

IS2A4 - Internet

CM3 - Visualisation de données

Thibault Liétard

Polytech' Lille

2022

Introduction

Pourquoi la visualisation ?

Une définition

La visualisation de données consiste à pouvoir fournir une représentation visuelle sémantiquement pertinente à partir d'un grand ensemble de données hétérogènes.

Une citation

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.” [Herb Simon]

L'ordinateur

L'ordinateur répond efficacement à des questions précises (quel est le taux de chômage, détection d'une caractéristique précise sur une image,...) et obtient une réponse rapide.

L'être humain

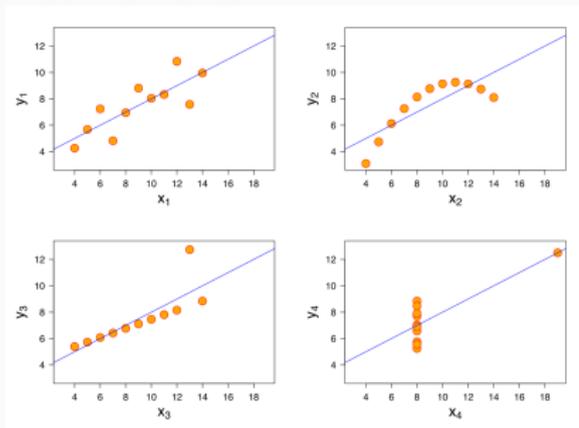
Plutôt efficace sur des questions complexes (optimisation d'emploi du temps, relation chômage/politiques économiques, est-ce un chien sur cette photo...) mais en un temps plus long.

Pourquoi visualiser ?

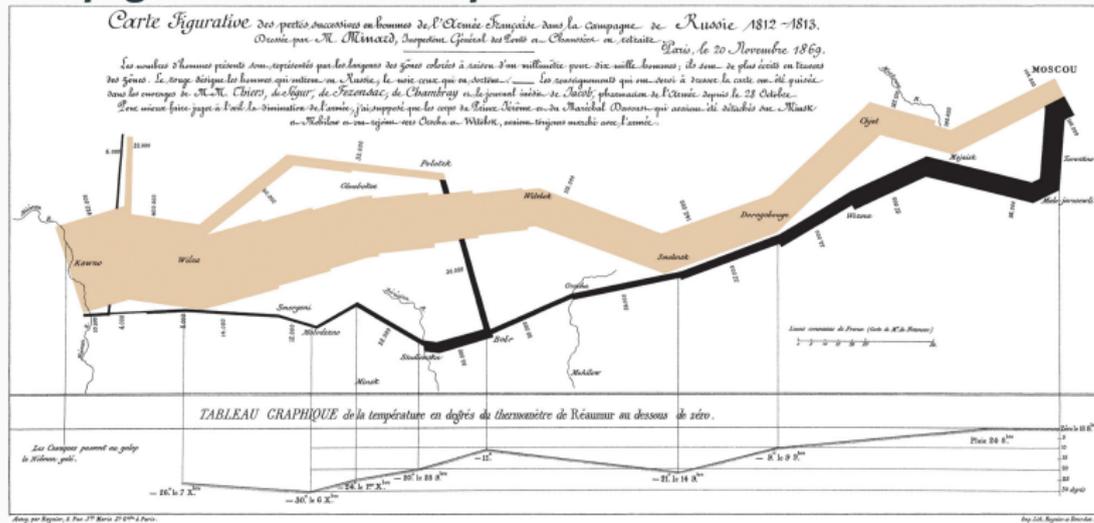
Avantage

- Enregistrer de l'information sous forme compacte
- Faciliter le raisonnement sur ces informations
- Transmettre de l'information

Inconvénients



Campagne de Russie de Napoléon

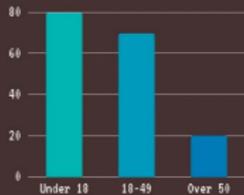


Les différents types de visualisation

Qui sont les gamers ?

LET'S PLAY ON! INFOGRAPHIE DU GAMING

RÉPARTITION ÂGE



LES JEUX VIDÉOS À LA MAISON



- ★ MULTITASKING
- ★ CREATIVITY
- ★ TEAMWORK
- ★ FINANCIAL SKILLS

32

MOYENNE D'ÂGE
D'UN GAMER



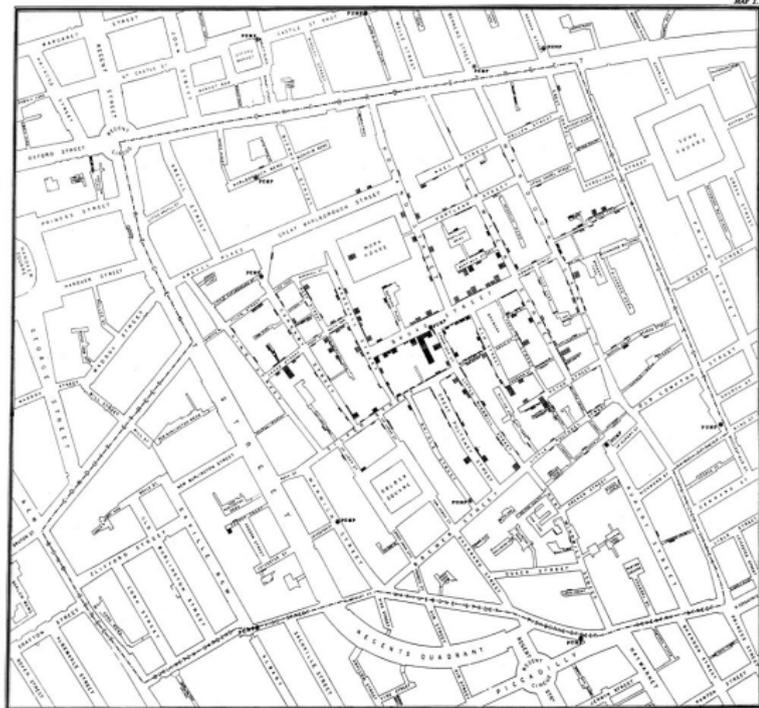
2 JOUEURS SUR 5
SONT DES FEMMES

LES VIDÉOS EN LIGNE EN FAVEUR DE LA SOCIALISATION

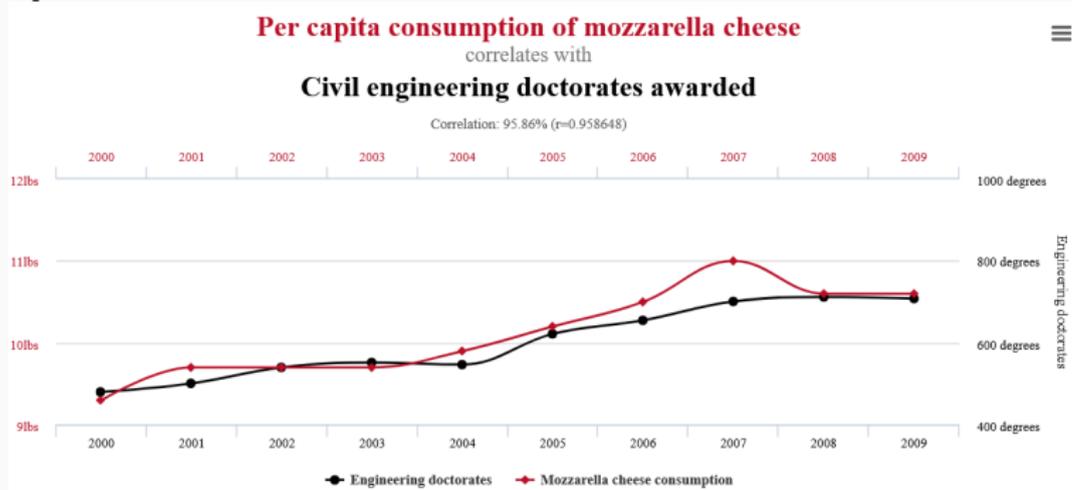
65% OF GAMERS
PLAY THEIR
GAMES WITH A
FRIEND



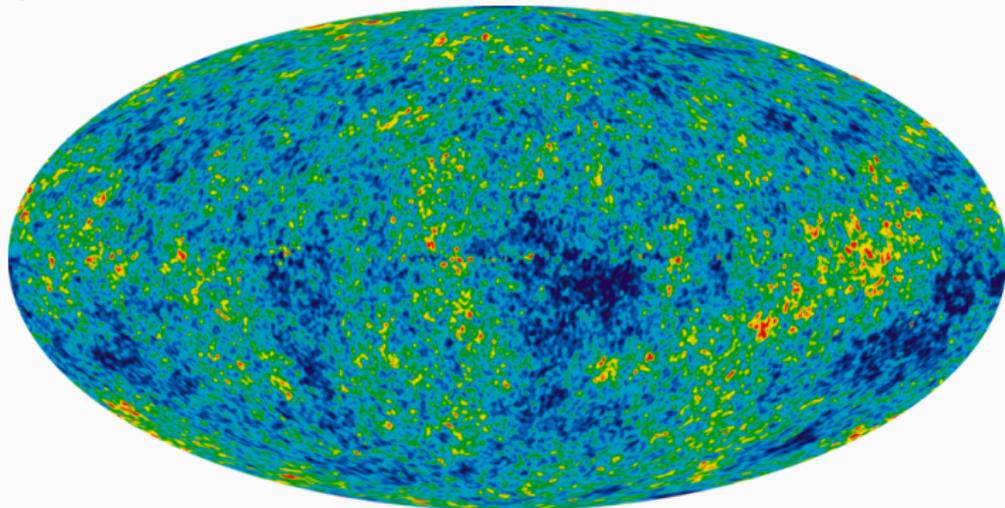
Épidémie de choléra de Broad Street (1854)



Spurious correlation



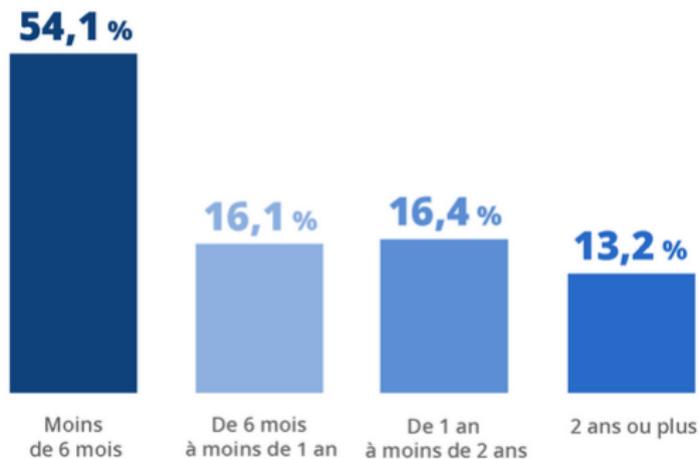
CMB



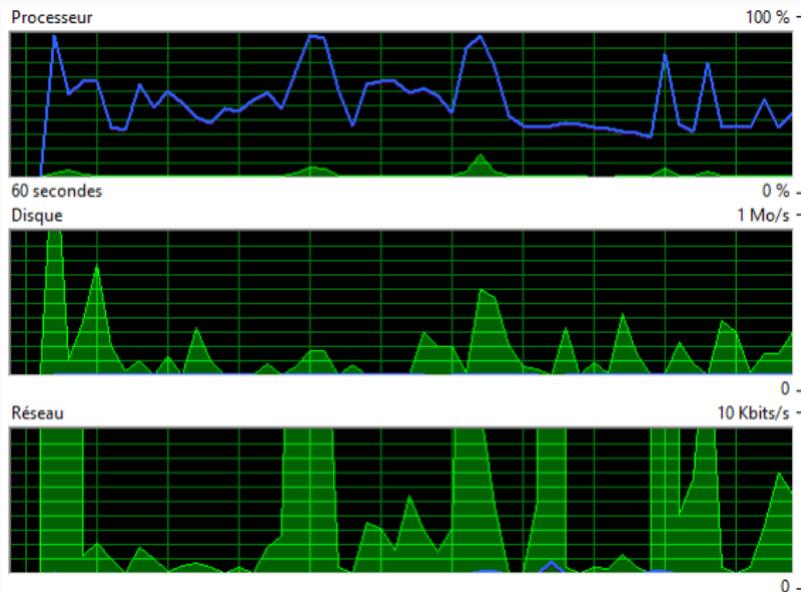
nécessite souvent du contexte...

INSEE

Répartition des chômeurs selon l'ancienneté au chômage en 2021



Moniteur de ressources



Comment choisir ?

Du contexte

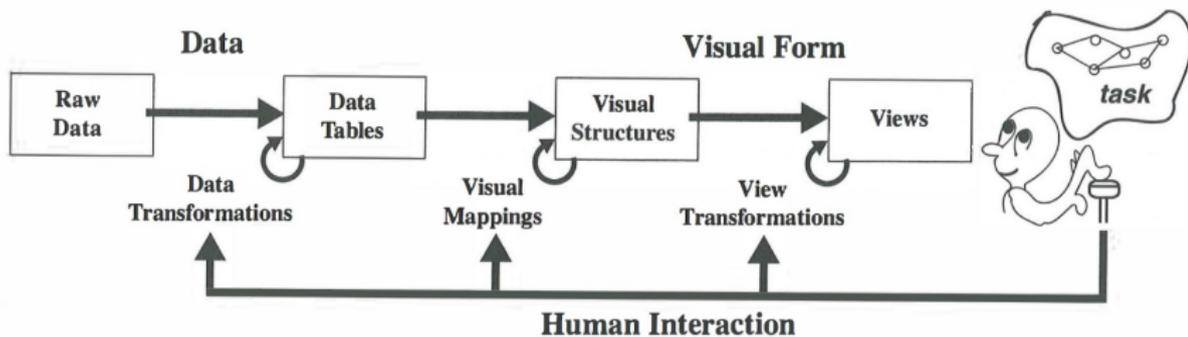
Le type de visualisation dépend de plusieurs choses définies par le contexte :

- Le type de données disponibles
- Le public cible
- Le contexte de présentation des données
- La quantité d'information à transmettre

Attention

On regarde d'abord ce que l'on a avant de décider ce qu'on va faire. Il n'est pas question de tordre les données pour arriver à un résultat bancaire.

Pipeline à suivre



Raw Data: idiosyncratic formats

Data Tables: relations (cases by variables) + metadata

Visual Structures: spatial substrates + marks + graphical properties

Views: graphical parameters (position, scaling, clipping, ...)

Tableaux de bord

Quoi qu'est-ce ?

Définition

Le tableau de bord est un ensemble d'indicateurs peu nombreux (5 à 10) permettant de prendre connaissance de l'état de l'évolution d'un système et éventuellement de prévoir les tendances de ces évolutions

Que trouver dedans ?

- Des chiffres clés
- Des graphiques
- Des KPI (Key Performance Indicator)
- Des indicateurs d'avancement / d'évolution

Exemple



Définition

Un indicateur est un paramètre ou une combinaison de paramètres qui représente l'état ou l'évolution d'un système, il est choisi en fonction des leviers d'action qui seront utilisés pour prendre d'éventuelles mesures correctives et donc en fonction de décisions à prendre dans le futur.

Nature

- les indicateurs d'activité (quantité produite, volume d'achat, chiffre d'affaires)
- les indicateurs financiers (salaires, achat, frais généraux)
- les indicateurs de rentabilité (résultat net, marge opérationnelle)
- les indicateurs de qualité (délai de fabrication, satisfaction des clients, réclamation)

Définition valable... dans le cas d'une entreprise

La notion de KPI est à adapter quand on s'intéresse à autre choses qu'à des organisations commerciales, voire à abandonner.

Dans le contexte d'analyses sociales ou scientifiques, la notion de KPI peut simplement ne pas avoir de sens.

Exemple

- Analyse des résultats d'une action sociale : les indicateurs de rentabilité sont complètement secondaire, voire non-pertinent
- Présentation des résultats d'une recherche scientifiques : les exemples de KPI vu précédemment n'ont pas de sens, il faut redéfinir des KPI spécifique au problème posé.

- évaluer la performance
- réaliser un diagnostic de la situation
- communiquer, informer et motiver
- permettre un suivi continue

1 - Partir des données

Étudier la base de données pour voir ce qu'il est possible de faire

2 - Définir les indicateur à présenter

En discutant avec les personnes concernées (décideurs, public cible, etc...)

3 - Réaliser un schéma

Et réitérer sur l'étape précédente autant de fois que nécessaire

4 - Réaliser le dashboard

Et le rendre disponible aux usagers

Indicateurs chiffrés

- Chiffres bruts
- Pourcentages
- Ratio
- Corrélation
- ...

Graphiques

- courbes (classique, cumulative, log,...)
- histogrammes (et similaires)
- séries
- Gant
- ...

Ce qu'on veut

- Toutes les infos accessibles en 1 clic
- Les données avant tout ! (logo et autre en petit)
- Informations claire et hiérarchisée
- Graphiques et tableaux minimalistes
- Possibilité d'avoir plus de détails **à la demande** grâce à des options ou d'autres pages
- Utilisation de filtres ou de liste déroulantes pour modifier les données affichées

Ce qu'on ne veut pas

- Devoir choisir 15 filtres pour avoir une info pertinente
- Plein de pub pour la boîte qui occupe la moitié de l'écran
- Toutes les infos présentées sur le même plan
- Des effets de style et des fioritures sur les graphiques
- Des graphiques avec des légendes et labels surchargés
- Toutes les infos d'un seul coup sans sélection pertinente

Axes de sélection

- Les objectifs
- La nature des données
- Les besoins

Axes de contrainte

- La qualité des données
- La quantité de données disponibles
- La significativité des résultats

Généralités

- Quel est le contexte du dashboard ? Est-il commun à tous les utilisateurs ?
- Les abréviations sont-elles connues de tou·te·s ?
- Quels éléments de contexte sont à rappeler ?
- Quelles connaissances techniques sont nécessaires ?

Inclusivité

- Quelle palette de couleurs utiliser ?
- Quelle police utiliser ?
- Quel est le sens de lecture habituel des utilisateurs ?
- ...

Comment créer un dashboard ?

Logiciels de créations assistées

- Facile à utiliser
- Systèmes propriétaires (payant)
- Maintenance parfois externalisées
- Personnalisation limitée

API Javascript

- Personnalisation infinie
- Plus difficile à mettre en œuvre
- Contrôle totale de l'application
- Maintenance et sécurité interne
- Solutions libres existantes

- Tableau
- PowerBI
- Plotly
- Mode
- ...

Plotly

Plotly Open Source Graphing Libraries

Interactive charts and maps for Python, R, Julia, Javascript, ggplot2, F#,
MATLAB®, and Dash.



Plotly Python
Open Source
Graphing Library

Star 15,006



Plotly R
Open Source
Graphing Library

Star 2,483



Plotly Julia
Open Source
Graphing Library

Star 403



Plotly Javascript
Open Source
Graphing Library

Star 16,438

Grace à l'API dédiée

<https://cdn.plot.ly/plotly-2.30.0.min.js>

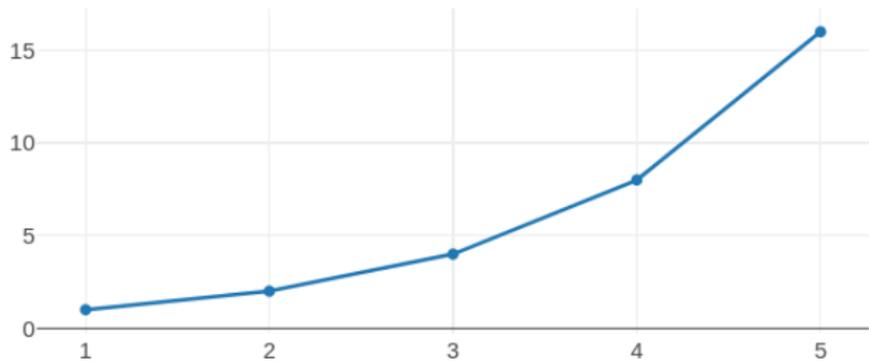
Dans le code

```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <meta charset="utf-8">
5     <title>Ma page</title>
6     <script src="https://cdn.plot.ly/plotly-2.30.0.min.js" charset="utf-8"></script>
7   </head>
8   <body>
9     <main>
10      <h1>Exemple 1</h1>
11      <div id="tester" style="width:600px;height:250px;"></div>
12    </main>
13    <script src="main.js"></script>
14  </body>
15 </html>
16
17
18
19
```

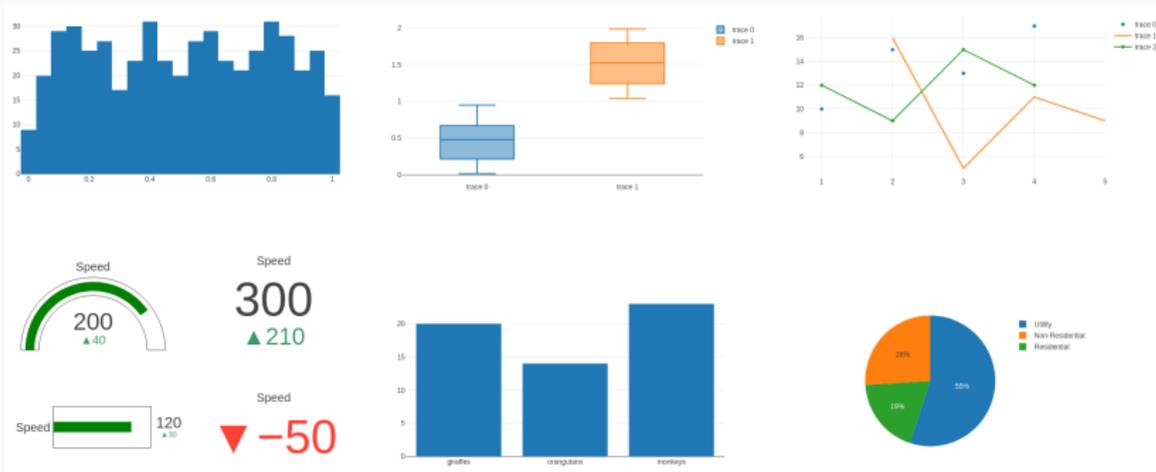
Premier exemple 1/2

```
1 TESTER = document.getElementById('tester');
2
3 Plotly.newPlot( TESTER, [{
4     x: [1, 2, 3, 4, 5],
5     y: [1, 2, 4, 8, 16] }], {
6     margin: { t: 0 } }
7 );
```

Exemple 1



Type de graphiques



Principe

Les graphiques sont décrits comme des objets JSON, c'est à dire avec un ensemble d'attributs ayant des valeurs.

Structures

Un graphique se décompose en général en plusieurs séries (les données) qui sont affichées dans un cadre (*layout*)

Pour les séries de données

- **type** : le type du graphique
- **x** et **y** : les séries de données
- **name** : le nom de la série de données

Pour les cadres (layout)

- **title** : titre du graphique
- **xaxis** et **yaxis** : attribut JSON pour paramétrer les axes
- **legend** : idem pour la légende

Pour les axes

- **range** : étendue de l'axe (liste de deux valeurs)
- **title** : titre de l'axe

Example

```
9 data = [  
10     {  
11         type: 'scatter',  
12         x: [1, 2, 3],  
13         y: [3, 1, 6],  
14     },  
15     {  
16         type: 'bar',  
17         x: [1, 2, 3],  
18         y: [3, 1, 6],  
19         name: 'bar chart example'  
20     }  
21 ];  
22  
23 layout = {  
24     title: 'simple example',  
25     xaxis: { title: 'time' }  
26 }
```

Réaliser des graphiques

Contexte

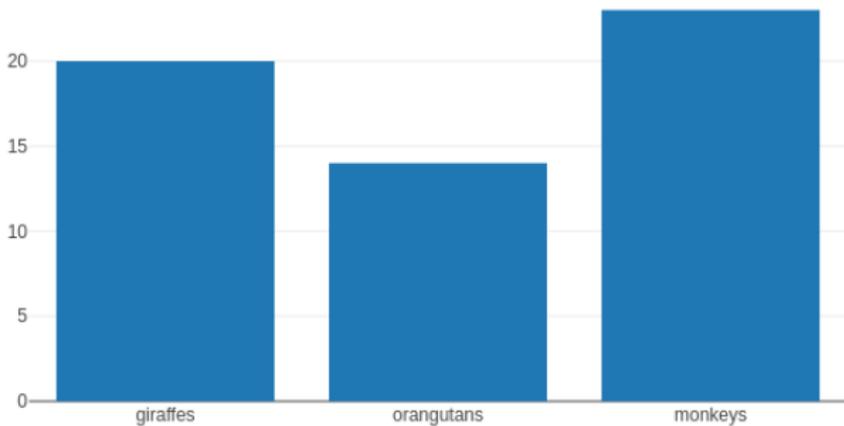
- Mise en place d'un dashboard dans un contexte numérique
- Récupération de données depuis des bases ou un stream

Réalisation

- Plotly : réalisation des graphiques
- Bootstrap : mise en forme de la page
- Javascript : dynamisme de la page, réponse aux événements*

(*même si plotly gère déjà beaucoup)

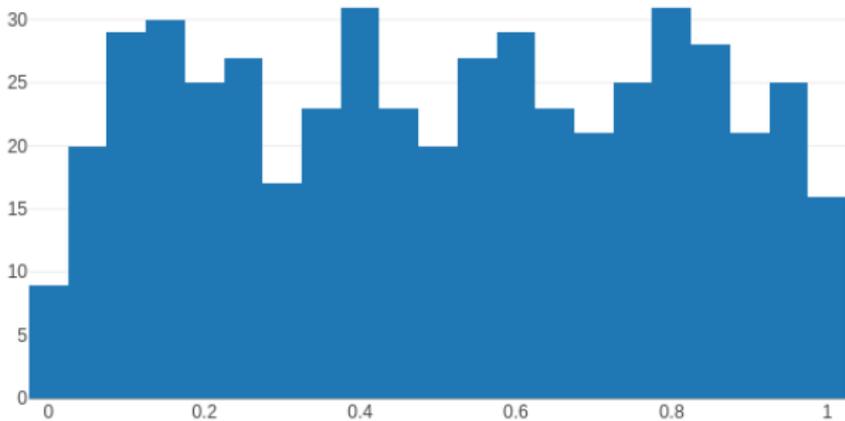
Diagramme en barre



Code du diagramme en barre

```
3 var data = [  
4   {  
5     x: ['giraffes', 'orangutans', 'monkeys'],  
6     y: [20, 14, 23],  
7     type: 'bar'  
8   }  
9 ];
```

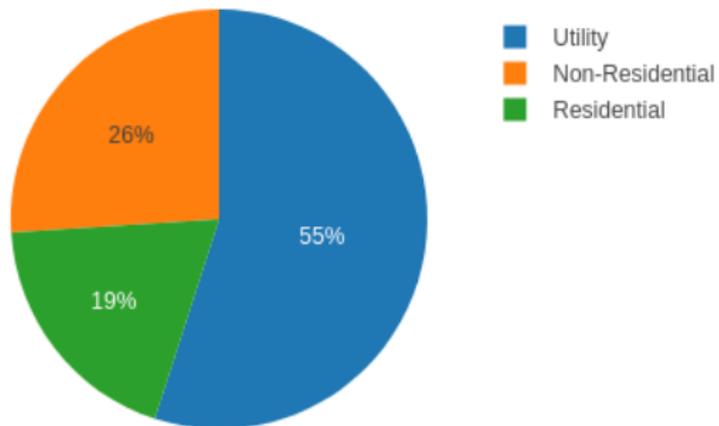
Histogramme



Code de l'histogramme

```
3 var x = [];  
4  
5 for (var i = 0; i < 500; i ++) {  
6     x[i] = Math.random();  
7 }  
8  
9 var trace = {  
10     x: x,  
11     type: 'histogram',  
12 };  
13  
14 var data = [trace];
```

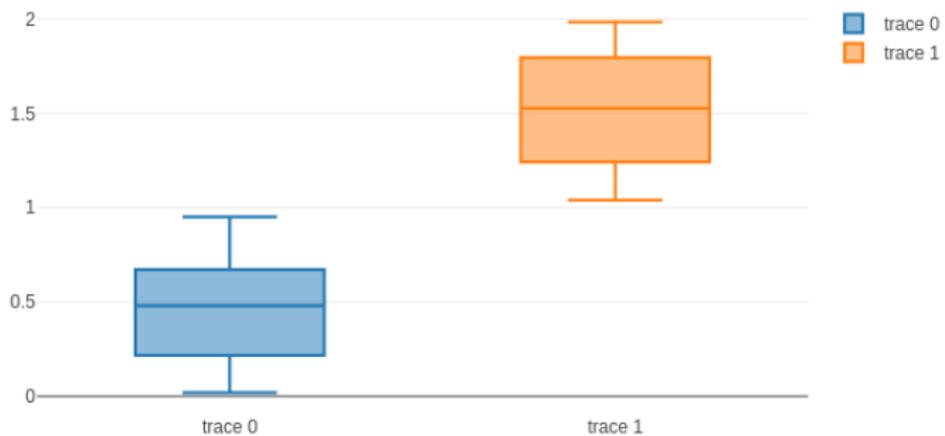
Piechart



Code du piechart

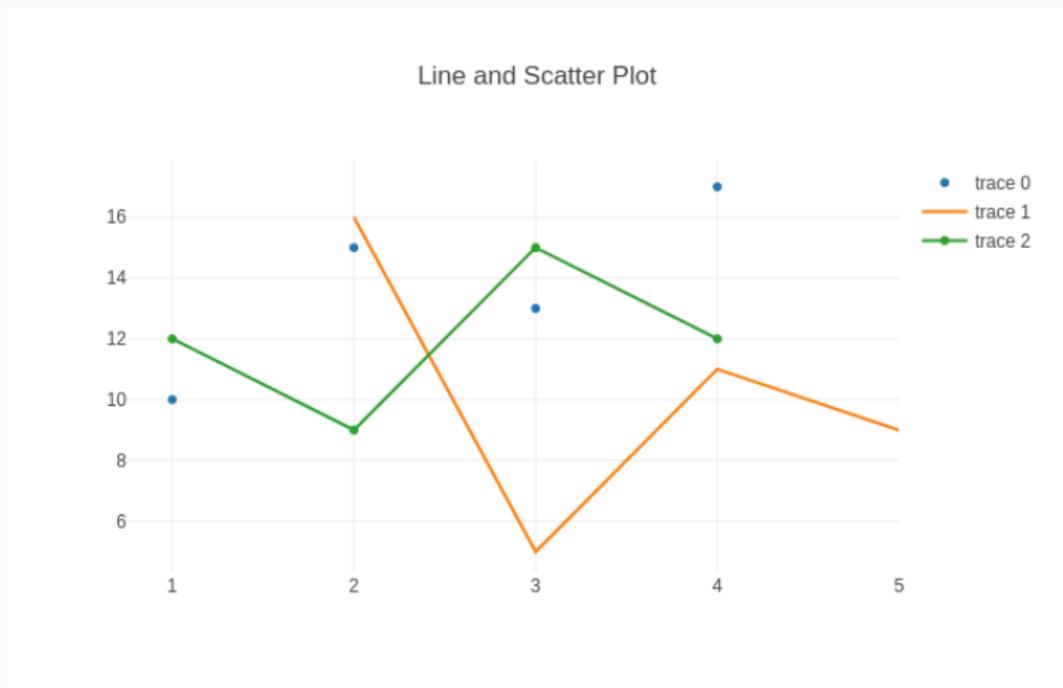
```
3 var data = [{
4   values: [19, 26, 55],
5   labels: ['Residential', 'Non-Residential', 'Utility'],
6   type: 'pie'
7 }];
8
9 var layout = {
10  height: 400,
11  width: 500
12 };
13
14 Plotly.newPlot(TESTER, data, layout);
```

Boxplot



Code du boxplot

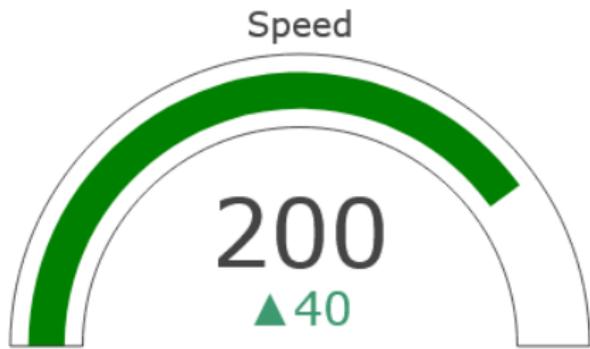
```
3  var x = [];  
4  
5  for (var i = 0; i < 500; i ++) {  
6      x[i] = Math.random();  
7  }  
8  
9  var trace = {  
10     x: x,  
11     type: 'box',  
12     };  
13  
14  var data = [trace];
```



Code pour les courbes

```
3 var trace1 = {
4   x: [1, 2, 3, 4],
5   y: [10, 15, 13, 17],
6   mode: 'markers',
7   type: 'scatter'
8 };
9
10 var trace2 = {
11   x: [2, 3, 4, 5],
12   y: [16, 5, 11, 9],
13   mode: 'lines',
14   type: 'scatter'
15 };
16
17
18 var trace3 = {
19   x: [1, 2, 3, 4],
20   y: [12, 9, 15, 12],
21   mode: 'lines+markers',
22   type: 'scatter'
23 };
24
25 var data = [trace1, trace2, trace3];
```

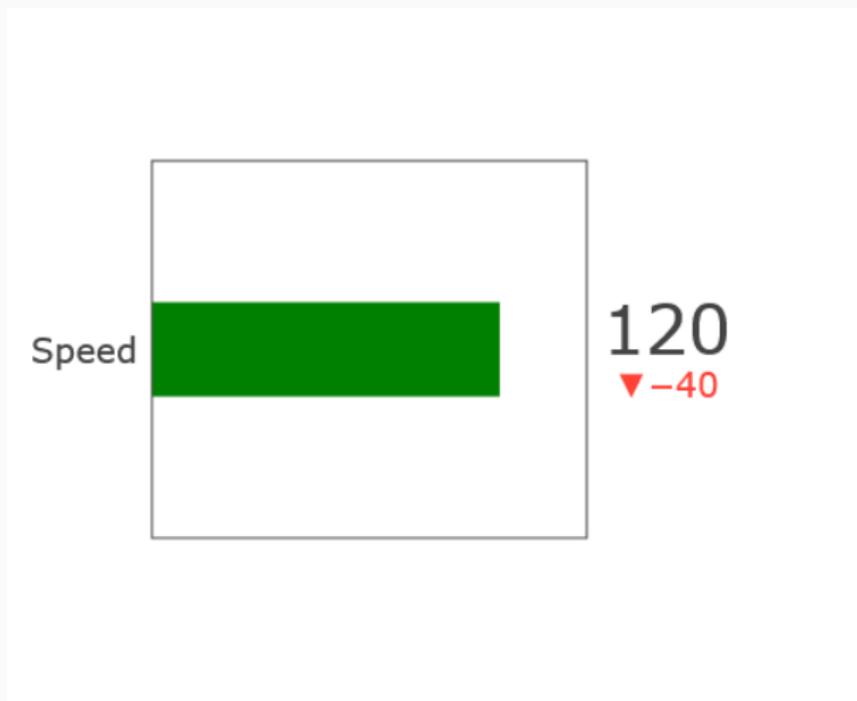
Indicateurs : jauge



Code des indicateur : jauge

```
3 var data = [{
4   type: "indicator",
5   value: 200,
6   delta: { reference: 160 },
7   gauge: { axis: { visible: false, range: [0, 250] } },
8   domain: { row: 0, column: 0 }
9   }];
10
11 var layout = {
12   height: 400,
13   width: 500,
14   template: {
15     data: {
16       indicator: [
17         {
18           title: { text: "Speed" },
19           mode: "number+delta+gauge",
20         }
21       ]
22     }
23   }
24 };
25
26 Plotly.newPlot(TESTER, data, layout);
```

Indicateurs : barre

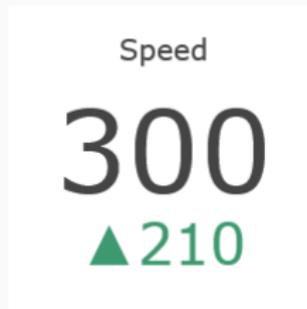


Code des indicateur : barre

```
3 var data = [{
4   type: "indicator",
5   value: 120,
6   delta: { reference: 160 },
7   gauge: {
8     shape: "bullet",
9     axis: {
10      visible: false,
11      range: [-200, 200]
12    }
13  },
14  domain: { row: 0, column: 0 }
15 }];
16
17 var layout = {
18   height: 400,
19   width: 500,
20   template: {
21     data: {
22       indicator: [
23         {
24           title: { text: "Speed" },
25           mode: "number+delta+gauge",
26         }
27       ]
28     }
29   }
30 };
```

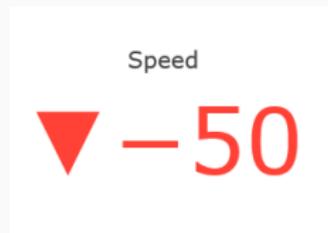
Indicateur : nombre

```
3  var data = [{
4      type: "indicator",
5      mode: "number+delta",
6      value: 300,
7      domain: { row: 0, column: 1 },
8      delta: { reference: 90 }
9  }
10 ];
11
12 var layout = {
13     height: 400,
14     width: 500,
15     template: {
16         data: {
17             indicator: [
18                 {
19                     title: { text: "Speed" },
20                     mode: "number+delta+gauge",
21                 }
22             ]
23         }
24     }
25 };
```



Indicateur : Delta

```
3 var data = [{
4   type: "indicator",
5   mode: "delta",
6   value: 40,
7   domain: { row: 1, column: 1 },
8   delta: {reference: 90} }
9 ];
10
11 var layout = {
12   height: 400,
13   width: 500,
14   template: {
15     data: {
16       indicator: [
17         {
18           title: { text: "Speed" },
19           mode: "number+delta+gauge",
20         }
21       ]
22     }
23   }
24 };
```



La doc

`https://plotly.com/javascript/`

This is it 😊