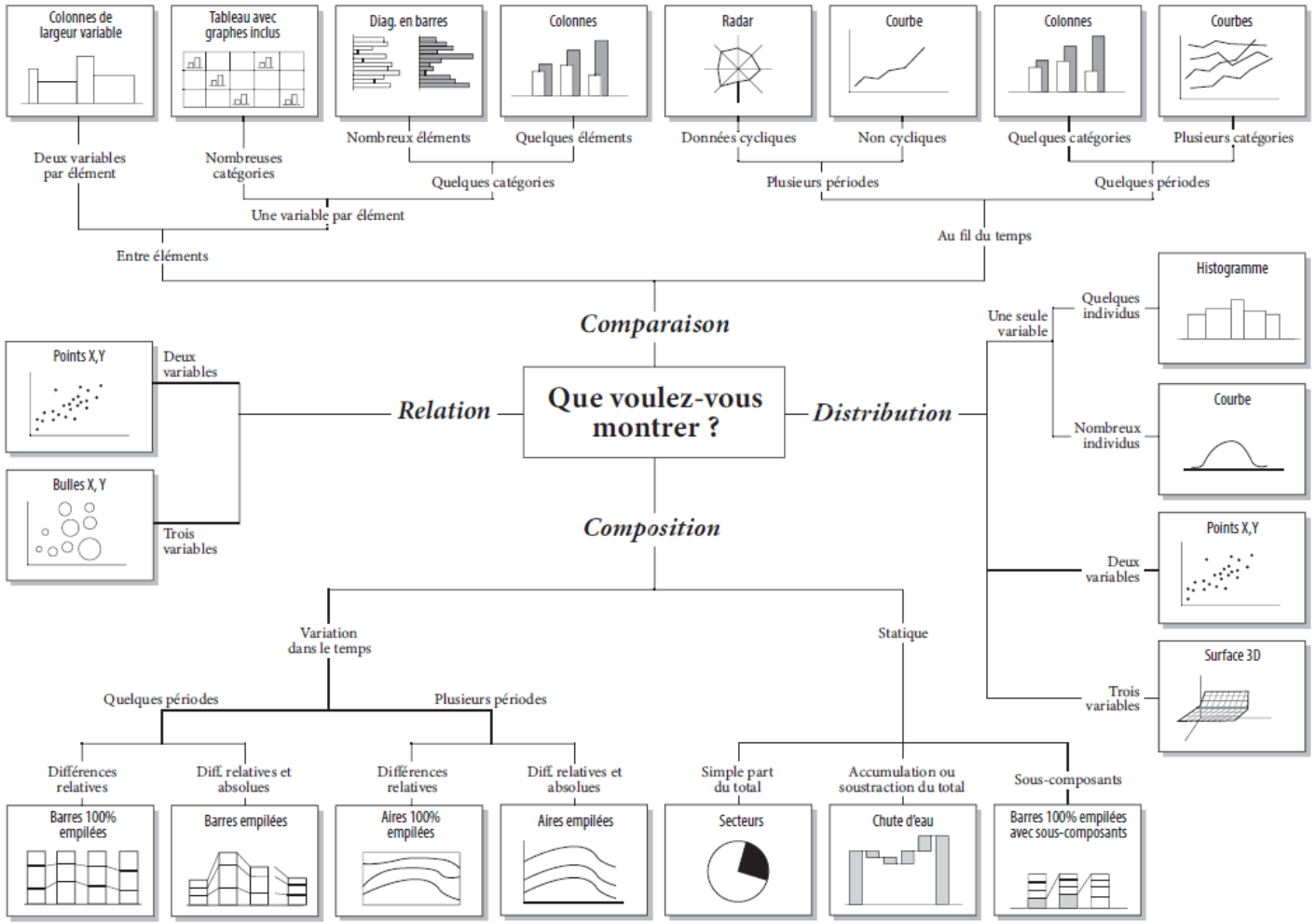
Quels types de dataviz ?

Différents types de données



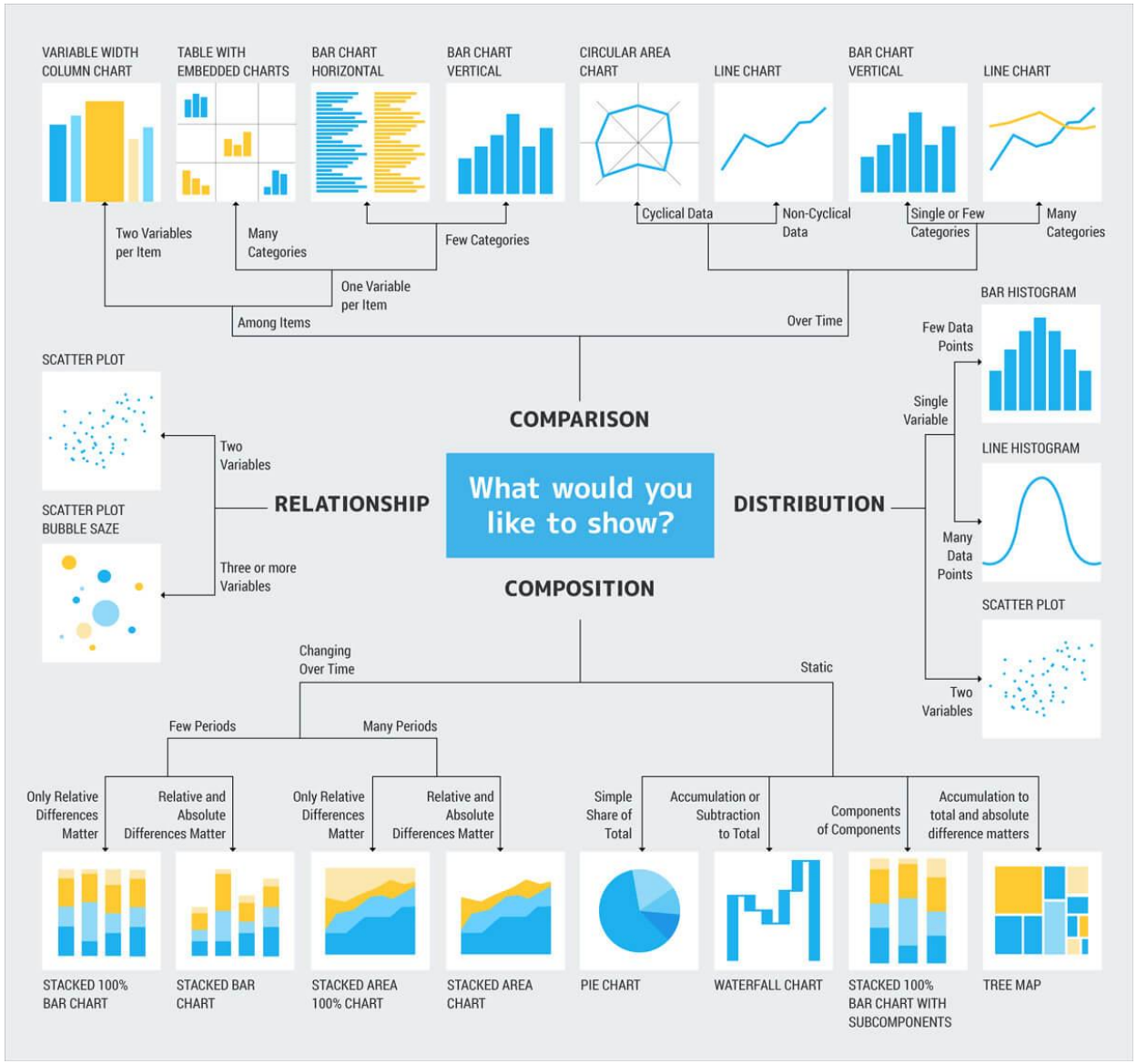
La base

Arbre de décision des graphiques



www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com
Traduction : L. Jégou, 2018

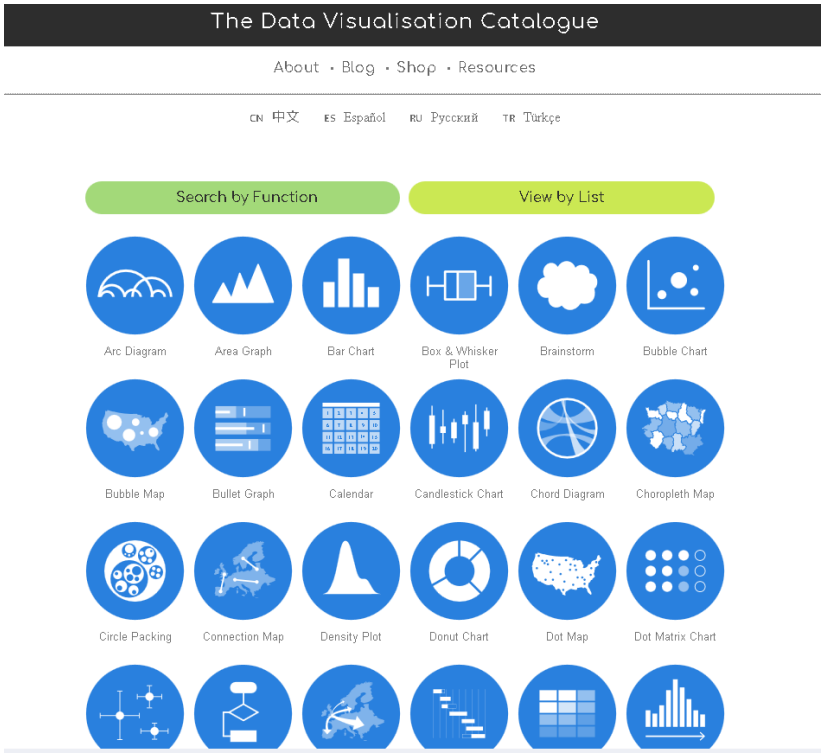
La base

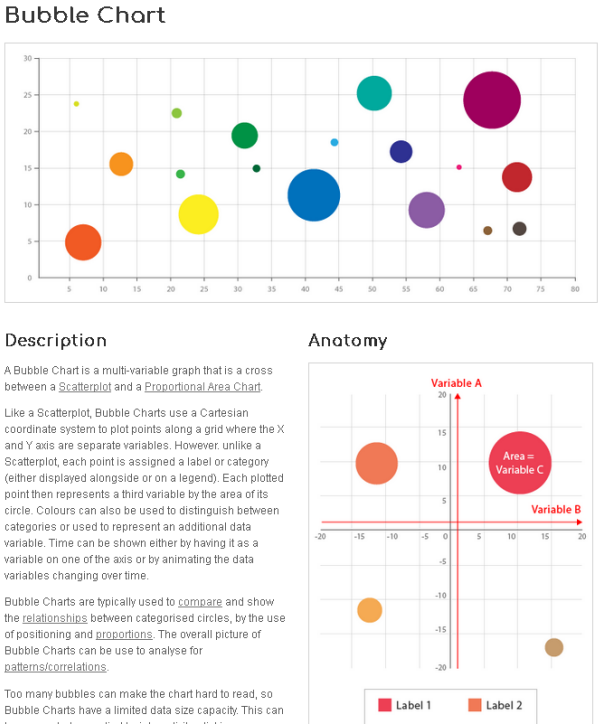

What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
Two Variables per Item
Many Categories
Few Categories
One Variable per Item
Among Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Over Time
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
SCATTER PLOT
BUBBLE SAZE
Three or more Variables
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP

Des outils pour choisir

Des outils pour choisir

- https://datavizcatalogue.com/search.html

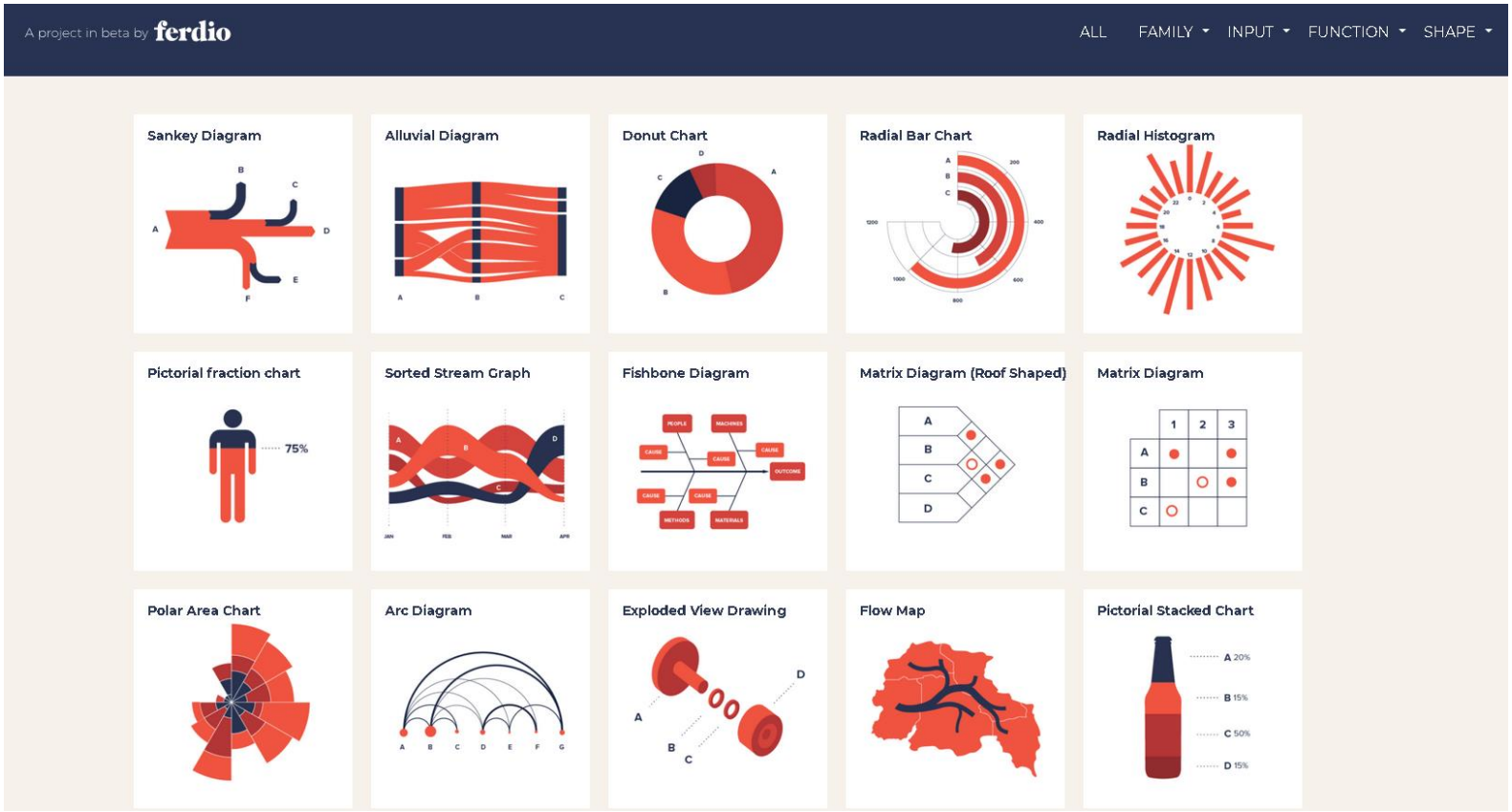




Des outils pour choisir

Des outils pour choisir

- https://datavizproject.com/



Des outils pour choisir

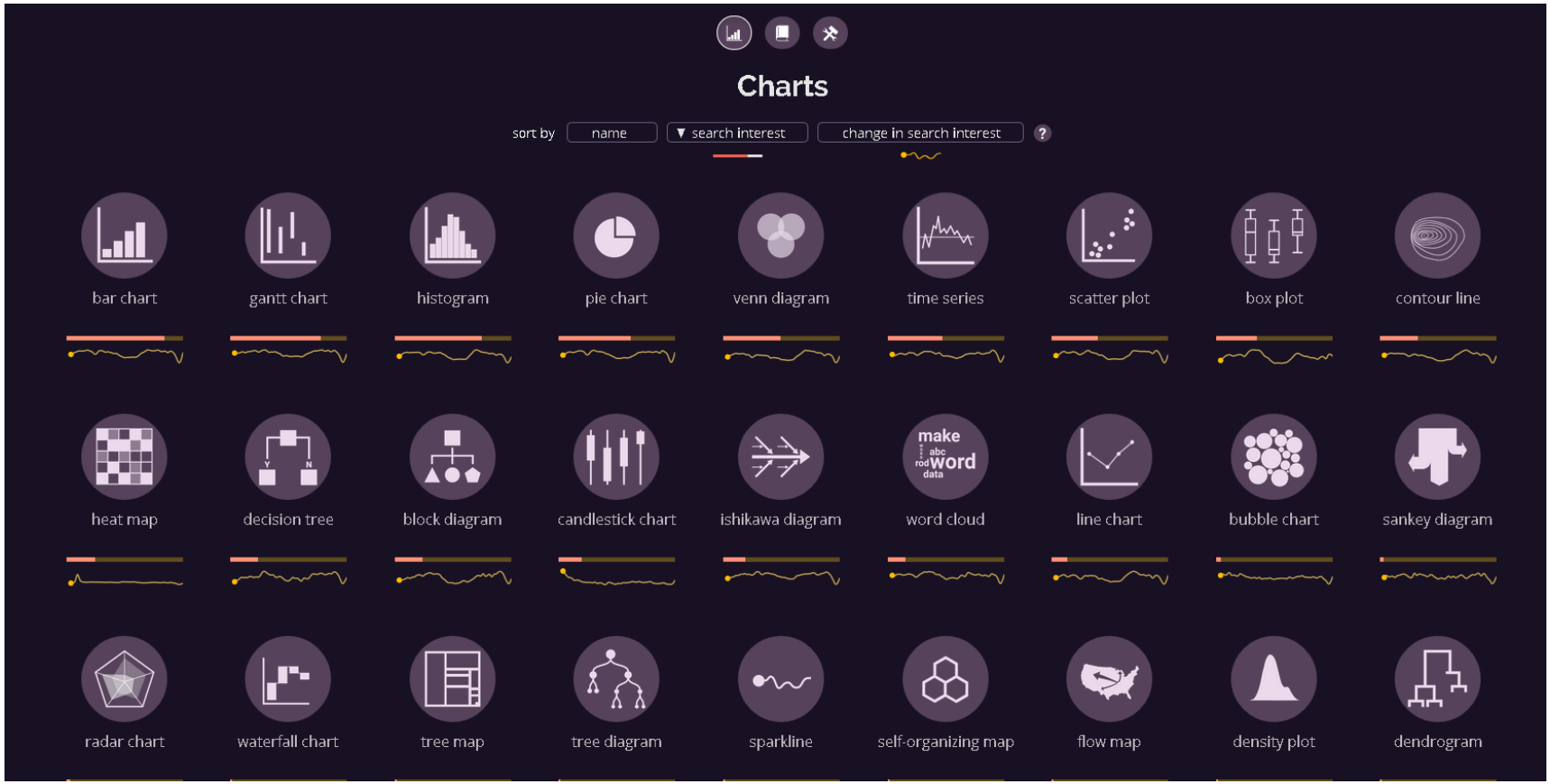
- https://datavizproject.com/

Des outils pour choisir

- https://datavizproject.com/

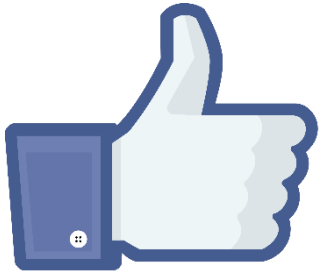
Des outils pour choisir

- http://visualizationuniverse.com/charts/

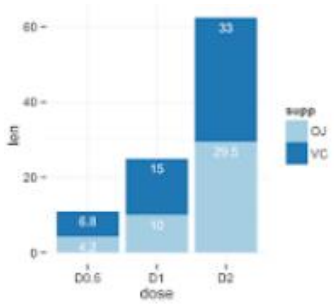


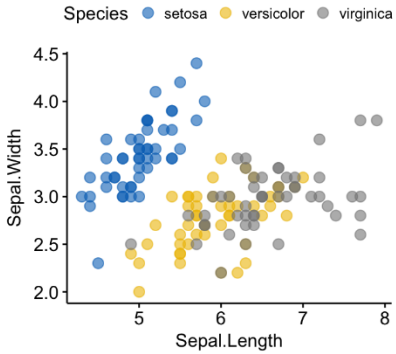
Les clef pour un bonne Dataviz

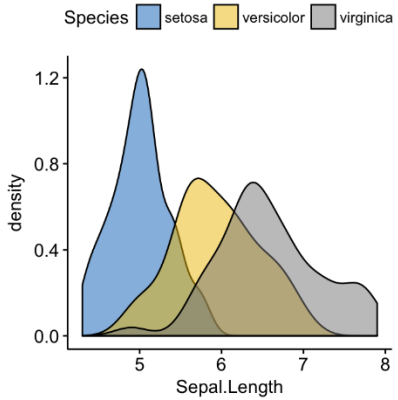
- Le choix des sources de données (primordial)
  - Fiabilité de la source des données
- Sélectionner les données les plus intéressantes
  - En fonction de la finalité visée par la dataviz
- Bien structurer les données
  - Variables qualitatives et/ou quantitatives
  - Jouer avec les tableaux croisés dynamiques !
- La construction d'indicateurs pertinents
  - Mobilisation de méthodes statistiques (descriptives, univariées,…)
- Un mode de visualisation adapté au message et aux données
  - Plusieurs type de visualisation
- Des couleurs attrayantes
- Le bon médium de diffusion (présentation, rapport, site Web, blog, réseaux sociaux,…)

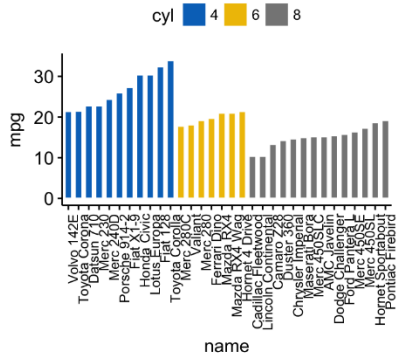


Quelles solutions techniques ?









Species	length	mean	sd
setosa	50	5.01	0.352
versicolor	50	5.94	0.516
virginica	50	6.59	0.636

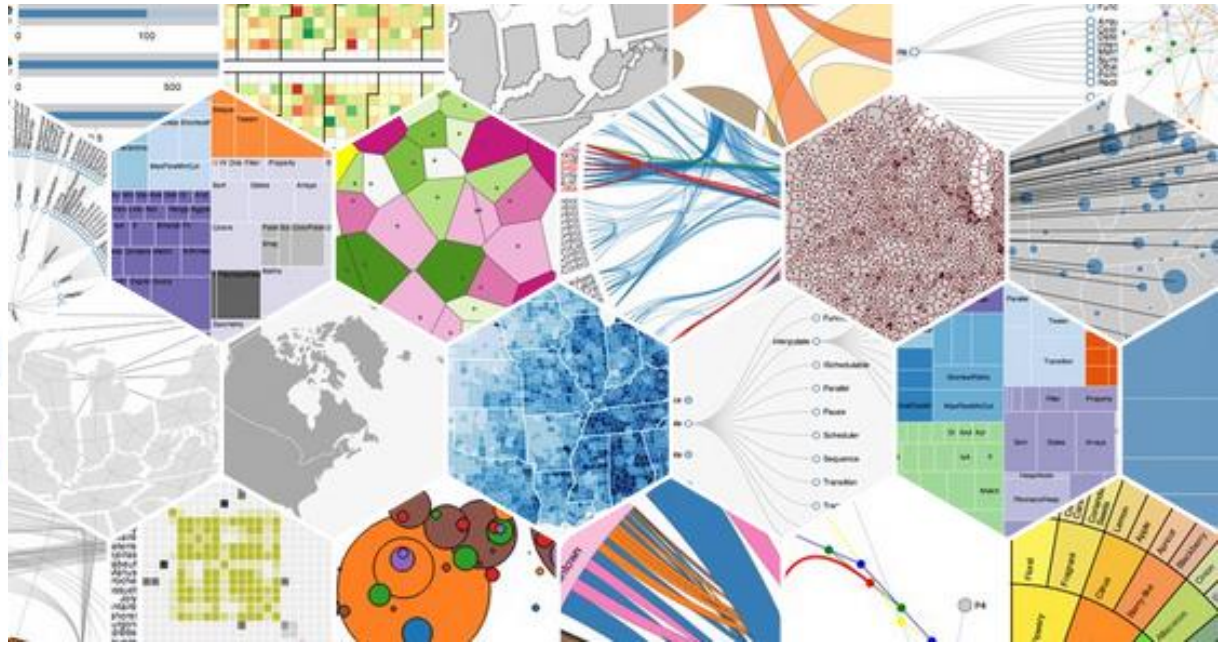
iris data set gives the measurements in cm of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.

Ggplot2

ggplot2 est une librairie R de visualisation de données développée par Hadley Wickham. La librairie est développée selon les principes développés par Leland Wilkinson dans son ouvrage The Grammar of Graphics. Wikipédia

Programmé en : R

Quelles solutions techniques ?



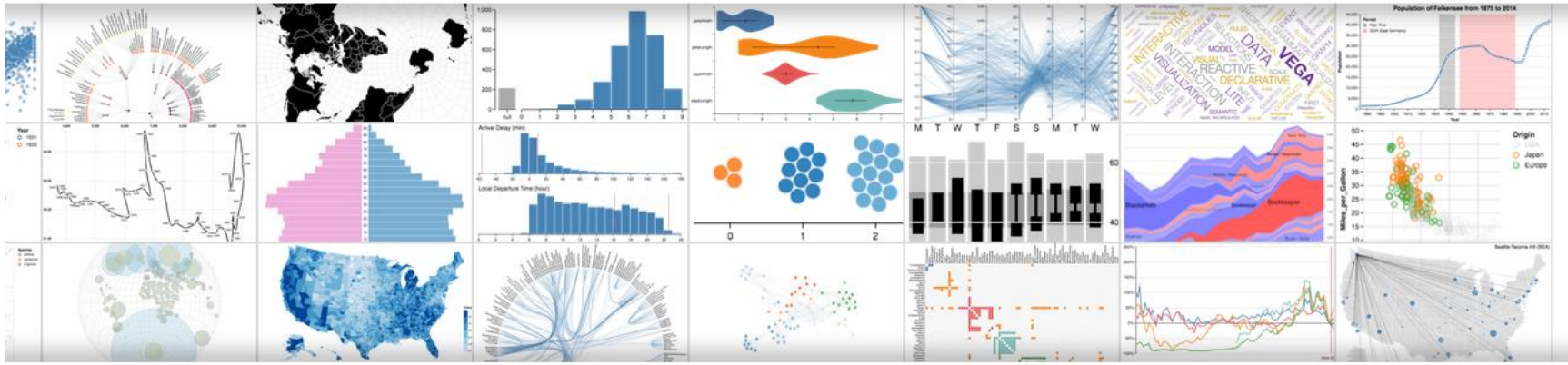
D3.js

D3.js est une bibliothèque graphique JavaScript qui permet l'affichage de données numériques sous une forme graphique et dynamique. Il s'agit d'un outil important pour la conformation aux normes W3C qui utilise les technologies courantes SVG, JavaScript et CSS pour la visualisation de données. Wikipédia

Quelles solutions techniques ?

Vega Examples Tutorials Documentation Usage About GitHub Try Online

Vega – A Visualization Grammar



Vega is a *visualization grammar*, a declarative language for creating, saving, and sharing interactive visualization designs. With Vega, you can describe the visual appearance and interactive behavior of a visualization in a JSON format, and generate web-based views using Canvas or SVG.

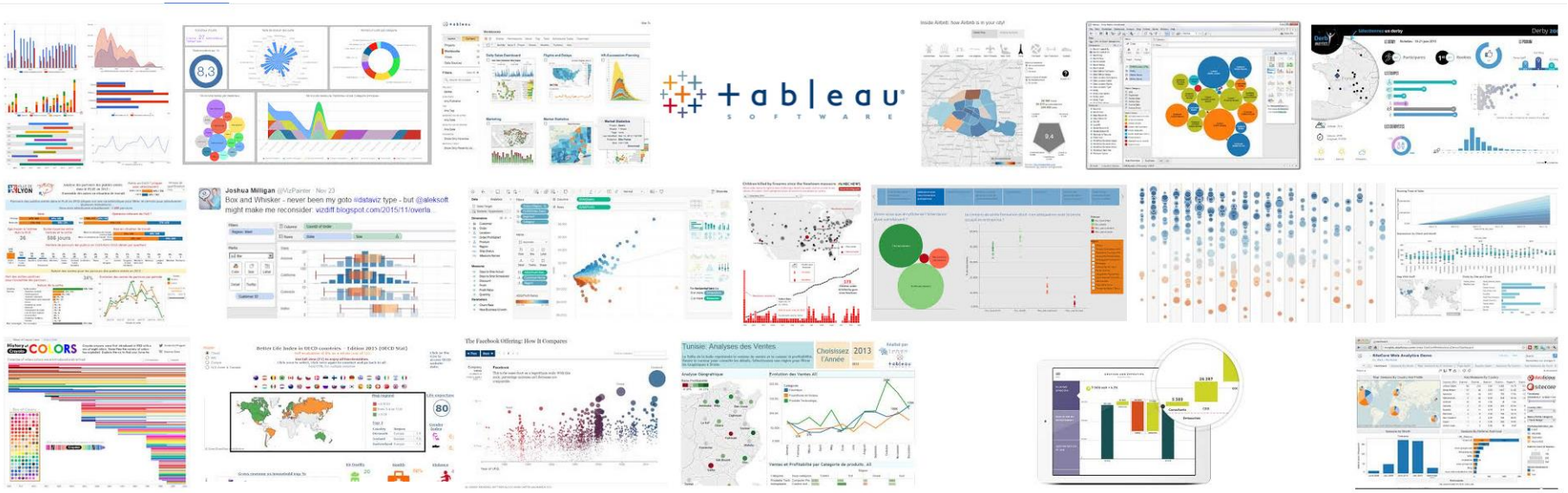
Vega provides basic building blocks for a wide variety of visualization designs: data loading and transformation, scales, map projections, axes, legends, and graphical marks such as rectangles, lines, plotting symbols, *etc.* Interaction techniques can be specified using reactive signals that dynamically modify a visualization in response to input event streams.

Version 4.4.0

A Vega *specification* defines an interactive visualization in a JSON format. Specifications are parsed by Vega's JavaScript *runtime* to generate both static images or interactive web-based views. Vega provides a convenient representation for computational generation of

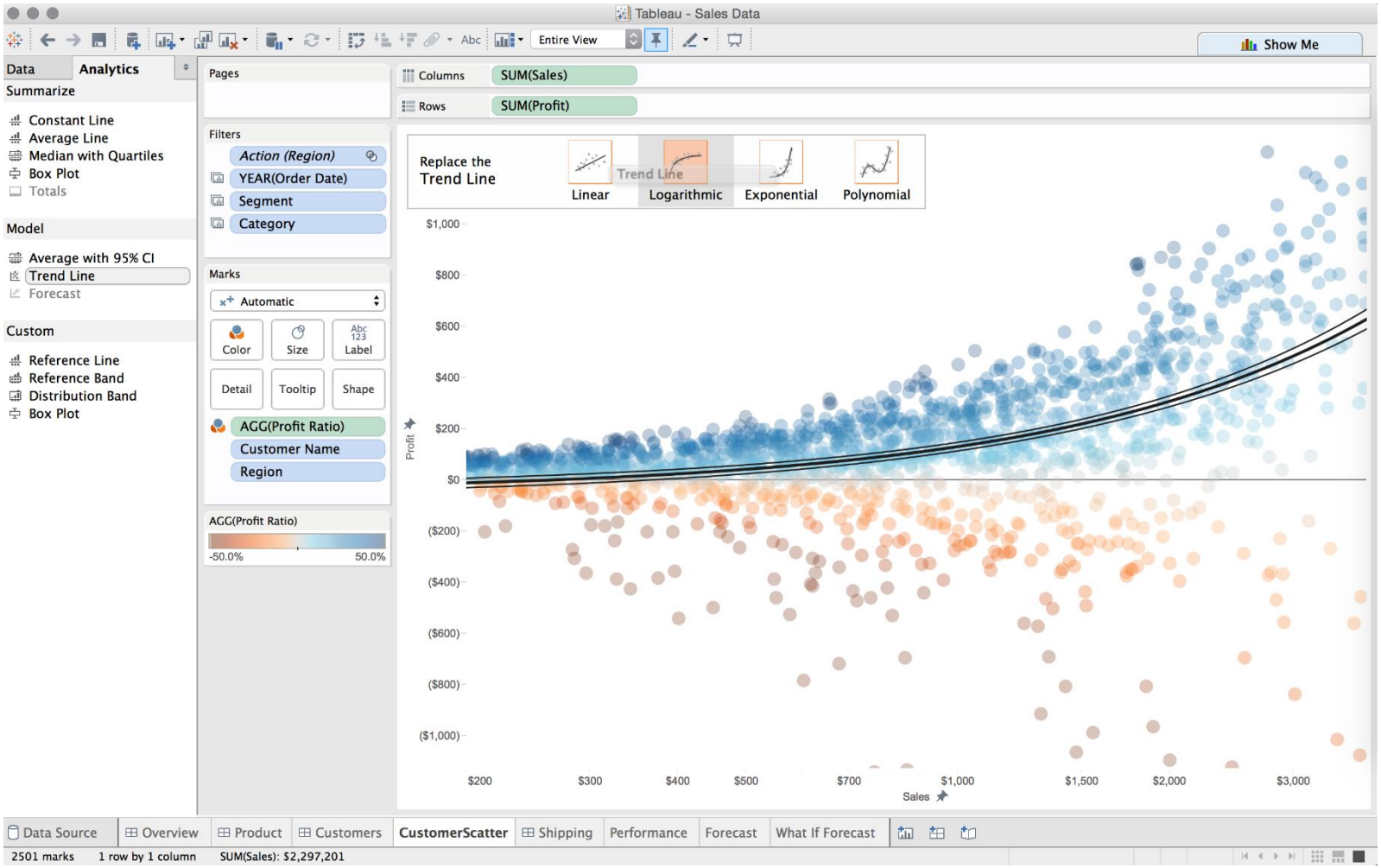
Quelles solutions techniques ?

- **Tableau : Le leader du marché, très intuitif et efficace**



https://www.tableau.com/fr-fr

Quelles solutions techniques ?



Quelles solutions techniques ?

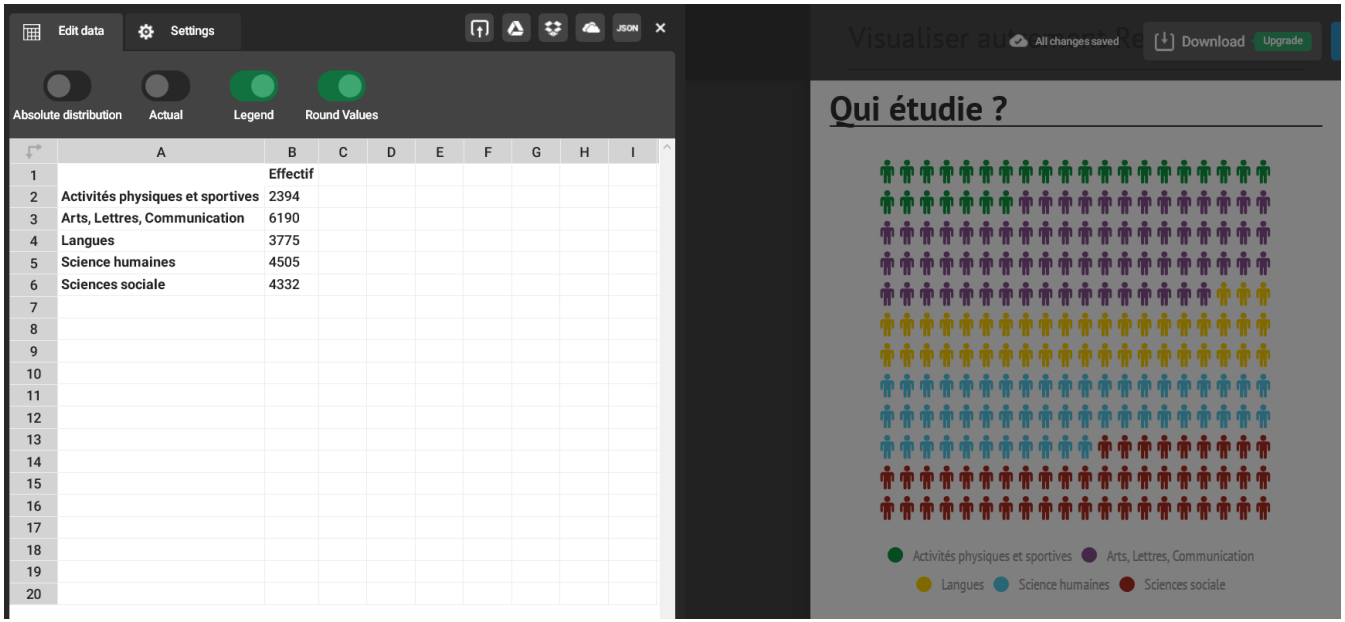
+ Mises en pratiques

#Infographie

infogr.am

Infogr.am

- **Infogr.am permet de créer rapidement une infographie en important ses propres données**
  - Le résultat est beaucoup plus esthétique et tendance que les graphiques proposés par les tableurs classiques

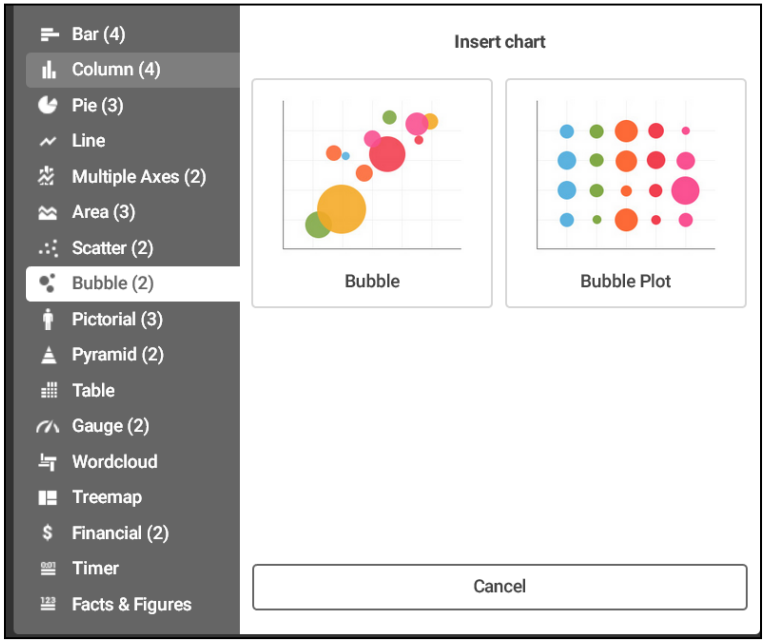


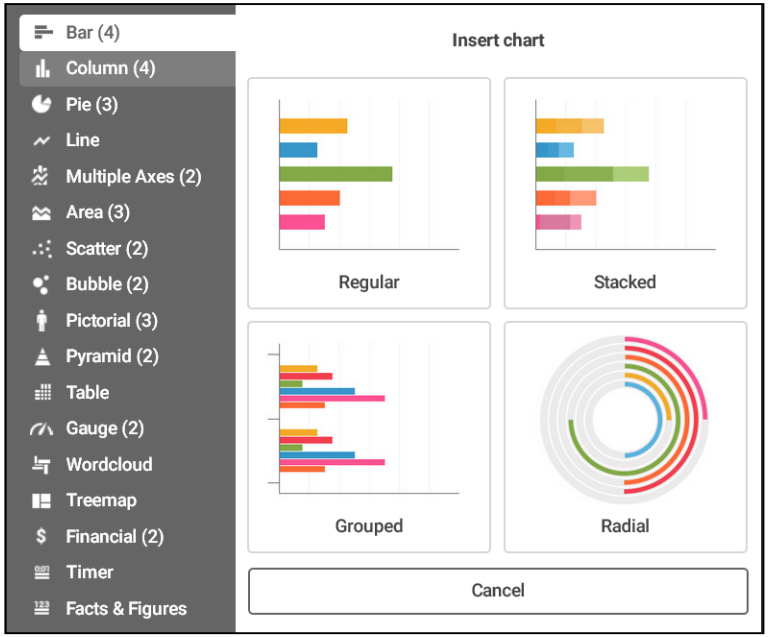
	A	B
1		Effectif
2	Activités physiques et sportives	2394
3	Arts, Lettres, Communication	6190
4	Langues	3775
5	Science humaines	4505
6	Sciences sociale	4332

https://infogr.am/

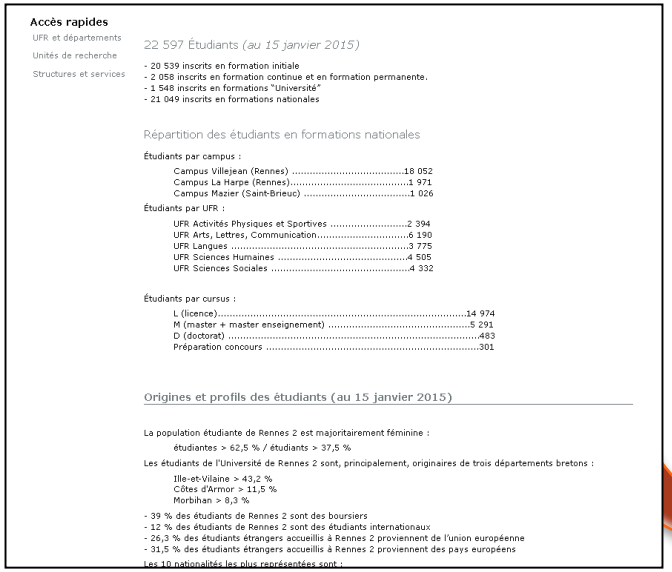
Infogr.am

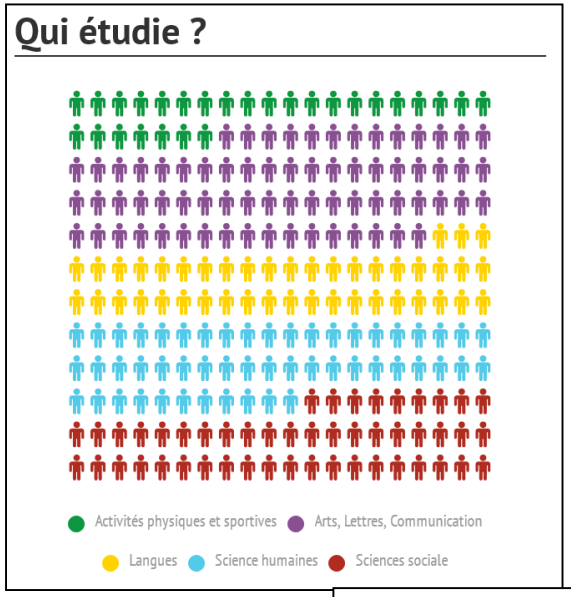
- **Créer des visualisation originales**
  - Collection riche de mode de représentation de données
  - Personnalisation (couleur, thèmes, taille, disposition,…)

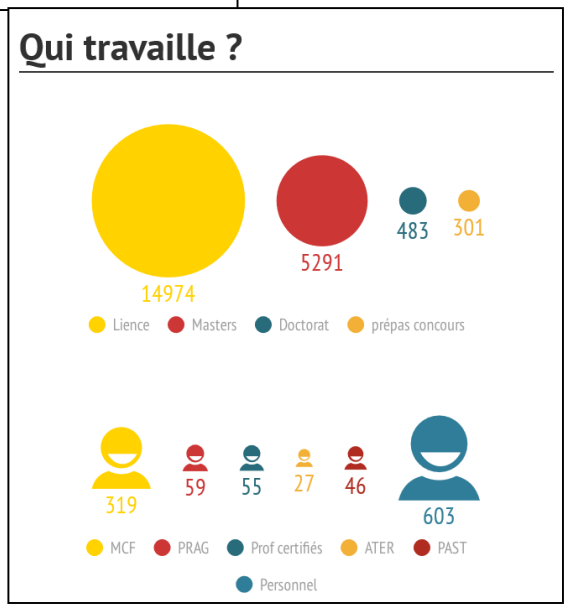




Infogr.am







Mon exemple :
https://infogr.am/dataviz_rennes_2

Les chiffres clefs de l'UCP

UNIVERSITÉ de Cergy-Pontoise

Formations | Recherche | Vie étudiante | International | Université | Actualités

Université

Accueil / Université / Découvrir l'UCP / Chiffres clés

Partager :

- Actualités de l'université
- **Découvrir l'UCP**
  - > Chiffres clés
  - > Identité graphique, logo
- Engagements
- Organisation
- Vie pratique
- Université Paris Seine
- Recrutements et concours
- Actes règlementaires
- Marchés publics
- Venir à l'université

Chiffres clés de l'UCP

L'université de Cergy-Pontoise en chiffres.

Télécharger le fichier « Chiffres clés de l'université de Cergy-Pontoise »

1991 : création de l'UCP, une des 4 universités nouvelles d'Île-de-France
1er janvier 2007 : intégration de l'institut universitaire de formation des maîtres (IUFM) de l'académie de Versailles
1er janvier 2009 : passage aux responsabilités et compétences élargies (RCE)
Février 2012 : membre fondateur du PRES UPGO (Université Paris Grand Ouest)
Juillet 2013 : transformation du PRES UPGO en COMUE (communauté d'universités et établissements)
Novembre 2014 : vote des statuts de la COMUE université Paris Seine

Une université pluridisciplinaire

- **5 UFR** (Unité de Formation et de Recherche) : droit, économie et gestion (EG), langues et études internationales (LEI), lettres et sciences humaines (LSH), sciences et techniques (ST)
- **1 institut universitaire de technologie** (IUT) certifié ISO 9001 en avril 2011 pour la qualité des formations dispensées
- **1 institut d'études politiques** (IEP) Sciences Po Saint-Germain-en-

https://www.u-cergy.fr/fr/universite/decouvrir/chiffrescles.html

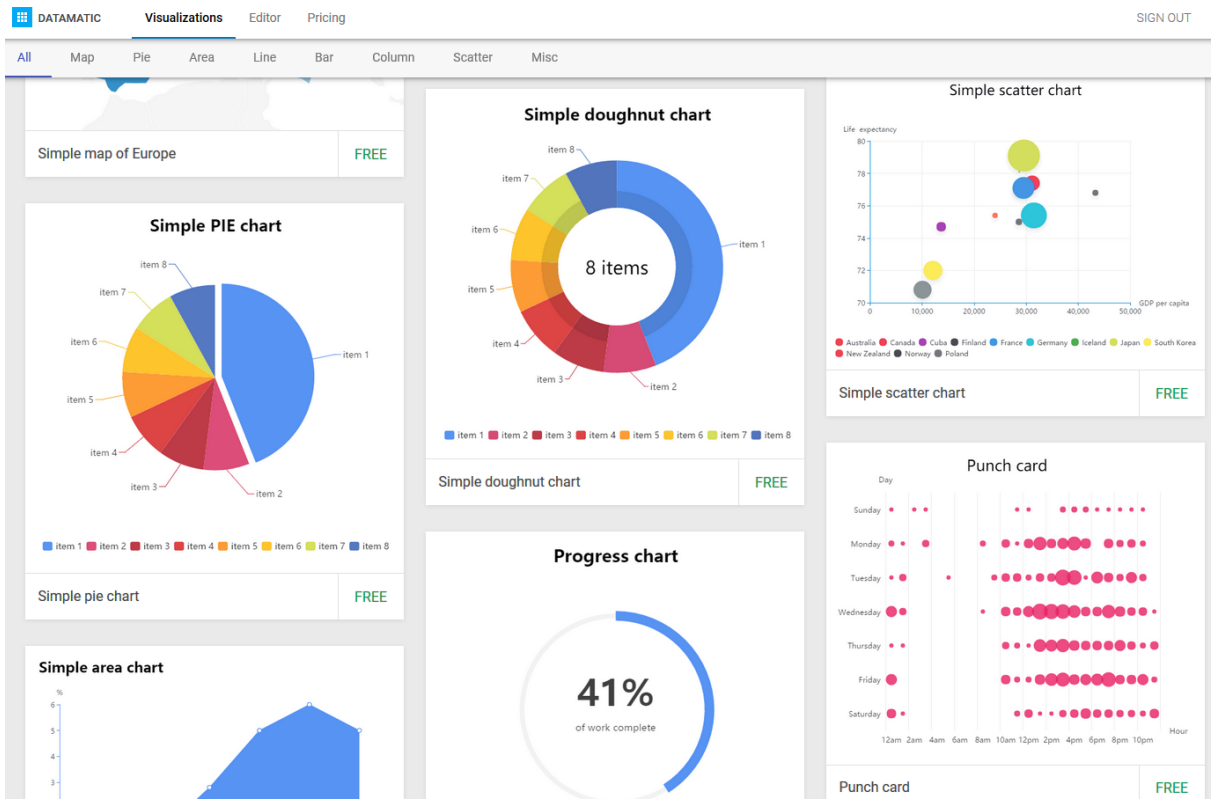
+ Mises en pratiques

#Dataviz Intro

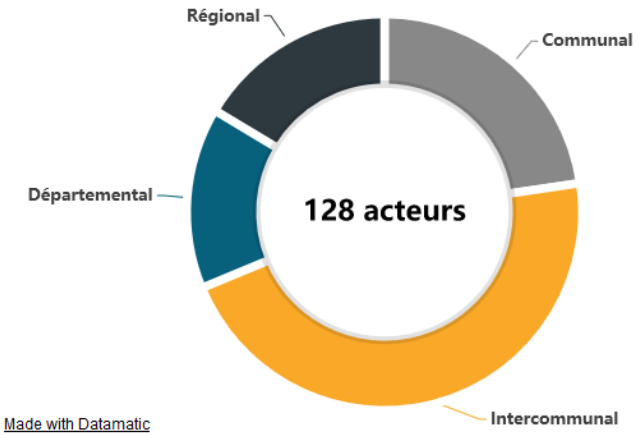
Datamatic

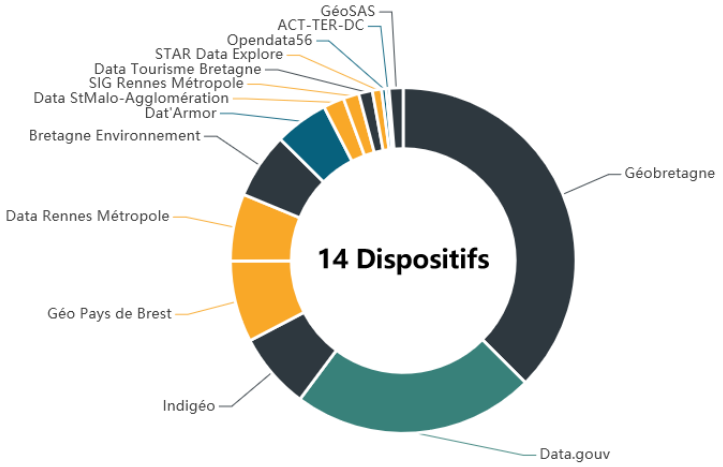
- Datamatic est une solution de dataviz (propulsée par Google) qui permet de réaliser assez simplement des dataviz en ligne interactives (choix assez limité)

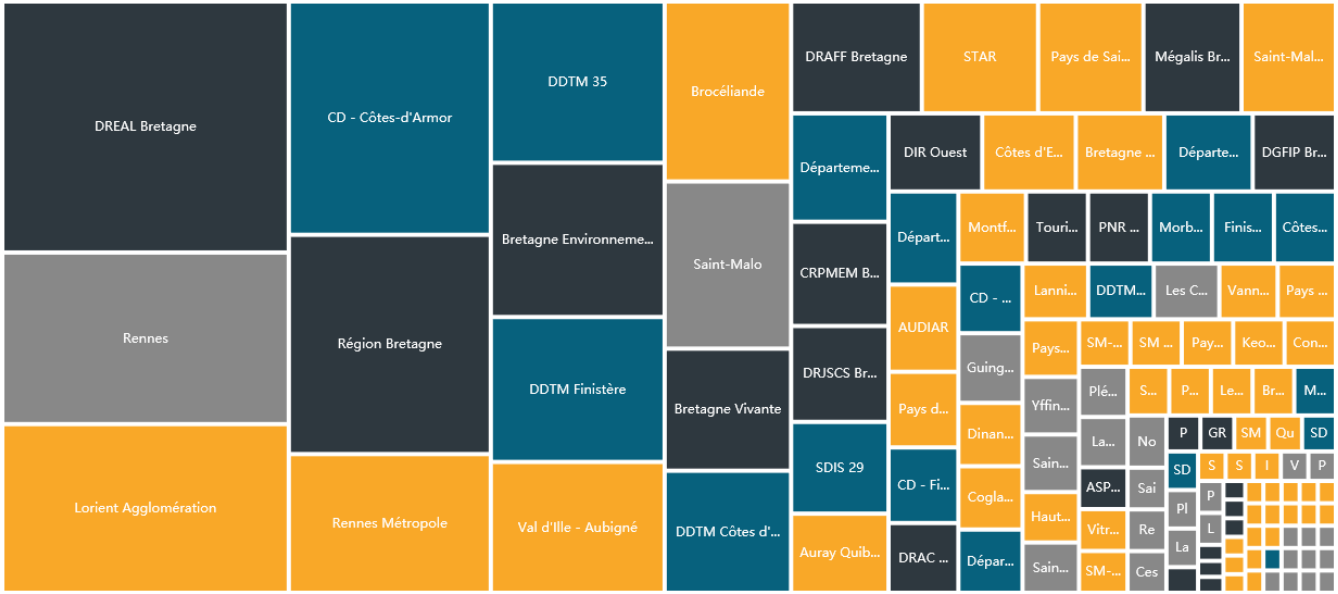
https://datamatic.io/



Datamatic - Exemples





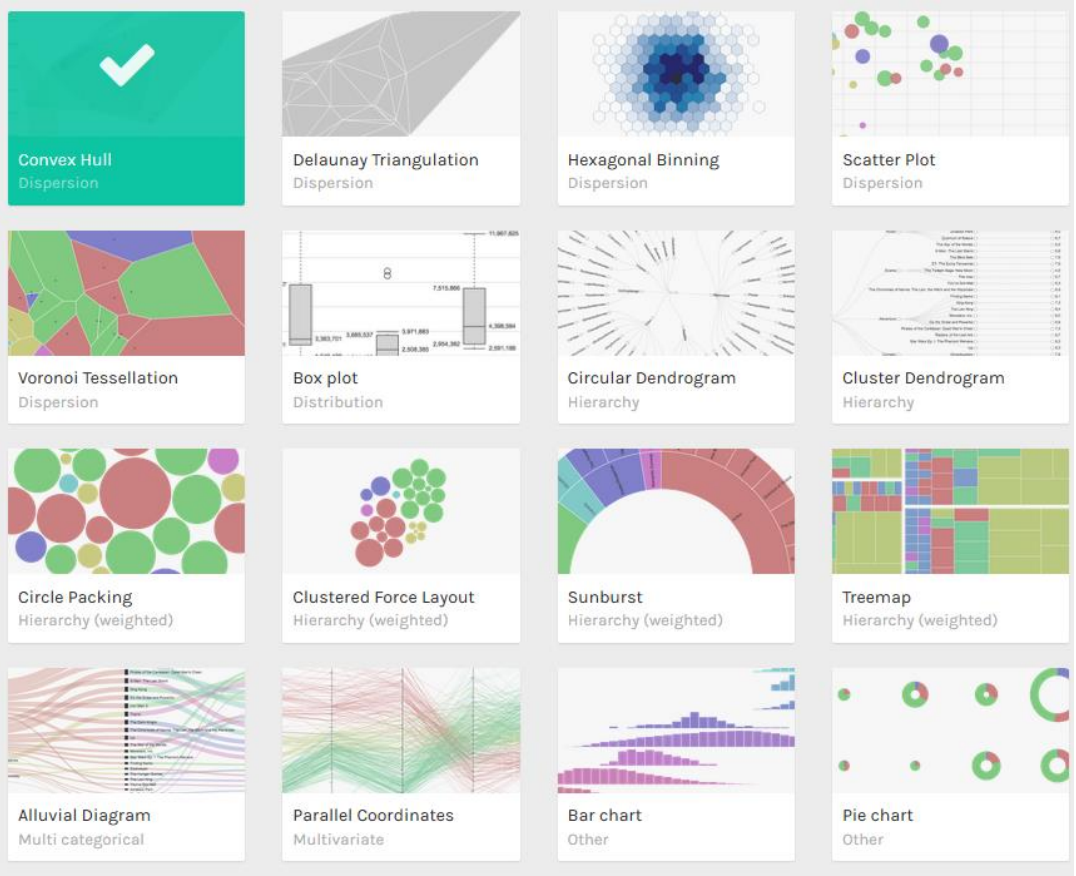


Mises en pratiques
#Dataviz Design1

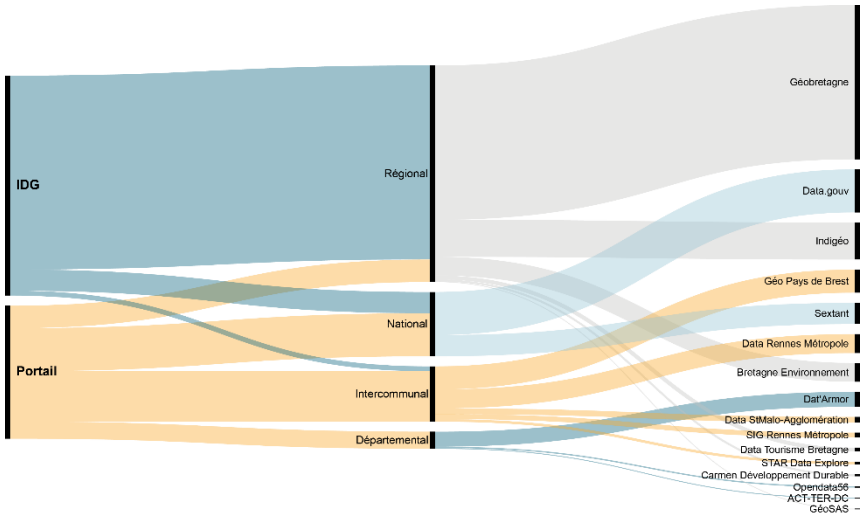
RawGraphs - Exemples

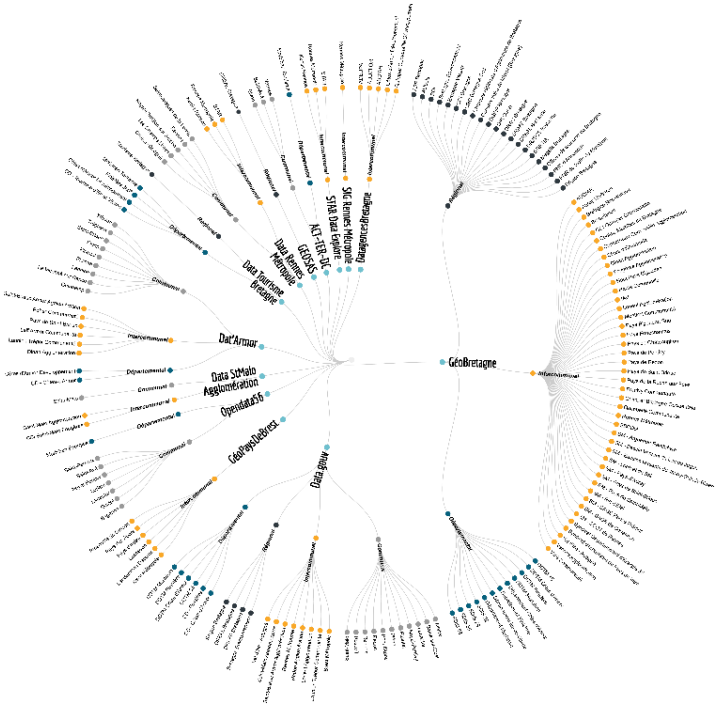
- Raw permet de créer des dataviz originales en quelques clics (surcouche interfacée de D3.js)

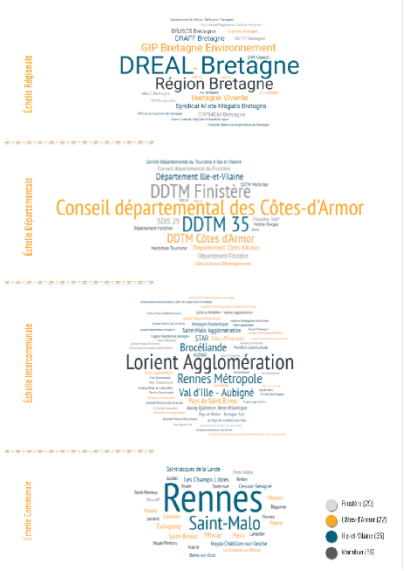
http://rawgraphs.io/

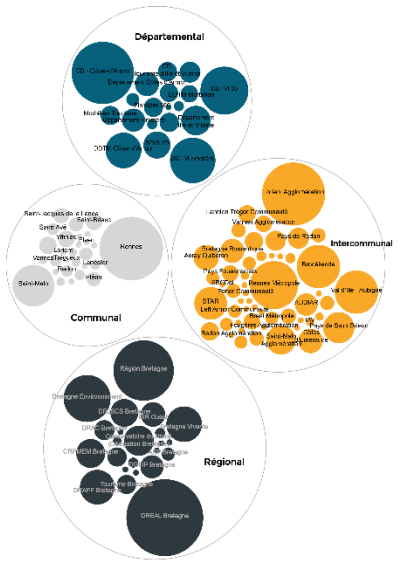


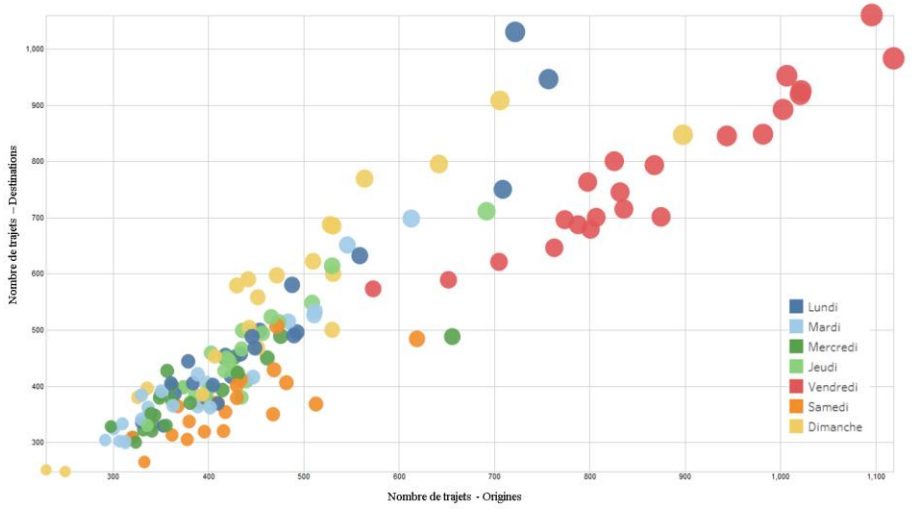
RawGraphs - Exemples











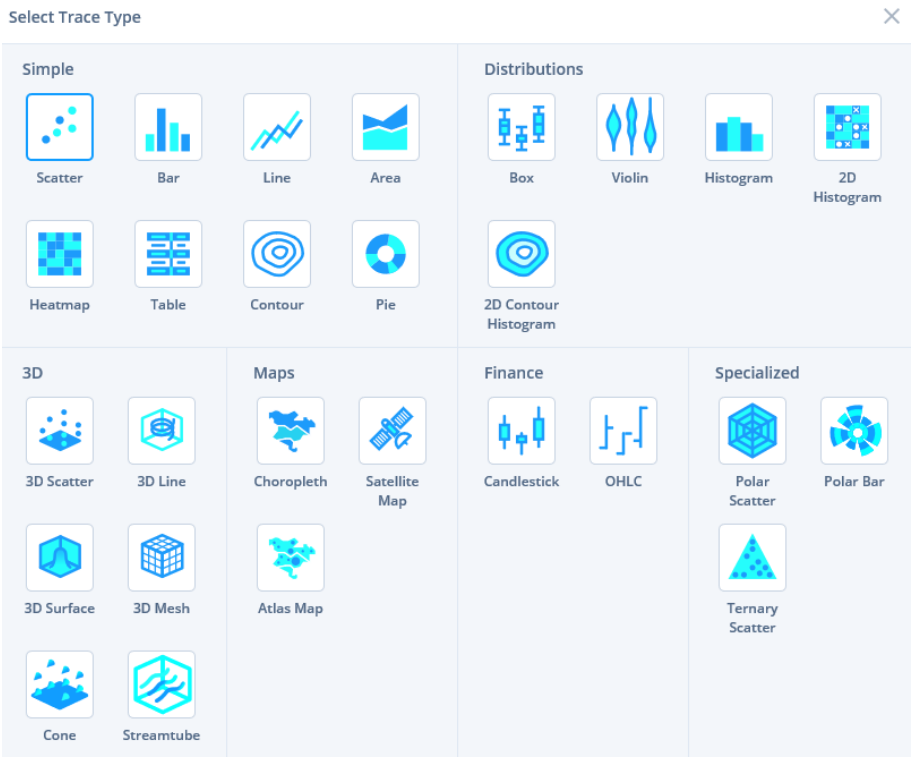
+ Mises en pratiques

#Dataviz Stat

Plotly

- Plotly permet de réaliser des dataviz orientées statistiques (Scatter Plot, Box, Violin, Polar Scatter,…)

https://plot.ly/Auth/login/?action=signup#/



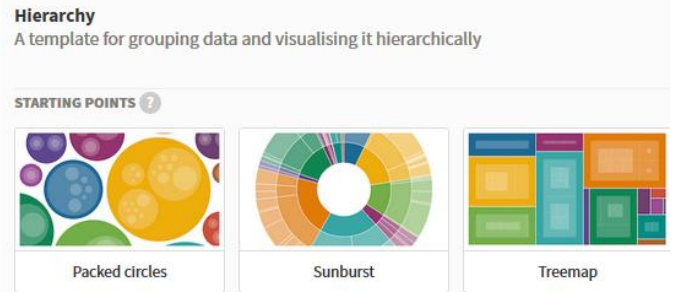
+ Mises en pratiques
#Dataviz Design2

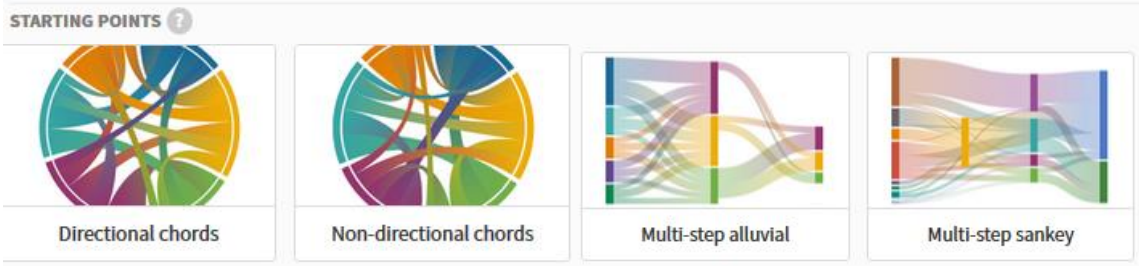
Flourish

Flourish

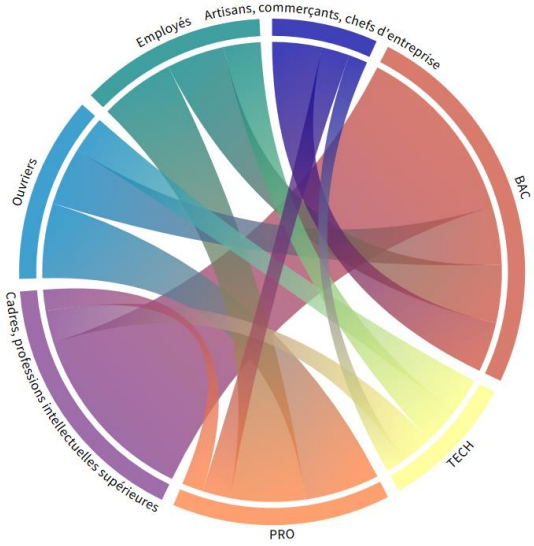
- Petit dernier des solutions Dataviz, Flourish permet de réaliser rapidement des dataviz originales et interactives dans un environnement en ligne intuitif et stable

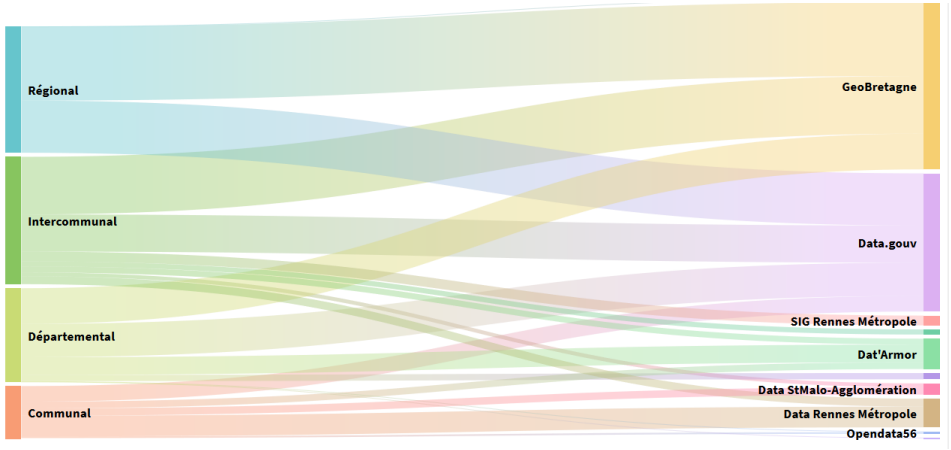
https://app.flourish.studio/register?noredirect=true

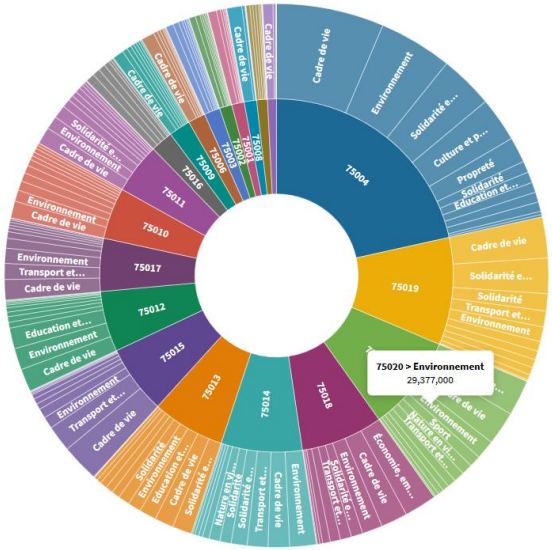




Flourish - Exemples





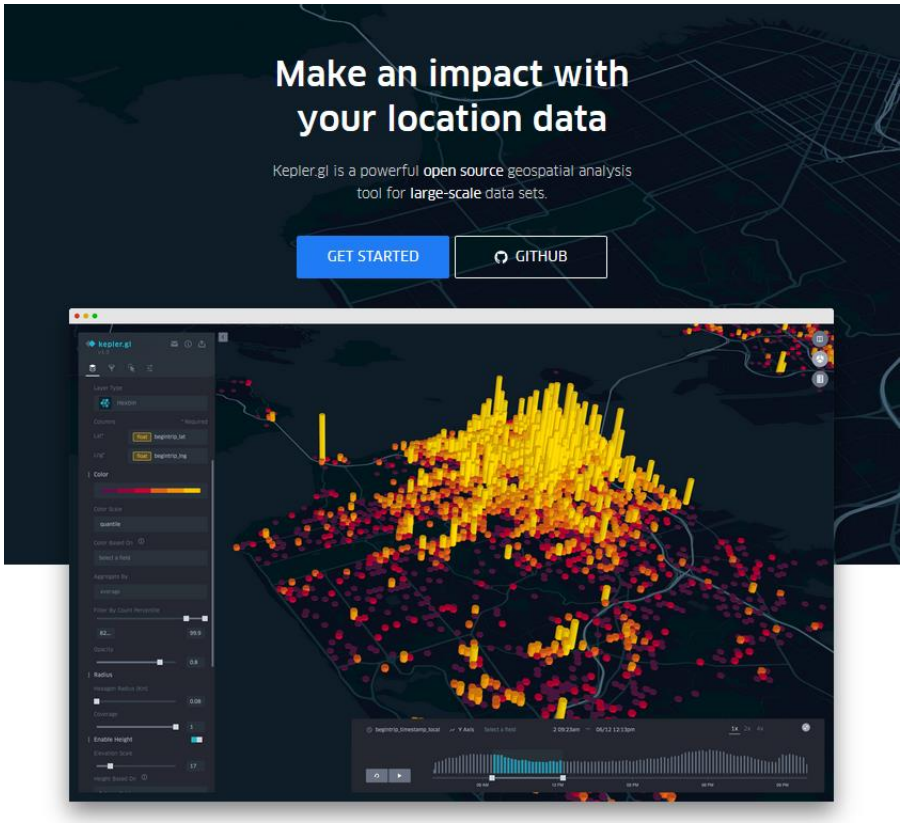


+ Mises en pratiques
#GeoDataVIZ

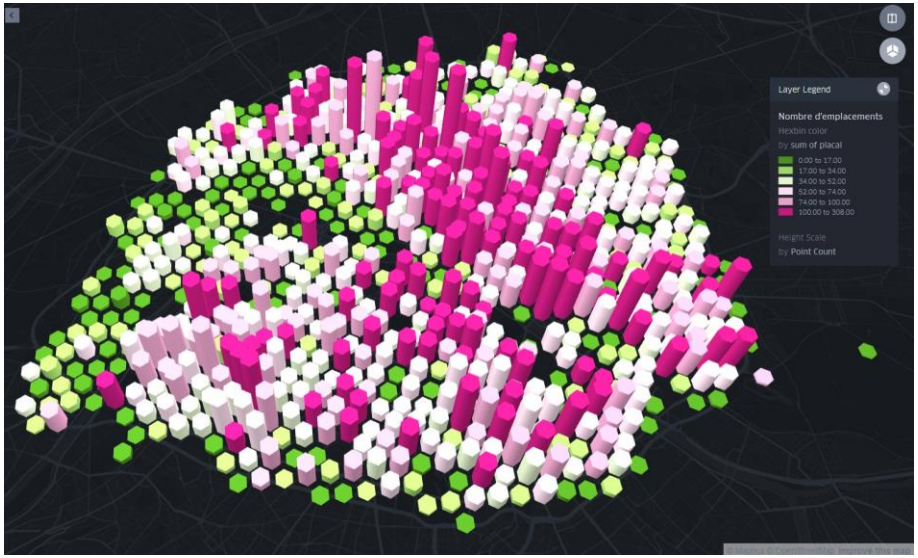
KeplerGL

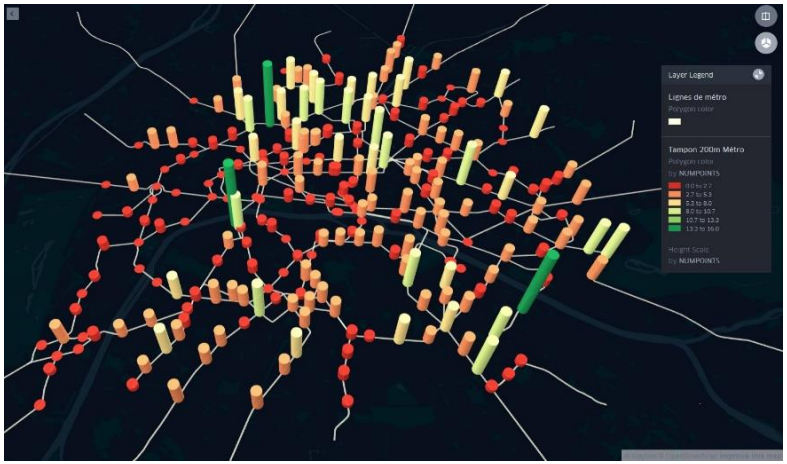
- **KeplerGL est une solution de géovisualisation de données (propulsée par Uber) qui permet de visualiser sous des formes originales de larges de jeux de données**

https://kepler.gl/#/



KeplerGL - Exemples

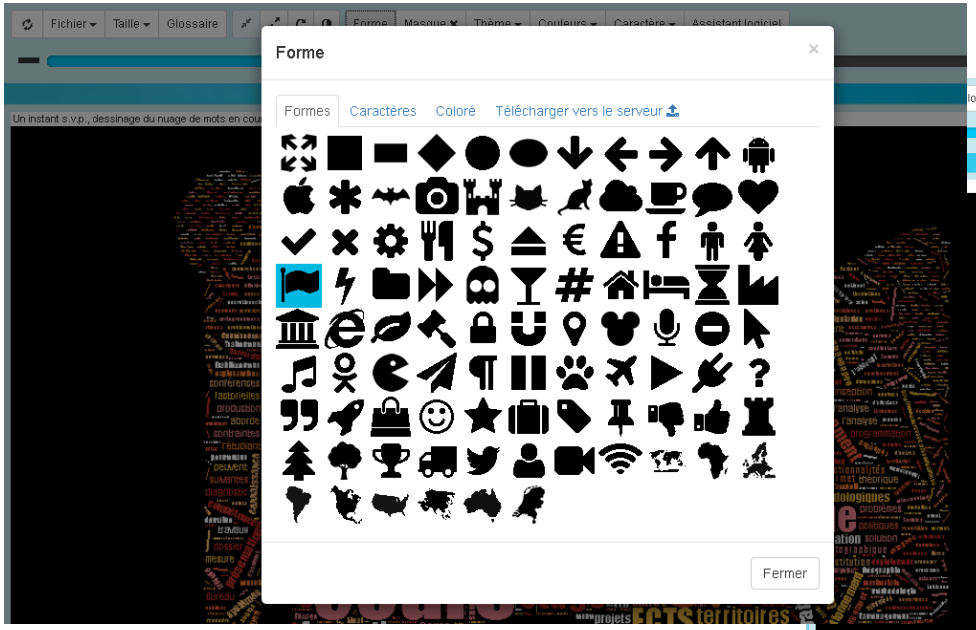


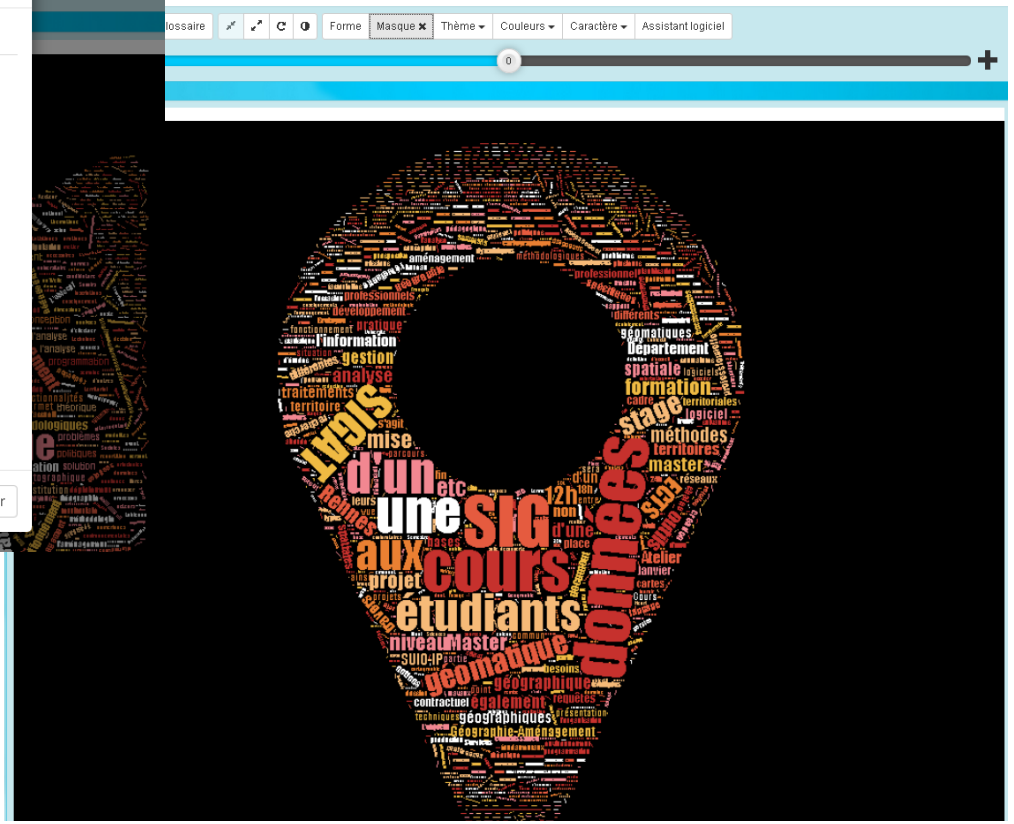


Autres outils

Nuage de mots

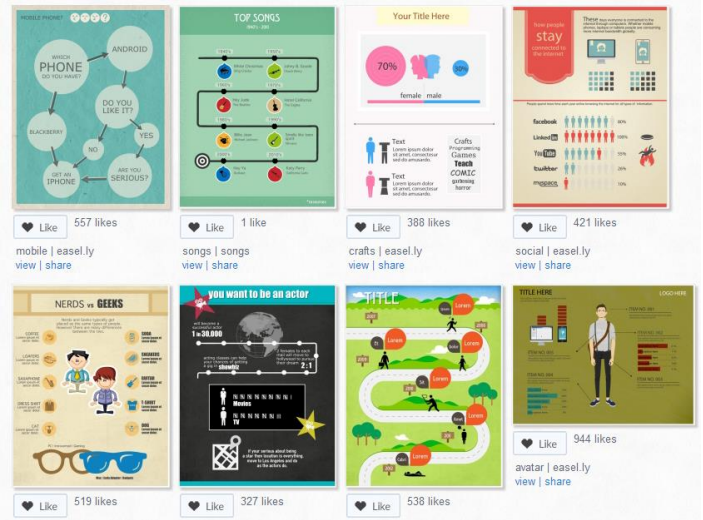
- http://www.nuagesdemots.fr/





Ease.ly

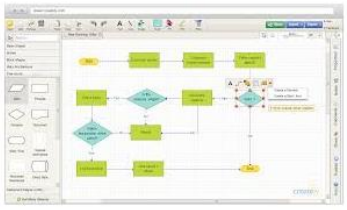
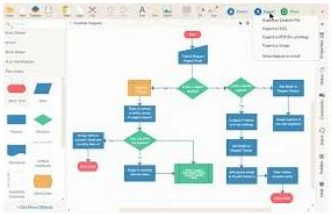
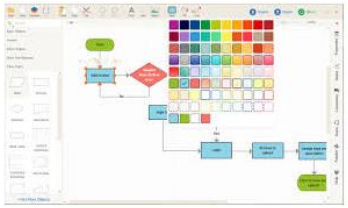
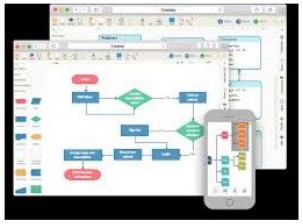
- **Créer des infographies http://www.easel.ly/**
  - Easel.ly propose gratuitement une douzaine de thèmes assez sommaires qui pourront ensuite servir de base à une infographie plus personnalisée
  - Après, il faut les éditer en ligne, en choisissant parmi de nombreux objets graphiques (icônes, cartes, paysages, formes, etc.).
    - Infographies en HTML5

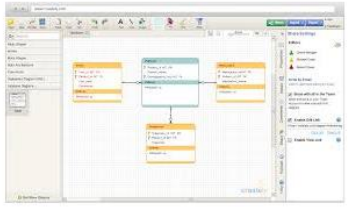


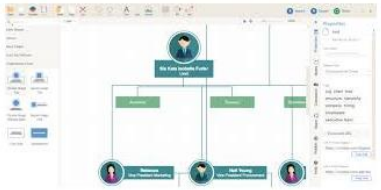
Creately

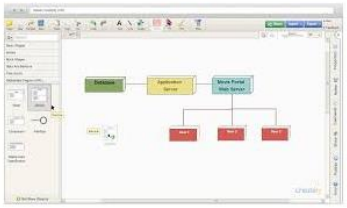
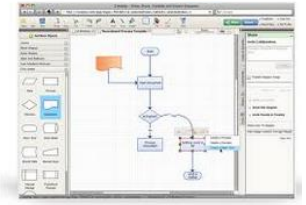
- **Plus orienté schémas, diagrammes,…**

https://creately.com

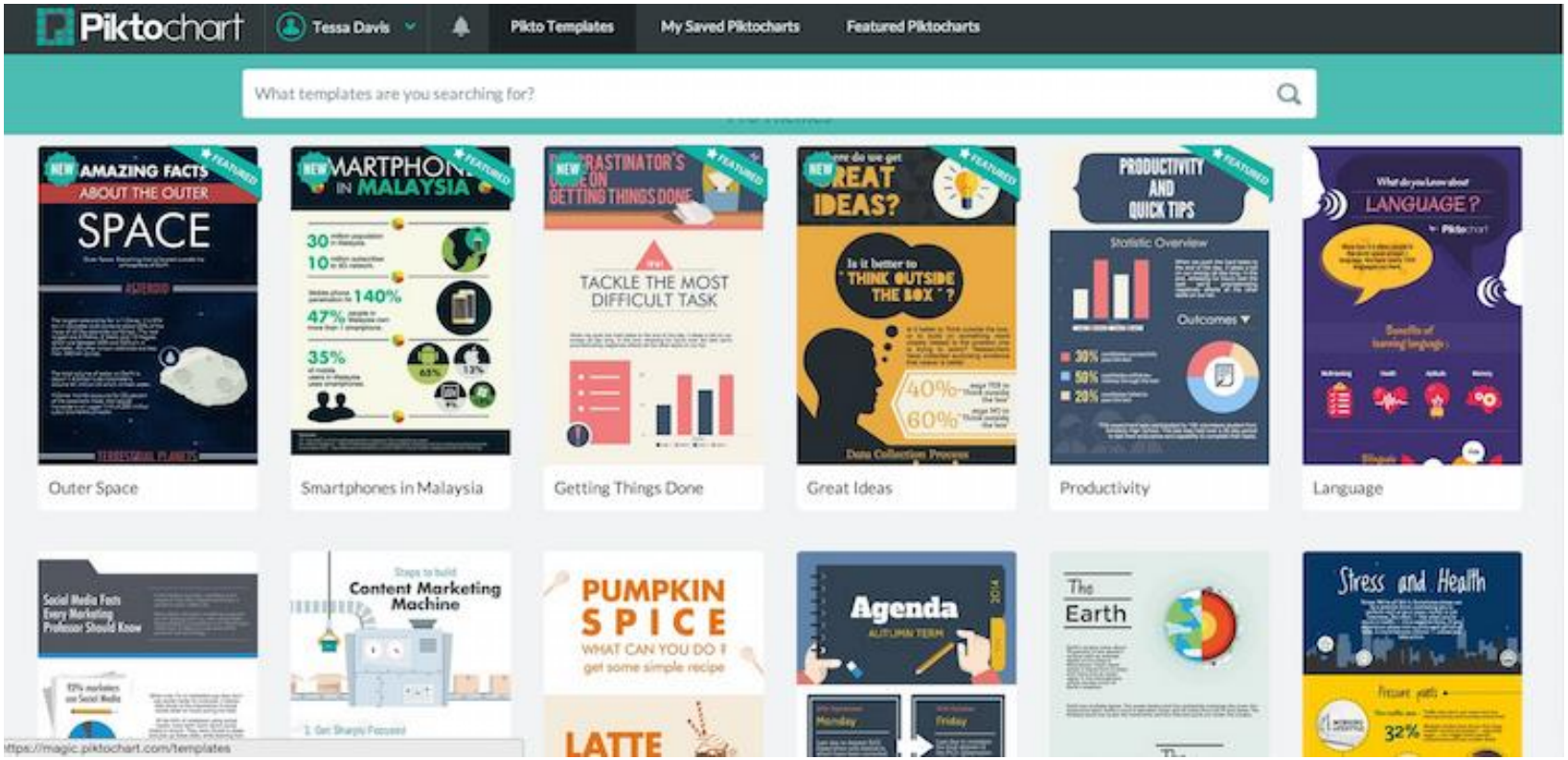




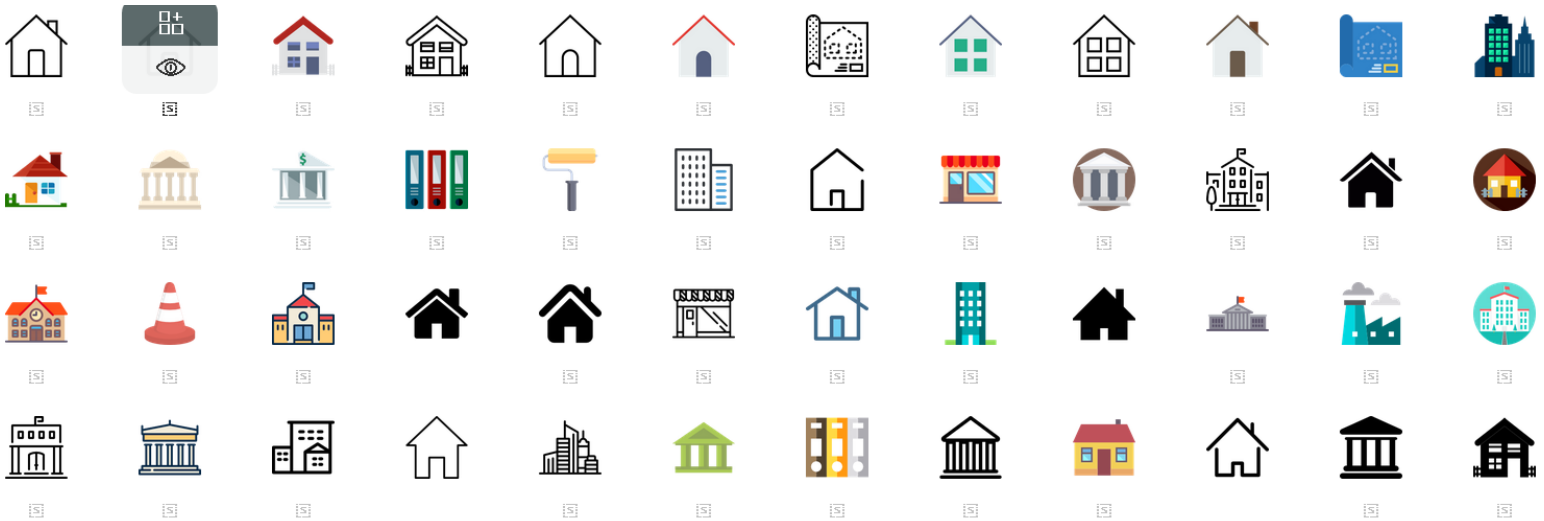
PikToChart

https://magic.piktochart.com/



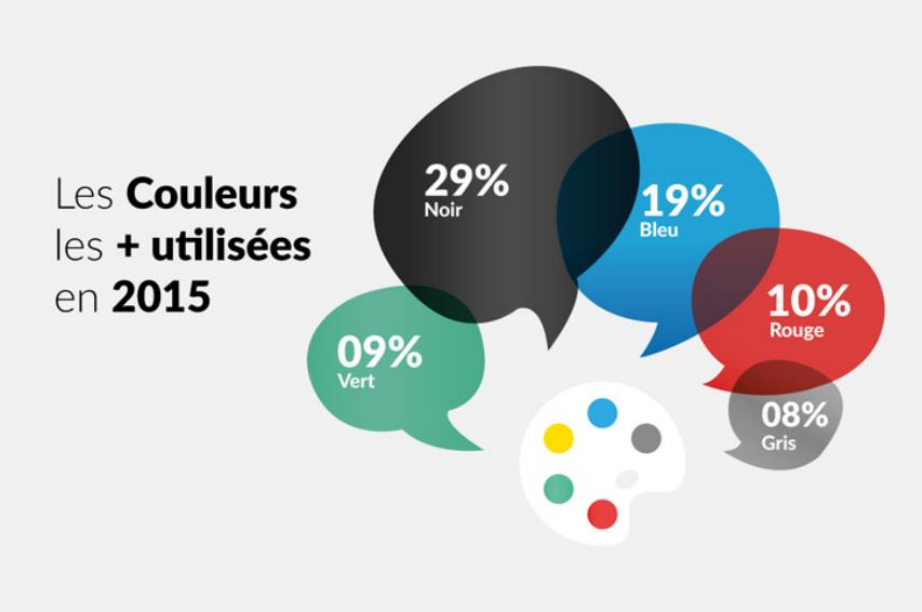
Aller plus loin

- Les pictogrammes : la clef d'une bonne visualisation !
  - https://thenounproject.com/
  - http://www.flaticon.com/
  - http://fr.freepik.com/
  - Google Image avec un filtre « Clipart »



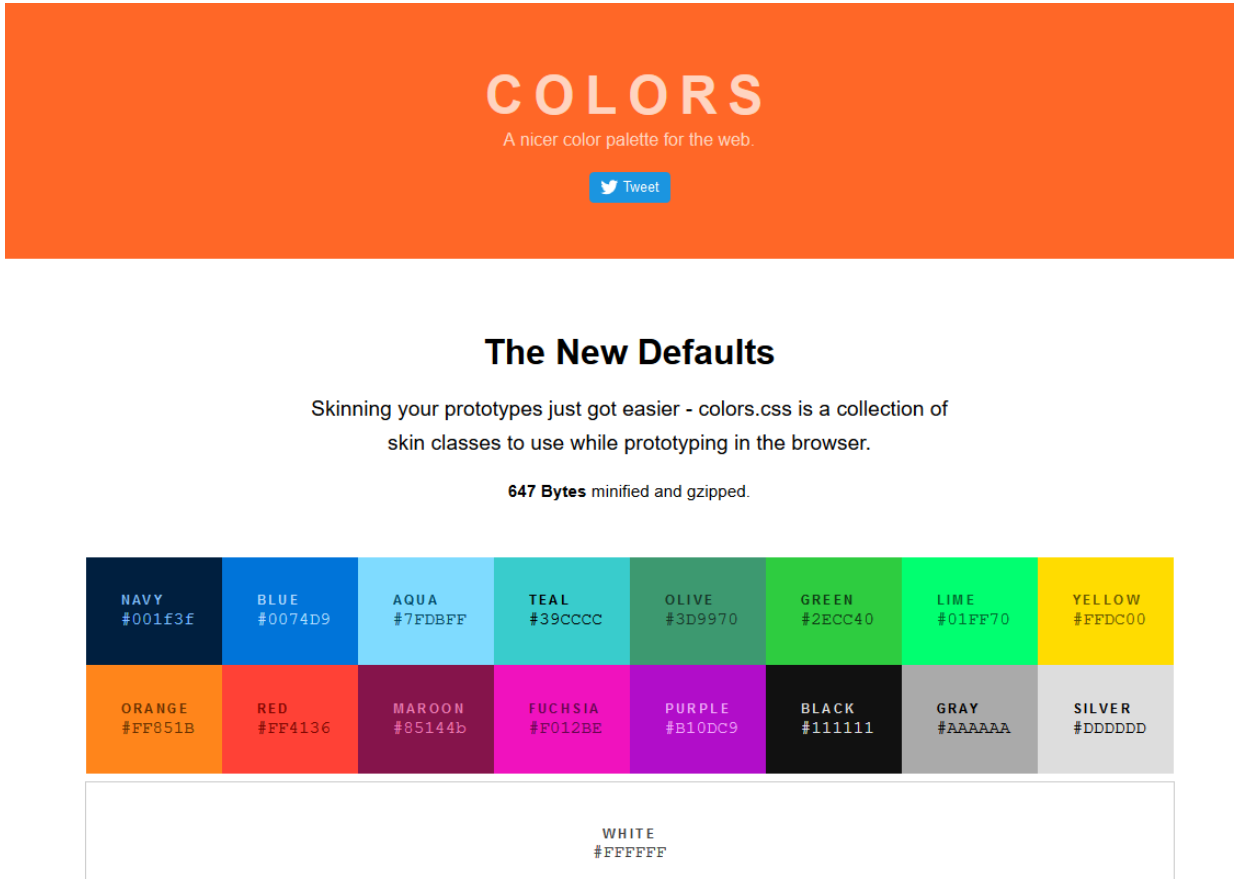
Aller plus loin

- Quelques pistes concernant les couleurs tendances



- Réutiliser une couleur sympa vue sur un site
  https://addons.mozilla.org/fr/firefox/addon/colorzilla/

Aller plus loin

- **Couleurs qui vont bien**

https://clrs.cc/