

*Santé environnement*

# Recommandations pour la représentation des résultats au DSE

# Sommaire

Abréviations	2	3.3 Représentation d'une démarche au moyen de schémas ou de tableaux synthétiques	97
<b>1. Introduction</b>	<b>4</b>	3.3.1 Exemples de schémas utilisés pour résumer une démarche dans une EIS	97
<b>2. Rappel des bonnes pratiques en matière de représentation graphique des résultats</b>	<b>5</b>	3.3.2 Utilisation de schéma pour résumer une démarche scientifique en biosurveillance	101
2.1 Choix de la méthode de représentation	5	3.3.3 Utilisation de schémas en EQRS pour s'adresser au grand public	101
2.1.1 Représenter la distribution statistique d'une variable ou plusieurs variables	7	<b>4. Les nouvelles possibilités de visualisation des résultats</b>	<b>107</b>
2.1.2 Représenter une relation entre des variables	17	4.1 De nouvelles possibilités techniques et de nouveaux outils pour des données qui évoluent	107
2.1.3 Effectuer une comparaison	20	4.2 Des modes de représentation à explorer dans le champ de la santé-environnement	113
2.1.4 Représenter une décomposition	21	4.3 Des modes de représentation spécifiques liés à l'émergence du « big data »	114
2.1.5 Représenter une évolution	25	4.4 Les limites des nouvelles possibilités	114
2.1.6 Représenter l'incertitude	32	<b>5. Conclusion</b>	<b>116</b>
2.2 Mise en œuvre technique	35	Références bibliographiques	117
2.2.1 Les composantes du graphique	35	Ressources complémentaires sur Internet	118
2.2.2 Choix des outils	40	Annexes	120
2.2.3 Précautions à prendre, écueils à éviter	40		
2.3 S'adapter à son public : un enjeu de taille	41		
<b>3. Etat de l'art au DSE : les modes de représentations graphiques couramment utilisés</b>	<b>43</b>		
3.1 Représentations graphiques	43		
3.1.1 Représentations graphiques des résultats dans le cadre de l'investigation de syndromes collectifs inexplicables	43		
3.1.2 Représentations graphiques dans le champ des études d'impact sanitaire dans les domaines de l'air et du climat	48		
3.1.3 La représentation graphique des résultats dans le champ de l'EQRS	57		
3.1.4 Exemples de représentations graphiques dans le champ de la biosurveillance et des indicateurs de santé	69		
3.1.5 Exemples de représentations graphiques dans le champ de la toxicovigilance	78		
3.1.6 Exemples de représentations graphiques dans le champ de la surveillance du climat	87		
3.2 Représentations cartographiques	92		
3.2.1 Combinaison d'informations dans une même carte	92		
3.2.2 Représentations cartographiques dans le champ des études d'impact sanitaire	93		
3.2.3 Représentation cartographique combinée d'une tendance et de la significativité de l'information	96		

# Recommandations pour la représentation des résultats au DSE

## **Auteur :**

Perrine de Crouy-Chanel — Département santé environnement (DSE), Institut de veille sanitaire (InVS).

## **Ont participé à ce rapport :**

Jamel Daoudi, Cécile Kairo, Florence Kermarec, Alain Le Tertre, Mathilde Pascal, Marie Pécheux, Annabel Rigou — DSE, InVS.

## **Relecture :**

Daniel Eilstein — Direction scientifique et de la qualité (Disq), InVS.

## **Remerciements :**

Sylvia Medina — DSE, InVS.

# Abréviations

<b>ACHIA</b>	Air Climate Health Impact Assessment (Évaluation d'impact sanitaire air et climat)
<b>AEE</b>	Agence européenne de l'environnement
<b>AMP</b>	Assistance médicale à la procréation
<b>Anses</b>	Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail
<b>AP</b>	Administration pénitentiaire
<b>ARS</b>	Agence régionale de santé
<b>BPA</b>	Bisphénol A
<b>CAP</b>	Centre antipoison
<b>CAP TV</b>	Centre antipoison et de toxicovigilance
<b>CG28</b>	Conseil général de l'Eure-et-Loir
<b>Cire</b>	Cellule de l'InVS en région
<b>Cire-IDF</b>	Cire Ile-de-France
<b>CLOI</b>	Commission locale d'orientation et d'information
<b>CSP</b>	Catégorie socioprofessionnelle
<b>CST</b>	Comité scientifique et technique
<b>DGCCRF</b>	Direction générale de la concurrence, de la consommation et de la répression des fraudes
<b>DSE</b>	Département santé environnement
<b>DST</b>	Département santé travail
<b>EHP</b>	Environmental Health Perspective
<b>EIS</b>	Évaluation d'impact sanitaire
<b>Elfe</b>	Étude longitudinale française depuis l'enfance
<b>EQRS</b>	Évaluation quantitative des risques sanitaires
<b>ETM</b>	Éléments traces métalliques
<b>HAP</b>	Hydrocarbures aromatiques polycycliques
<b>HTML</b>	Hypertext Markup Language
<b>Iris</b>	Îlots regroupés pour l'information statistique
<b>Isee</b>	International Society for Environmental Epidemiology
<b>MoE</b>	Marge of exposure

<b>NOAEL</b>	No Observed Adverse Effect Level
<b>OMS</b>	Organisation mondiale de la santé
<b>OSCOUR<sup>®</sup></b>	Organisation de la surveillance coordonnée des urgences
<b>PDF</b>	Portable Document Format
<b>PSAS</b>	Programme de surveillance air et santé
<b>QD</b>	Quotient de danger, utilisé dans les EQRS
<b>RFID</b>	Radio-Frequency IDentification
<b>SICAP</b>	Système d'information du réseau des CAP TV
<b>SIG</b>	Système d'information géographique
<b>Sniir-AM</b>	Système national d'information inter-régimes de l'Assurance maladie
<b>VTR</b>	Valeur toxicologique de référence

# 1. Introduction

Le Département santé environnement (DSE) de l'Institut de veille sanitaire (InVS) produit de nombreux travaux dans le domaine de la santé environnementale. Les résultats de ces travaux sont publiés sous des formes diverses et variées et s'adressent, en fonction des thématiques, des enjeux et des supports, à des publics également variés. De plus, de la diversité et de la complexité du domaine et des thématiques impliquées, découle une multiplicité de résultats produits et, de ce fait, des possibilités pour les représenter. Or souvent, de par la complexité du domaine, les résultats de ces travaux sont difficiles à présenter et à communiquer aux publics non-spécialistes.

Parallèlement, l'émergence dans l'univers des données en général, du *big data* (données massives) d'une part, et de l'*open-data* (ouverture des données) d'autre part, a entraîné, ces dernières années, le développement de nouveaux outils de représentation graphique des données. Certains sont accessibles à tous, librement, sur internet, et d'autres ne le sont pas. Mais, quand bien même un mode de représentation est accessible à chacun, il demeure souvent difficile au niveau individuel de savoir lequel est le plus approprié à une situation donnée et comment l'utiliser simplement. De manière générale, le choix d'une représentation graphique plutôt qu'une autre, en vue de communiquer des résultats, implique de s'être posé un certain nombre de questions préalables quant aux objectifs de communication que vise cette représentation, de maîtriser la réalisation de la représentation en elle-même – ce qui passe éventuellement par la maîtrise technique d'un outil spécifique – et de s'être assuré que le public destinataire sera pleinement en mesure de comprendre et d'interpréter la représentation adoptée.

Dans ce contexte, l'objectif de ce travail est de proposer un recueil d'exemples d'utilisation des outils les plus adaptées aux situations spécifiques les plus fréquemment rencontrées en santé environnement, de présenter pour chacun la méthode et les outils qui en permettent la réalisation.

Prévu initialement pour ne couvrir que le champ des graphiques *stricto-sensu*, la réflexion s'est élargie également aux schémas, ainsi que, dans une moindre mesure, aux cartes, puisque ces formalisations de l'information constituent une manière parmi d'autres de communiquer un résultat. Pour les cartes cependant, l'existence, par ailleurs, de documents disponibles concernant les bonnes pratiques [9;15] fait que le sujet n'est que brièvement abordé ici.

Un recensement de situations variées, propres aux problématiques de travail des différentes unités composant le DSE a également été réalisé, dans l'idée de mieux connaître les besoins des scientifiques, et de leur proposer, quand c'était possible ou nécessaire, des améliorations.

En rappelant les bonnes pratiques en matière de représentation graphique, ce travail vise également à harmoniser les représentations graphiques produites au sein du département. Enfin, en présentant de nouvelles possibilités de représentation, ce travail permet un élargissement du champ des possibilités en matière de communication des résultats. De ce fait, l'émergence de nouveaux modes de représentations se faisant en continu, il est difficile d'imaginer que le document issu de ces réflexions soit un document figé. Aussi, il est prévu, dans le cadre de ce travail, de mettre également à disposition le contenu sous la forme d'un outil de consultation alternatif au guide papier (type jeu de pages HTML) qu'il sera possible de faire évoluer en continu.

## 2. Rappel des bonnes pratiques en matière de représentation graphique des résultats

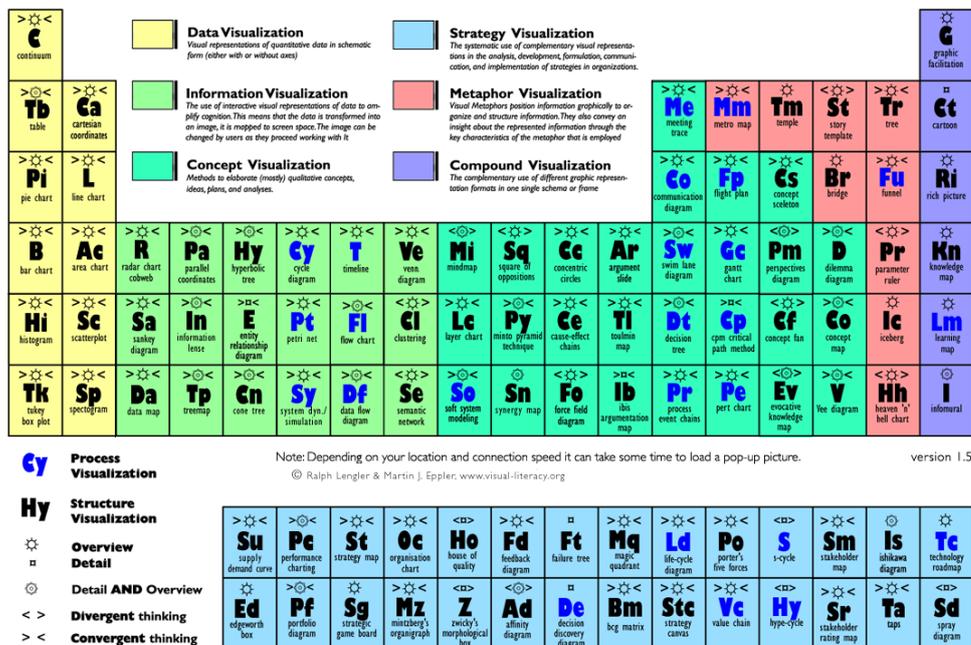
### 2.1 Choix de la méthode de représentation

La littérature propose un certain nombre de schémas synthétiques regroupant une très large variété de modes de représentation graphique, catégorisés afin d'aider au choix d'un graphique adapté aux données et à ce que l'on souhaite montrer. On peut citer, entre autres, la table périodique des méthodes de visualisation et de représentation de Lengler et Eppler<sup>1</sup> (figure 1).

Figure 1 |

Sélection d'un mode de représentation en tableau périodique des éléments

#### A PERIODIC TABLE OF VISUALIZATION METHODS



En se rendant sur la page web de ce site, on accède aux types de graphiques avec un exemple en passant le pointeur de la souris sur chacune des cases de la table<sup>2</sup>.

Le spectre couvert est très large, la représentation simple des données est abordée tout comme l'illustration de concepts ou d'idées maîtresses.

Le choix d'un mode de représentation graphique est pleinement déterminé par les objectifs de la représentation et l'information que l'on souhaite mettre en avant. Comme le montre le schéma

<sup>1</sup> <http://www.visual-literacy.org/>

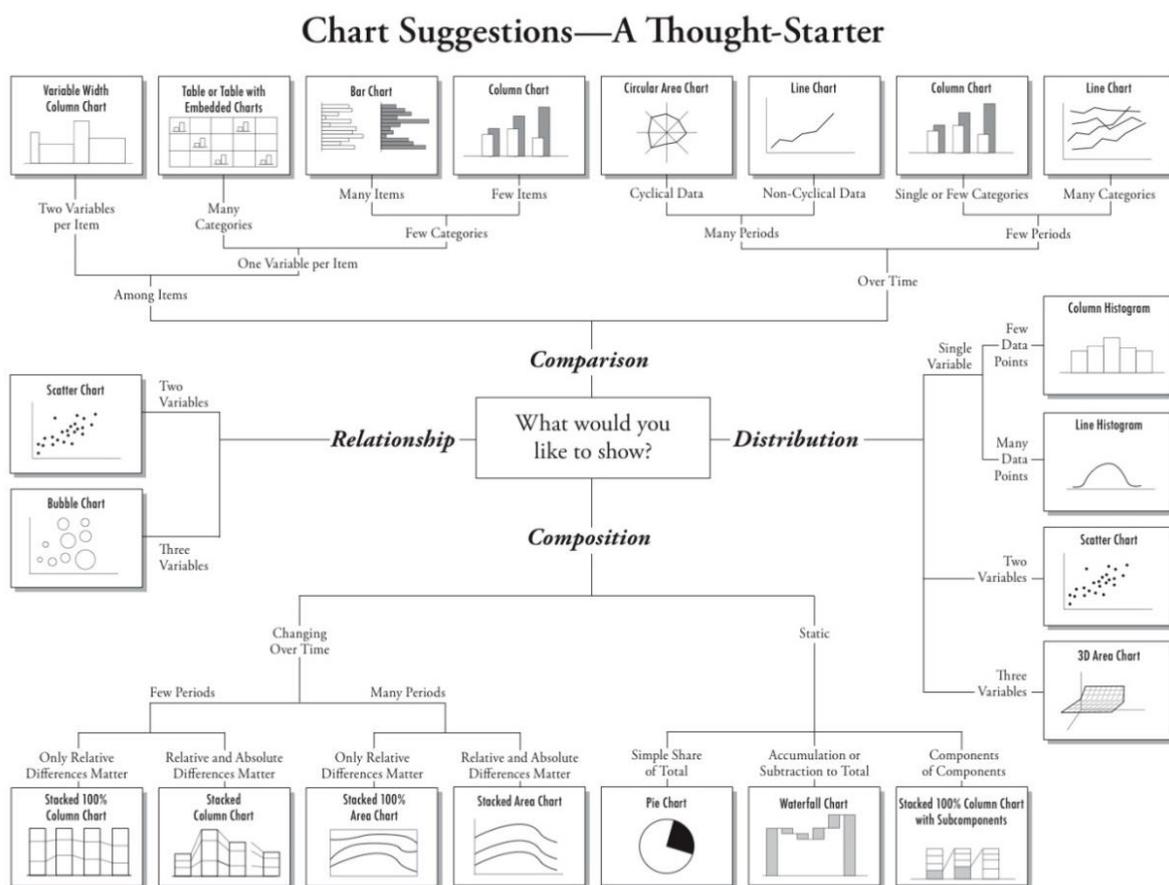
<sup>2</sup> [http://www.visual-literacy.org/periodic\\_table/periodic\\_table.html](http://www.visual-literacy.org/periodic_table/periodic_table.html)

heuristique ci-dessous (figure 2), une question essentielle préalable à ce choix est donc : que veut-on montrer ?

Ces schémas synthétiques peuvent être mobilisés en amont du choix d'un mode de représentation graphique. Les possibilités qui s'offrent sont souvent multiples, c'est d'ailleurs ce que montrent ces schémas. C'est là que la question du public destinataire du graphique peut être envisagée : en effet, le fait d'être familier ou pas d'un mode de représentation graphique va souvent compter pour beaucoup dans la capacité à le lire et à le comprendre correctement ou non.

Figure 2 |

### Choisir un mode de représentation



(Source : © Conception A. Abela a.v.abela@gmail.com – www.extremepresentation.com)

Dans le choix entre les différentes possibilités qui s'offrent lorsque l'on a répondu à la question « que veut-on montrer ? », il est nécessaire de s'interroger également sur le public destinataire : sera-t-il familier du mode de représentation choisi ou non, va-t-il être en mesure de l'interpréter correctement ou non ? Si l'interlocuteur n'est pas familier du mode de représentation choisi, comment accompagner le graphique d'une explication qui lui permettra de mieux le comprendre, ou encore, comment envisager de choisir un autre mode de représentation ?

Des rappels de bonnes pratiques pour la représentation des graphiques les plus couramment utilisés au DSE sont illustrés ci-dessous par des exemples tirés de productions du DSE. Plusieurs outils permettent de réaliser ces graphiques de manière interchangeable (R, Stata, SAS, etc.), et le choix d'un outil dépend des habitudes de la personne qui l'utilise. On verra par la suite que, bien que cet outil ait l'avantage d'être relativement simple à prendre en mains et à utiliser, il convient de se méfier

du logiciel bureautique Excel qui, dans ses fonctions standard, offre moins de possibilités techniques tout en augmentant le risque d'erreur.

## 2.1.1 Représenter la distribution statistique d'une variable ou plusieurs variables

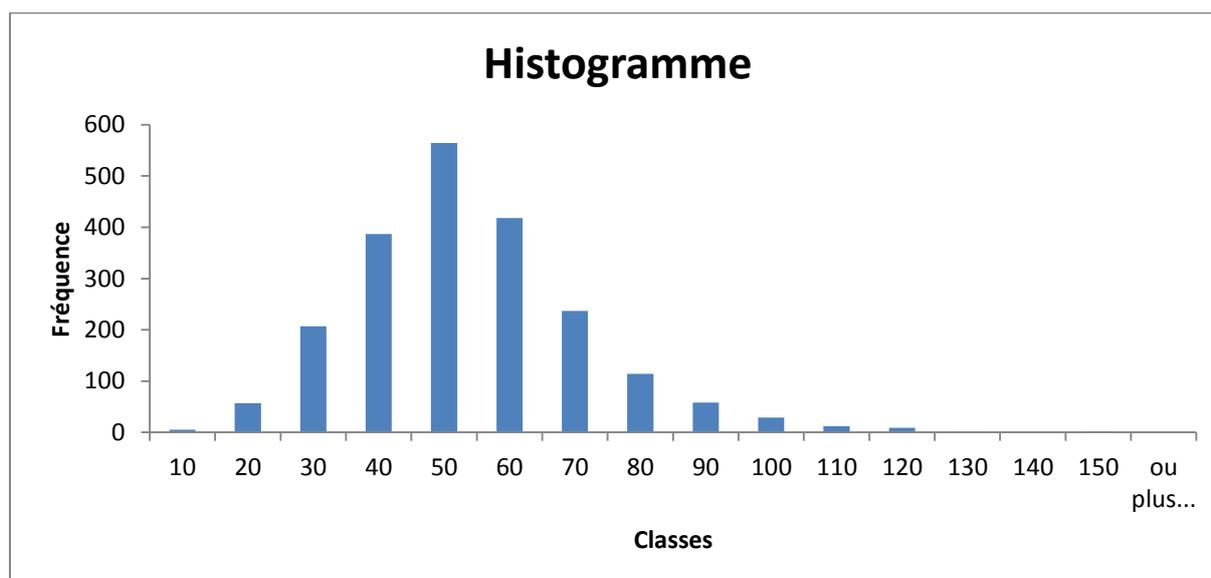
### Distribution statistique d'une variable

#### Histogramme

La méthode de représentation la plus usuelle pour représenter une distribution statistique des valeurs prises pour une variable quantitative continue est l'histogramme. Dans un histogramme, la surface des barres est proportionnelle à la **fréquence** de classe. Les barres sont contiguës de manière à illustrer formellement le caractère continu de la variable. On privilégiera une représentation avec une largeur de barres égale, cependant, lorsque l'amplitude des classes représentées sur le graphique n'est pas constante il est nécessaire de l'ajuster sur le graphique de manière à maintenir la proportionnalité de la **surface** des barres et non pas de la **hauteur** des barres.

I Figure 3 I

L'histogramme obtenu avec Excel



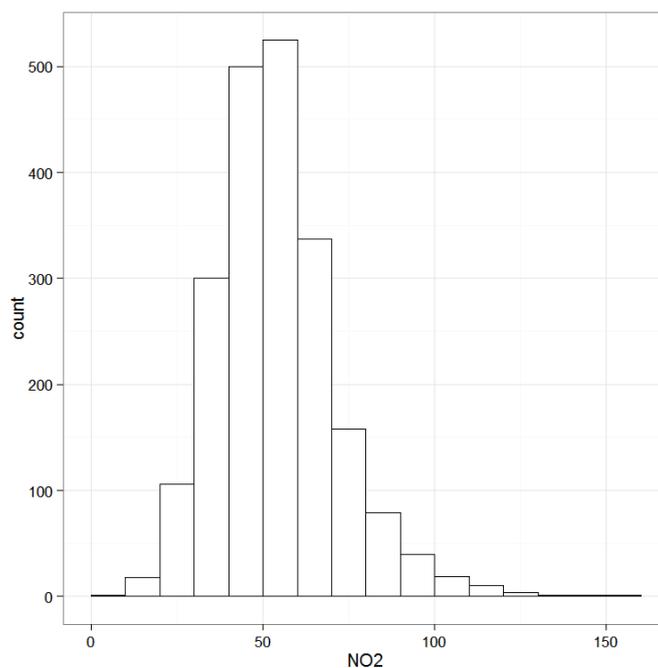
Excel ne calcule pas automatiquement un histogramme. Soit la procédure implique un comptage des valeurs, soit il faut passer par une extension qui réalise cette étape. Le graphique nécessite de définir des classes. Par ailleurs, comme le montre l'exemple ci-dessus (figure 3), alors que la série est continue, Excel génère des blancs entre les différentes barres. Il convient donc d'éviter d'utiliser Excel pour réaliser un histogramme, car ce logiciel ne calcule pas d'histogramme mais des diagrammes en barres. À l'inverse, un graphique équivalent produit sous R rend aisée la lecture de la distribution de la variable. Les classes sont délibérément identiques à celles d'Excel afin de favoriser la comparaison (figure 4).

D'autres histogrammes, produits dans Stata, sont également présentés. Le premier graphique (figure 5) représente la distribution des niveaux en  $PM_{2,5}$  modélisés par GAZEL [16] dans les communes d'Ile-de-France, sur l'ensemble de la région. Un rendu équivalent (couleurs, mises en

forme, etc.) aurait tout aussi bien pu être obtenu avec R. Le second graphique (figure 6) quant à lui est en fait une combinaison d'histogrammes qui effectue un « focus » sur les départements de Paris et de la petite couronne.

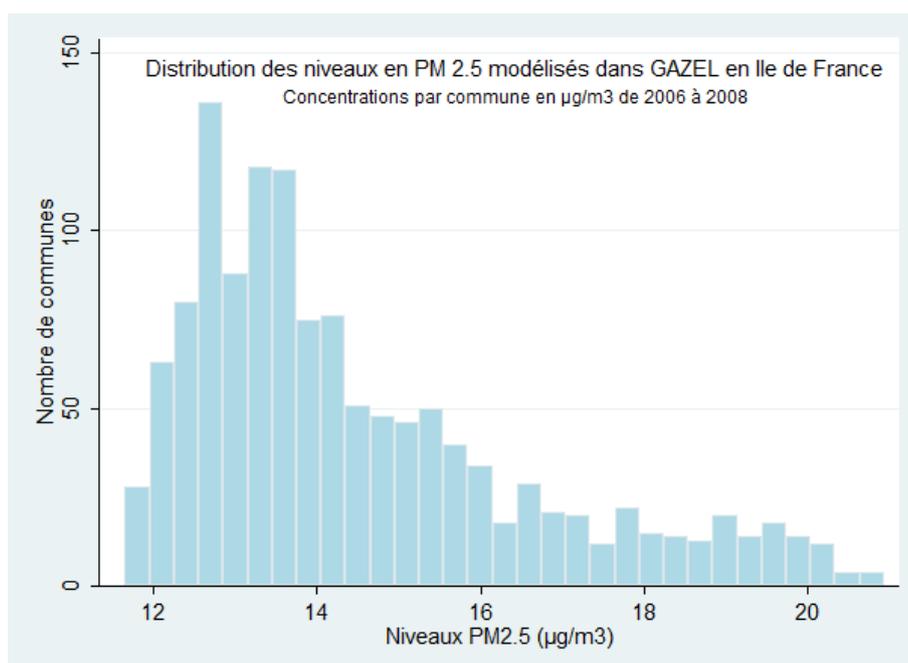
#### I Figure 4 I

Exemple d'histogramme obtenu dans R



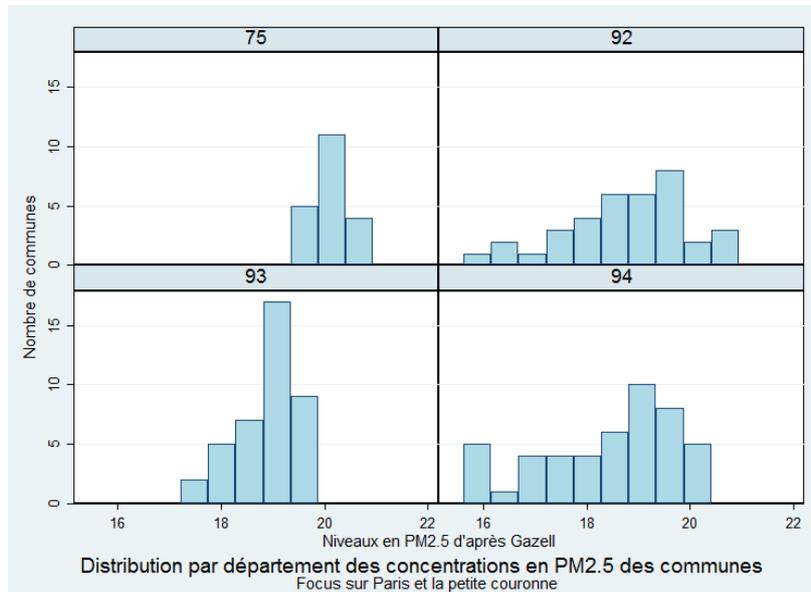
#### I Figure 5I

Exemple d'histogramme obtenu dans STATA



## I Figure 6 I

Collection d'histogrammes pour illustrer la distribution d'une même variable dans différents départements

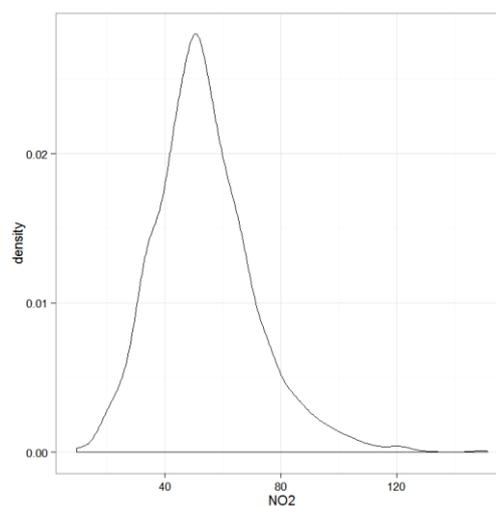


### Courbe de distribution

Une autre manière classique de représenter la distribution d'une variable est de recourir à une courbe de distribution dite « de densité ». La courbe de distribution est une conversion en courbe de l'histogramme de distribution. Elle est particulièrement adaptée dans le cas où le nombre d'observations est élevé et/ou l'étendue de la série est très importante. L'exemple ci-après figure la conversion en courbe de l'histogramme de distribution de la figure 4, réalisée dans le logiciel R. Excel, quant à lui, ne permet pas de réaliser ce genre de graphique (figure 7).

## I Figure 7 I

Courbe de densité de distribution

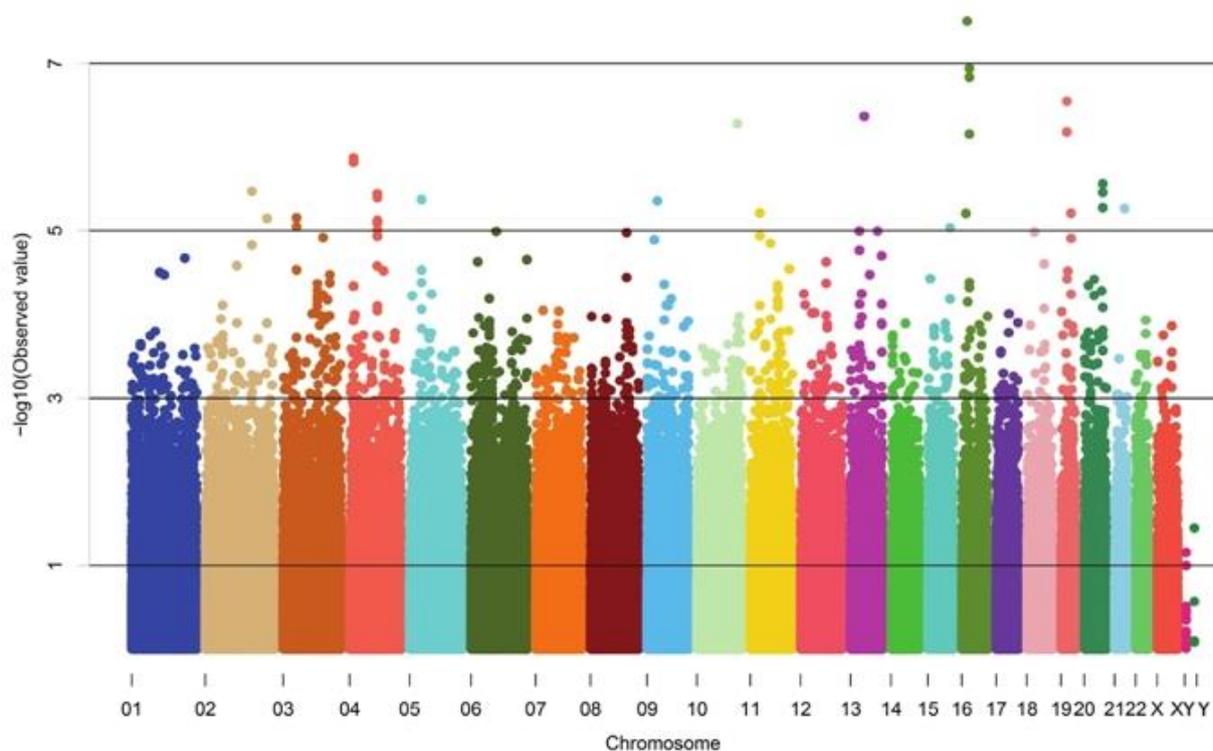


## Diagramme de Manhattan (Manhattan Plot)

Les diagrammes en Manhattan plot (figure 8) sont utilisés essentiellement en génétique. Ils sont en effet particulièrement adaptés à la visualisation de la distribution des très nombreuses informations compilées dans un génome afin d'étudier le lien entre les variations génétiques et le risque de développer une maladie. Dans ce type de représentation, chaque valeur de chaque individu pour chacun des caractères étudiés est représentée par un point. La largeur de chaque bande correspond au nombre de caractères étudiés pour chaque chromosome. Les points isolés dont les valeurs en Y sont les plus élevées correspondent aux positions, sur les chromosomes, pour lesquelles l'association avec la pathologie étudiée est la plus significative (l'axe des Y dans ce type de graphique utilisé en génomique étant le plus souvent une conversion logarithmique de la p-value). L'exemple présenté ici est extrait du site de l'université de Glasgow (Royaume-Uni) et porte sur la confrontation des génomes de près de 37 000 cas et contrôles dans le cadre d'une étude portant sur les facteurs génétiques de l'hypertension artérielle<sup>3</sup>.

I Figure 8 I

Exemple de Manhattan plot



## Graphique en box plot

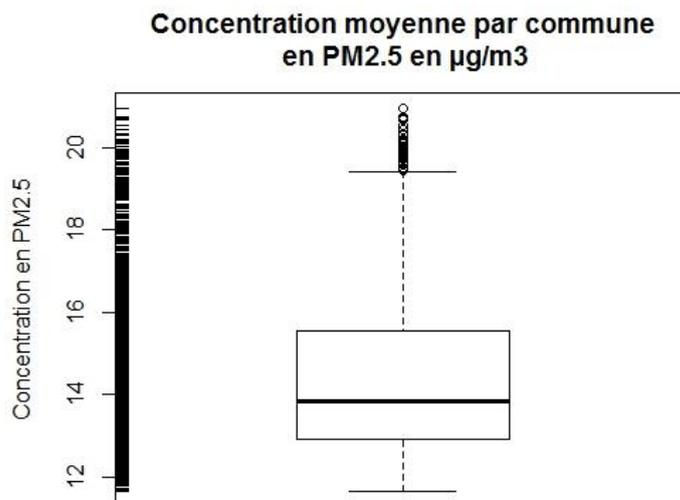
Les graphiques en box plot ou « boîte à moustaches » fournissent des informations sur la distribution statistique d'une série de données en permettant la visualisation de certains indicateurs de la dispersion. Ils sont souvent utilisés pour comparer entre elles les distributions de plusieurs variables ayant trait aux mêmes individus. Ils sont une version enrichie du simple « dot plot » qui « empile » sur même axe l'ensemble des valeurs prises par une même variable. Voici ci-après un exemple simple de

<sup>3</sup> <http://www.gla.ac.uk/researchinstitutes/icams/research/geneticsgenomicsandsystemsmedicineresearchtheme/ourresearch/>

box plot, réalisé sur une série de données de concentration en  $PM_{2,5}$  modélisées et agrégées à la maille de la commune, sur la région Ile-de-France. Ce graphique a été réalisé dans R. Le premier visuel (figure 9) fournit le graphique « brut », le second (figure 10) donne au lecteur quelques clefs pour une meilleure compréhension de ce type de graphique. Les barbules (segments de droite) figurées sur l'axe des ordonnées représentent les individus considérés (ici, il s'agit donc des communes d'Ile-de-France).

I Figure 9 I

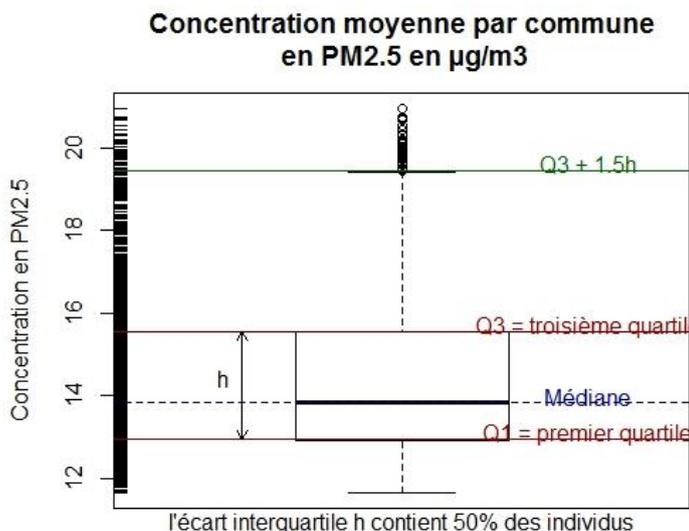
Exemple de graphique en « box plot » ou boîte à moustaches



Le graphique en box plot se lit de la manière suivante :

I Figure 10 I

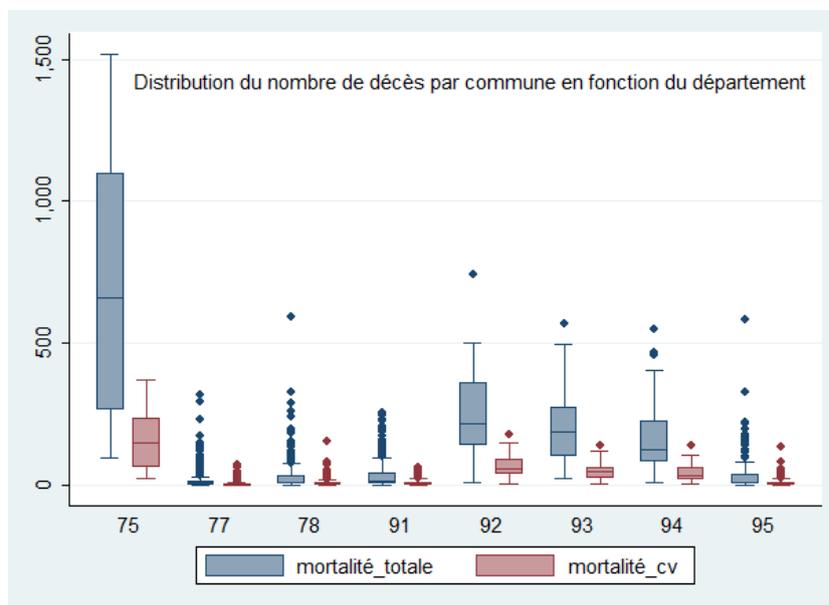
Clefs de lecture d'un graphique en box plot (source des données : GAZEL)



Des graphiques en box plot peuvent bien entendu également être réalisés dans STATA, comme c'est le cas de l'exemple ci-après (figure 11).

### I Figure 11 I

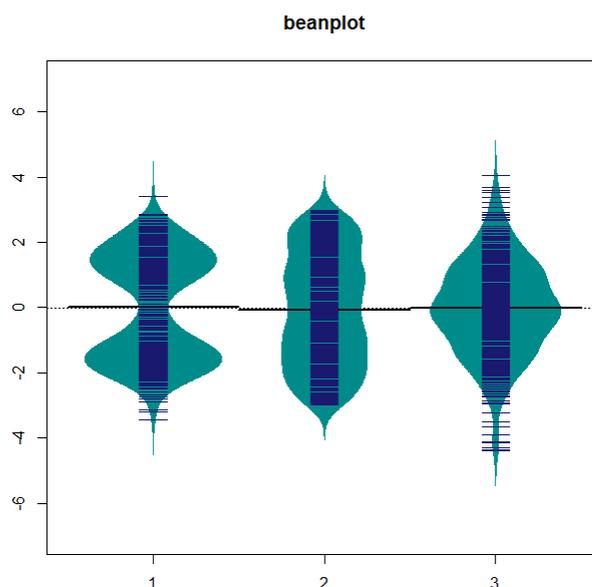
Autre exemple de graphique en box plot réalisés dans STATA



Variante du graphique en box plot : le bean plot (figure 12), est réalisé dans cet exemple dans R. Sur ce graphique, chaque individu de la série est figuré par une barre de couleur bleu foncé.

### I Figure 12 I

Graphique en bean plot



Les violin plots (figure 13) sont également une variante des box plots. Sur ce graphique, la valeur médiane de chacune des 5 séries considérées est représentée par un point rouge foncé, le rectangle

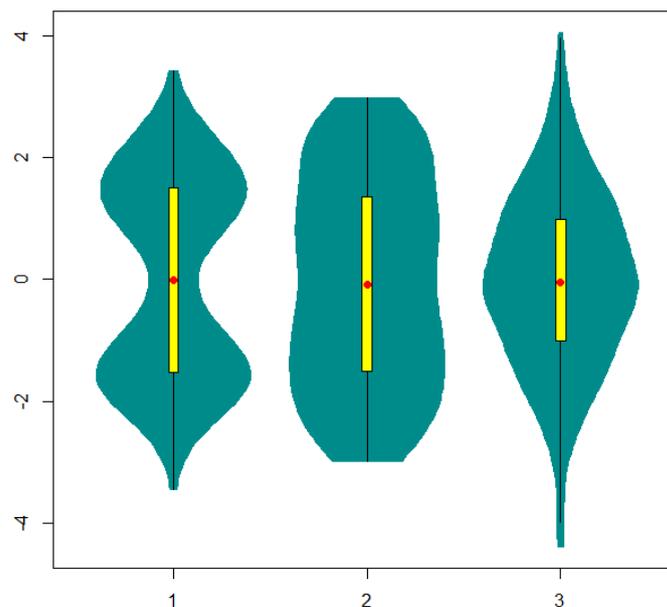
jaune figurant l'écart entre le 1<sup>er</sup> et le 3<sup>e</sup> quartile. Les valeurs prises par les variables étudiées sont mentionnées sur l'axe des ordonnées.

Les box plots sont les plus usités mais, en général, ils sont ceux qui présentent le moins d'information sur la distribution d'une variable. En effet, plus celle-ci s'écarte d'une distribution normale, moins le box plot est à même de la représenter de manière adéquate. Ainsi, la distribution bimodale (groupe 1) et uniforme (groupe 2) du graphique bean plot (et du violon plot) auront exactement la même représentation dans un box plot.

Les bean plot et violon plot n'axent plus leur représentation, à l'inverse des box plots, sur les quantiles d'une distribution, mais sur sa densité (la forme) de celle-ci. Le bean plot se différencie du violon plot par l'information secondaire ajoutée en plus de cette densité. Pour les premiers, il s'agit de marques pour chaque observation alors que pour le second, un box plot est intégré. Pour de faibles effectifs, le bean plot peut être plus pertinent, et le violon plot pour un effectif élevé. Le choix doit aussi s'effectuer en fonction de son souhait de représentation : soit on souhaite conserver le caractère individuel des données soit on souhaite en faire un résumé.

I Figure 13 I

Exemple de graphique en violon plot réalisé sur les mêmes données



### Distribution statistique de plusieurs variables

#### **Nuage de points**

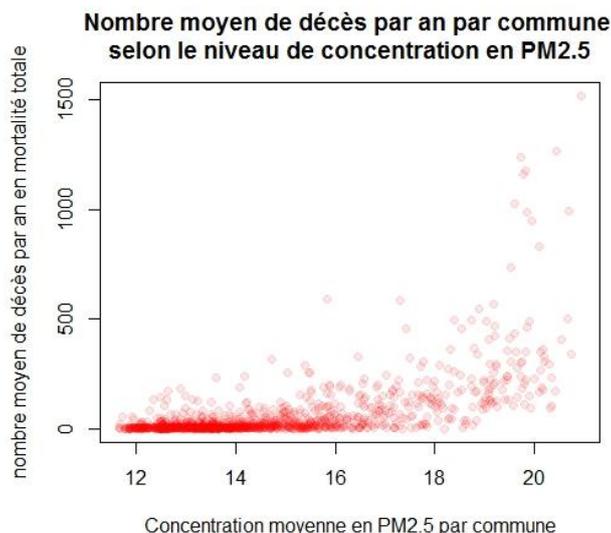
Pour visualiser la distribution de deux variables, la représentation par un nuage de points est la plus simple. Elle est également utilisée pour visualiser une éventuelle relation entre les deux variables représentées simultanément. Ici, l'exemple choisi (figure 14) est un nuage de points réalisé dans R mettant en relation le nombre moyen annuel de décès des 30 ans et plus (toutes causes confondues), dans les 1 300 communes d'Ile-de-France avec la distribution de la concentration moyenne en PM<sub>2,5</sub> de ces mêmes communes modélisées dans GAZEL. On remarque sans surprise que les communes pour lesquelles le nombre de décès est le plus élevé sont aussi les communes où l'on observe

les niveaux en  $PM_{2.5}$  les plus importants, ce qui n'a rien d'anormal puisque les communes où les concentrations sont les plus fortes sont aussi les plus peuplées (communes de Paris et de la première couronne).

On a utilisé ici un procédé de transparence afin que les points en superposition, très nombreux dans les plus faibles valeurs, soient néanmoins bien visualisables et donnent une idée assez précise de la réalité de la distribution.

### I Figure 14 I

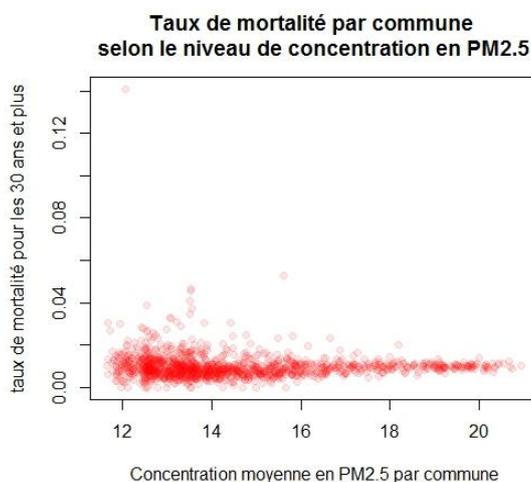
Exemple de nuage de points réalisé dans R



En réalisant le même graphique avec, non plus le nombre de décès par an, mais le taux de mortalité (le nombre de décès des 30 ans et plus, rapporté à la population), on obtient le graphique suivant (figure 15).

### I Figure 15 I

Nuage de points sur la variable précédente convertie en taux



## Distribution géographique d'une variable

La représentation cartographique d'une variable permet d'en visualiser la distribution spatiale et de faire apparaître la diversité d'un territoire donné, à une échelle donnée : ainsi si l'on dispose de l'incidence d'une maladie par département sur le territoire métropolitain, la représentation cartographique de celle-ci va apporter une information visuelle sur ses variations. Ce mode de représentation d'une (ou plusieurs) variable(s) à une échelle donnée sur un territoire donné s'appelle la cartographie thématique, qui diffère de la cartographie descriptive laquelle vise à présenter un territoire de la façon la plus exhaustive en fonction d'un thème donné et en tenant compte d'une certaine échelle : la cartographie descriptive regroupe, par exemple, les cartes topographiques de l'IGN (Institut national de l'information géographique et forestière) que l'on utilise pour organiser une randonnée, ou encore les cartes routières.

La réalisation d'une carte thématique suppose de disposer d'une information localisée (îlots regroupés pour l'information statistique (Iris), commune, canton, département, région, etc.). Elle implique également de connaître les règles de cartographie élémentaire qui la régissent, la première d'entre elles étant qu'on ne représente jamais par un aplat coloré un effectif brut sur un découpage géographique dont les unités ne sont pas homogènes en termes de taille<sup>4</sup>.

Enfin, elle suppose généralement de mettre en œuvre un processus de discrétisation de la variable que l'on souhaite représenter, ce processus ayant son importance car il pèse assez lourd dans le résultat visuel de la représentation et donc dans l'interprétation qui pourra être faite de la lecture de la carte.

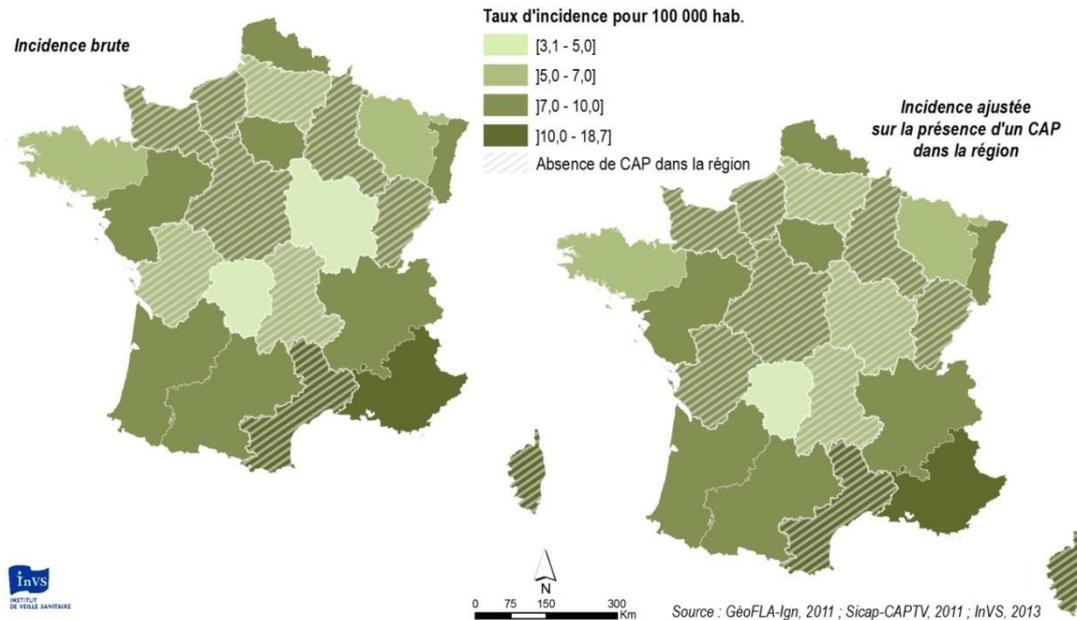
La carte ci-après (figure 16) présentée en exemple est une carte d'incidence des expositions à l'eau de javel répertoriées dans les bases de données des Centres antipoison (CAP) : la carte de gauche représente l'incidence observée alors que celle de droite, représente l'incidence prédite si l'ensemble des régions disposait d'un CAP.

---

<sup>4</sup> Le fait de représenter un effectif brut par un aplat dégradé sur une unité géographique de taille variable entraîne une distorsion du message du fait de ce qu'on appelle l'effet de taille. Il est effectivement aisé de comprendre en prenant l'exemple de la population représentée à l'échelle du département : les départements de Paris ou de la petite couronne sont très peu étendus en superficie pour autant, ils sont très peuplés. Si l'on représente leur population par un aplat dégradé, celui-ci pourra certes être dans une nuance foncée, mais sur une très petite surface : l'effet visuel de l'information concernant cette population ne sera pas à la hauteur du poids réel de la population de ces départements. En utilisant des cercles (par exemple) de taille proportionnelle, cet effet de taille est gommé et les informations plus comparables visuellement.

## Représentation cartographique de la distribution géographique d'une variable

Incidence par région des expositions à l'eau de Javel en 2011



Pour plus d'informations concernant la cartographie ainsi que les règles de sémiologie et de représentation qui la régissent, il est recommandé de se reporter à :

- la « Charte cartographique » de l'InVS [9] ;
- l'ensemble des modèles de cartes et modes opératoires qui sont proposés aux utilisateurs de l'information géographique pour la réalisation de cartes thématiques simples<sup>5</sup>.

### Distribution géographique de deux variables

La cartographie thématique permet de combiner la représentation de la distribution géographique de deux variables ou plus ; cette représentation combinée est notamment recommandée lorsque les différentes variables sont liées par le sens. La limite de cette représentation est conditionnée par la lisibilité de la carte qui en résulte. Généralement, il est relativement aisé de représenter simultanément deux informations qui s'éclairent l'une l'autre, mais sont de nature et de type suffisamment différents pour que la superposition ne rende pas la lecture impossible. Au-delà de deux variables, dans tous les cas, une vigilance particulière devra être portée à la lisibilité.

Voici ci-après (figure 17) un exemple de cartographie thématique associant la représentation de deux variables liées par le sens (un nombre de cas d'une part, représentée par des cercles de taille

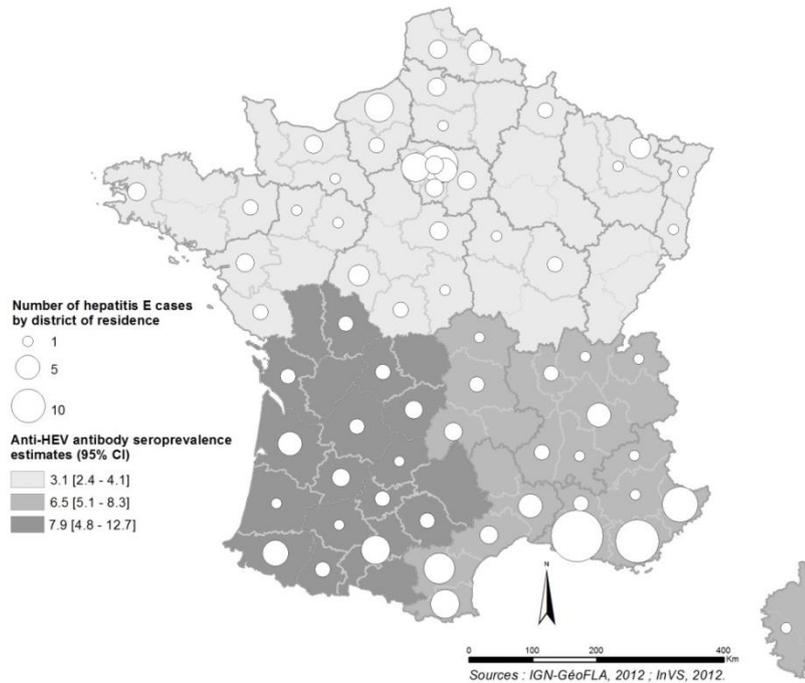
<sup>5</sup> Documents disponibles sur l'espace collaboratif relatif aux Systèmes d'information géographique (SIG) : [http://intranet\\_vosespaces/dse/SIG/Documents%20partages/Forms/AllItems.aspx?RootFolder=%2Fde%2FSIG%2FDocuments%20partages%2FDocuments%20m%C3%A9thodologiques&FolderCTID=0x012000E4F09009A87E4345A0036E22BF09650D&View={3A95747A-64AD-4E53-B05A-BB5ED9B8FB72}](http://intranet_vosespaces/dse/SIG/Documents%20partages/Forms/AllItems.aspx?RootFolder=%2Fde%2FSIG%2FDocuments%20partages%2FDocuments%20m%C3%A9thodologiques&FolderCTID=0x012000E4F09009A87E4345A0036E22BF09650D&View={3A95747A-64AD-4E53-B05A-BB5ED9B8FB72}) - (Saint-Maurice)

<http://vosespaces.invs.sante.fr/dse/SIG/Documents%20partages/Forms/AllItems.aspx?RootFolder=%2Fde%2FSIG%2FDocuments%20partages%2FDocuments%20m%C3%A9thodologiques&FolderCTID=0x012000E4F09009A87E4345A0036E22BF09650D&View={3A95747A-64AD-4E53-B05A-BB5ED9B8FB72}> - (Cires et personnels délocalisés)

proportionnelle aux effectifs, et, d'autre part, une estimation de séroprévalence, pour la même pathologie).

I Figure 17 I

Représentation combinée de la distribution géographique de deux variables



La représentation combinée d'un taux et d'un effectif ne pose *a priori* pas de difficulté en termes de faisabilité et de lisibilité. Par contre il peut être plus délicat de représenter deux ou plusieurs taux sur une même carte. Dans certains cas, la réalisation d'une « collection » (série) de cartes sera plus indiquée. Là encore, pour plus d'information, il sera utile de se reporter à la charte cartographique.

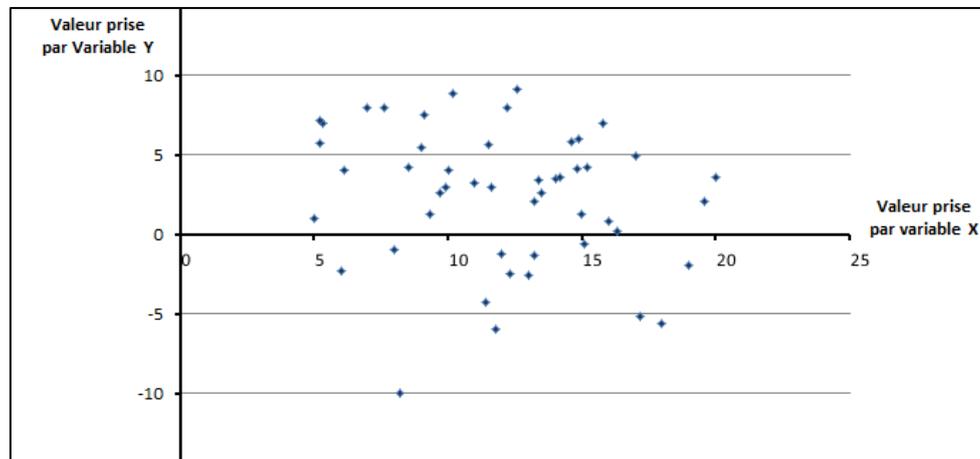
### 2.1.2 Représenter une relation entre des variables

Ce sont les nuages de points qui sont généralement utilisés pour représenter la relation entre deux variables, comme cela a pu déjà être mentionné pour l'exemple de la représentation de la distribution de deux variables. Ils permettent de déceler des valeurs aberrantes, de visualiser d'éventuelles corrélations, ou d'isoler des grappes d'observations.

## Nuage de points (deux variables)

I Figure 18 I

Le nuage de points pour illustrer la relation entre deux variables, réalisé dans Excel

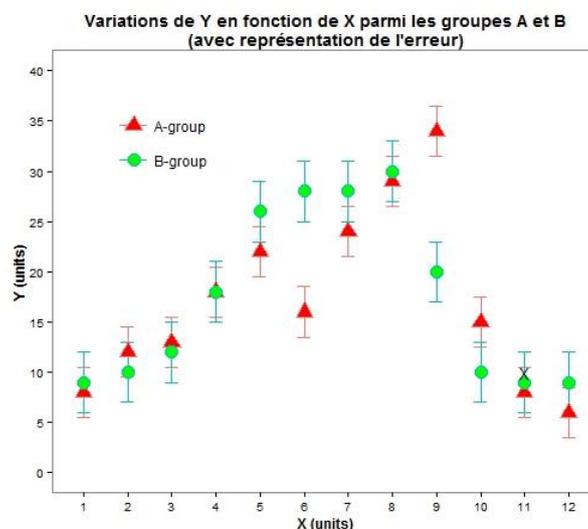


Pour représenter la relation entre deux variables pour deux sous-groupes d'une même population, on utilisera également un graphique en « nuage de points », en ayant recours à des éléments graphiques qui permettront au lecteur de distinguer facilement les deux sous-populations étudiées.

L'exemple ci-dessous, réalisé dans R, représente les variations d'une variable Y en fonction des valeurs de X pour deux groupes d'individus A et B d'une même population ; une « erreur », qui faisait partie des données d'entrée, a également été ajoutée au graphique, ce sont donc trois variables que l'on représente sur un même graphique, sans difficulté majeure de lisibilité. Ce graphique a été réalisé dans R.

I Figure 19 I

Représentation par un nuage de points des variations de X et de Y dans deux sous-groupes distincts A et B, réalisée dans R

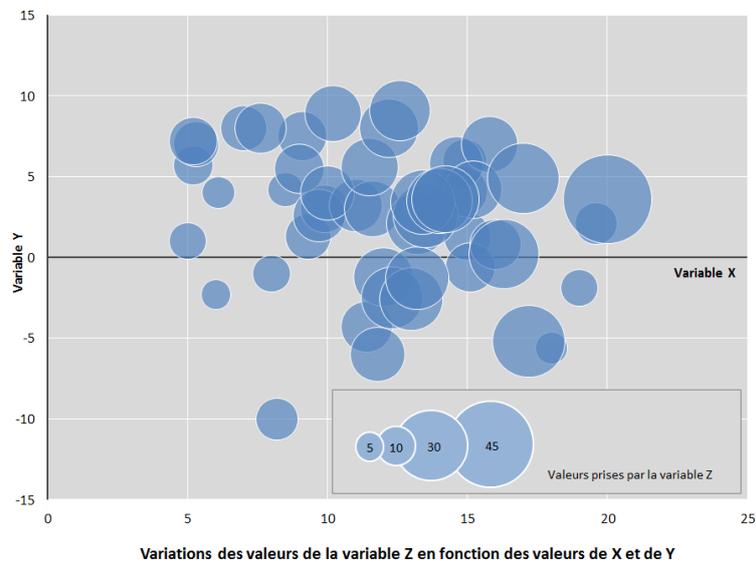


## Bubble chart (graphique en bulles)

Il est utilisé pour illustrer les co-variations de trois variables : la variation de la surface des cercles représente les variations de la troisième variable (figure 20).

### I Figure 20 I

Représentation conjointe des variations de trois variables X, Y et Z sur un graphique en bulles, réalisée dans Excel

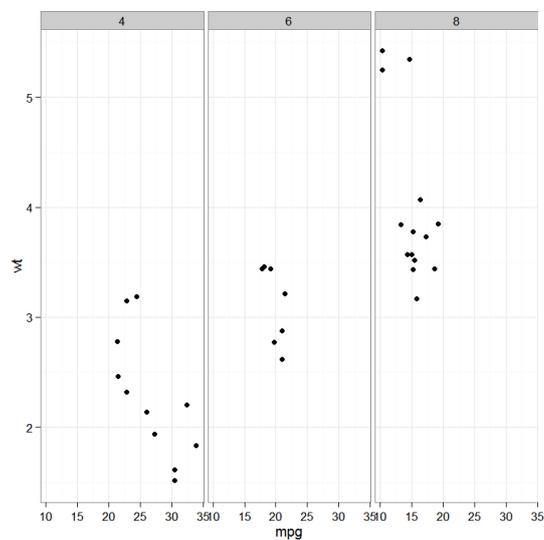


Une autre option peut consister à représenter un nuage de points par niveau d'une autre variable.

Ci-dessous un graphique montrant pour chacun des 3 niveaux d'une variable (ici : 4, 6 et 8), le nuage de points entre les deux autres (figure 21).

### I Figure 21 I

Nuages de points représentant les co-variations de trois variables, réalisés dans R

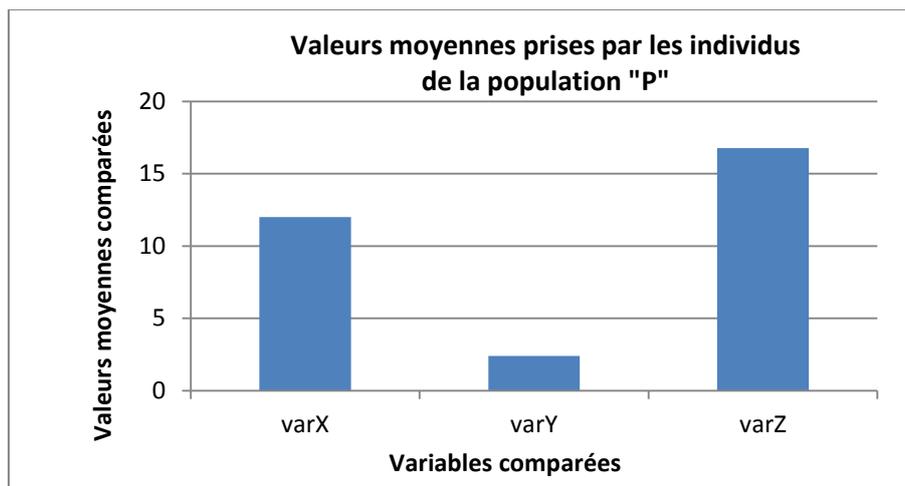


## 2.1.3 Effectuer une comparaison

On effectue une comparaison lorsque l'on considère plusieurs variables comparables sur une même population. Ce sont les graphiques en barres qui sont le plus couramment utilisés pour cet objectif. Pour que des variables soient comparables il est nécessaire, *a minima*, que leurs échelles soient les mêmes (on ne placera pas sur une même aire graphique des représentations d'échelles différentes, la comparaison serait alors fautive) et qu'il y ait un sens à les mettre en regard les unes des autres.

I Figure 22 I

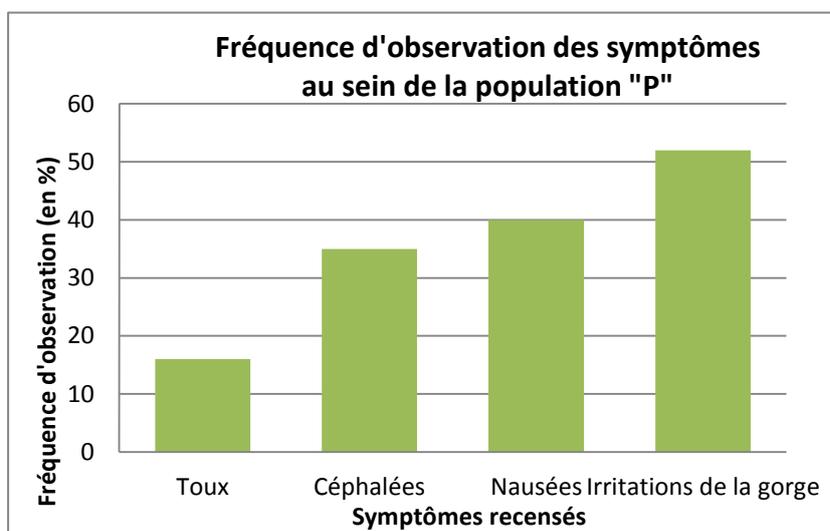
Représentation d'une comparaison à l'aide d'un diagramme en barres, réalisée dans Excel



On peut utiliser ce type de graphique pour comparer, par exemple, les fréquences d'observation de différents symptômes dans une même population (figure 23).

I Figure 23 I

Représentation des variations des fréquences d'observation de certains symptômes au sein d'une population P à l'aide d'un diagramme en barres

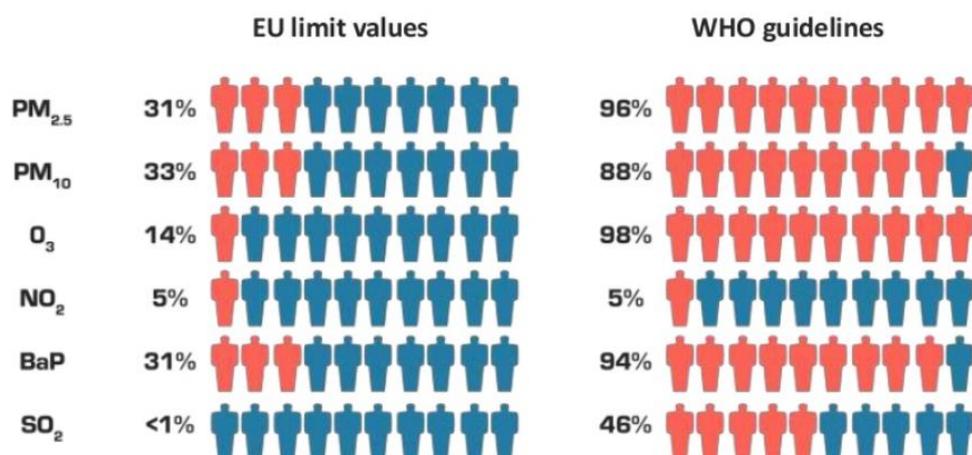


L'Agence européenne de l'environnement (AEE) utilise un mode de représentation analogue pour la visualisation rapide de la part de la population exposée à des niveaux élevés de polluants présents dans l'air et dont voici un extrait<sup>6</sup> :

I Figure 24 I

Infographie reprenant le principe du diagramme en barres pour représenter la part de la population urbaine exposée à des niveaux de pollution nocifs, selon les substances, et d'après les limites réglementaires européennes d'une part, et les valeurs guides de l'Organisation mondiale de la santé (OMS) d'autre part

EU urban population exposed to harmful levels of air pollution, according to:



Sur ce graphique les personnages colorés en rouge représentent le pourcentage de population exposée dans la population totale (figurée par la somme des personnages rouges et bleus, les bleus représentant donc la part non exposée). Ce type de graphique présente l'avantage d'être aisément lisible et compréhensible, le caractère frappant du message porté par le graphique atteint pleinement son objectif. L'utilisation des personnages place ce graphique à la limite de l'infographie. Mais la philosophie est analogue à celle du diagramme en barres.

Les graphiques « en barres » ne doivent pas être confondus avec les histogrammes<sup>7</sup>. Une inversion des axes peut être envisagée pour améliorer la lisibilité de la légende des axes, par exemple.

## 2.1.4 Représenter une décomposition

Représenter la décomposition, c'est figurer le poids des différentes modalités d'une même variable au sein d'une population. En d'autres termes, la population d'étude est décomposée en différentes sous-populations induites par les différentes modalités d'une même variable (décomposition par sexe, par catégorie socioprofessionnelle (CSP), par exemple).

<sup>6</sup> Source : <http://www.slideshare.net/fullscreen/EUEnvironmentAgency/air-quality-report-2013/12>

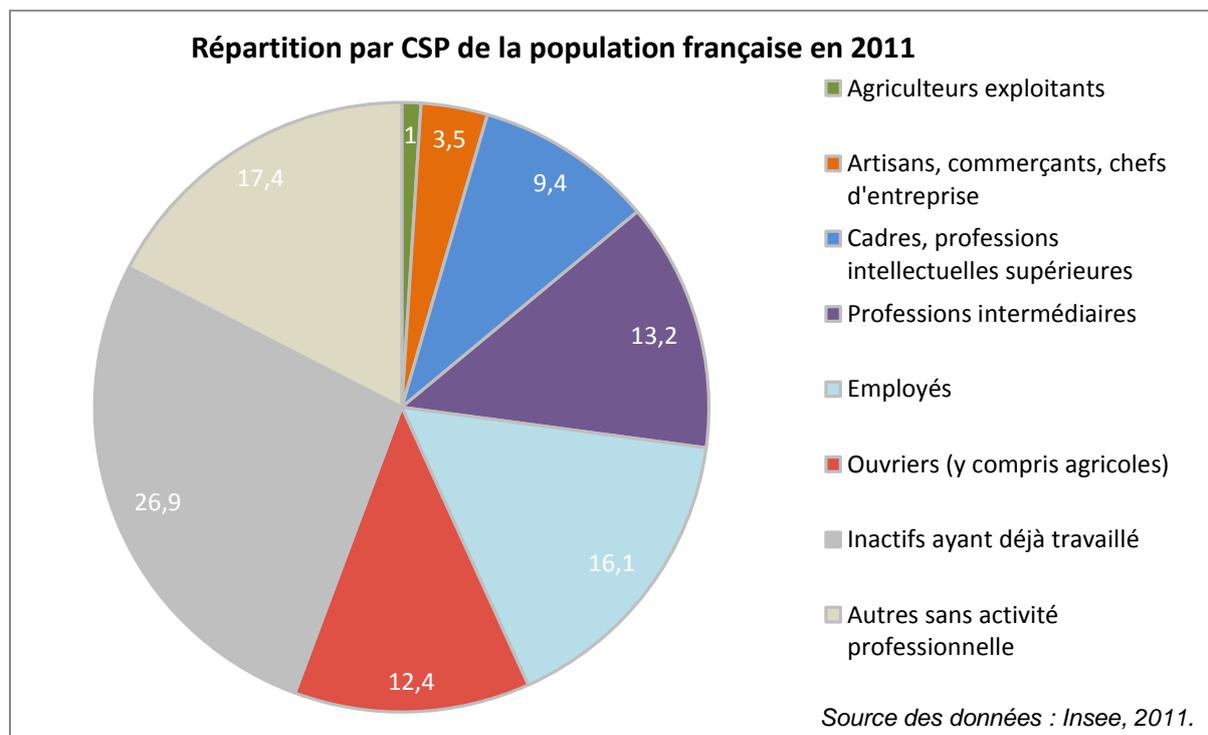
<sup>7</sup> Cf. 2.1.1 Représenter la distribution statistique d'une variable ou plusieurs variables

## Diagramme circulaire

Ce sont les diagrammes circulaires (les habituels « camemberts ») qui sont le plus couramment utilisés pour représenter ce type d'information.

### I Figure 25 I

Représentation d'une décomposition par un diagramme circulaire



Pourtant, ce choix de représentation n'est, et de loin, pas le plus efficace, bien que le plus fréquemment employé, et il implique de prendre quelques précautions : il n'est vraiment lisible qu'autour de 3 à 6-7 classes/modalités différentes. Au-delà, le morcellement du « camembert » risque de conduire à une lisibilité plus faible et ce, d'autant plus que l'on a des classes ou des modalités qui regroupent une très faible part de la population totale que l'on étudie (un angle inférieur à 30° est souvent mal estimé par l'œil humain). Par ailleurs, nombre d'outils proposent souvent, pour ce type de graphiques, des visualisations « 3D » qui, en recherchant plus ou moins une visée esthétique, rendent en fait les graphiques illisibles et faux puisque la surface représentée pour chacune des classes n'est plus proportionnelle à la part de chacune des classes dans la population totale étudiée.

D'autres modes de représentation graphique méritent qu'on s'y attarde car leur lisibilité est souvent bien meilleure.

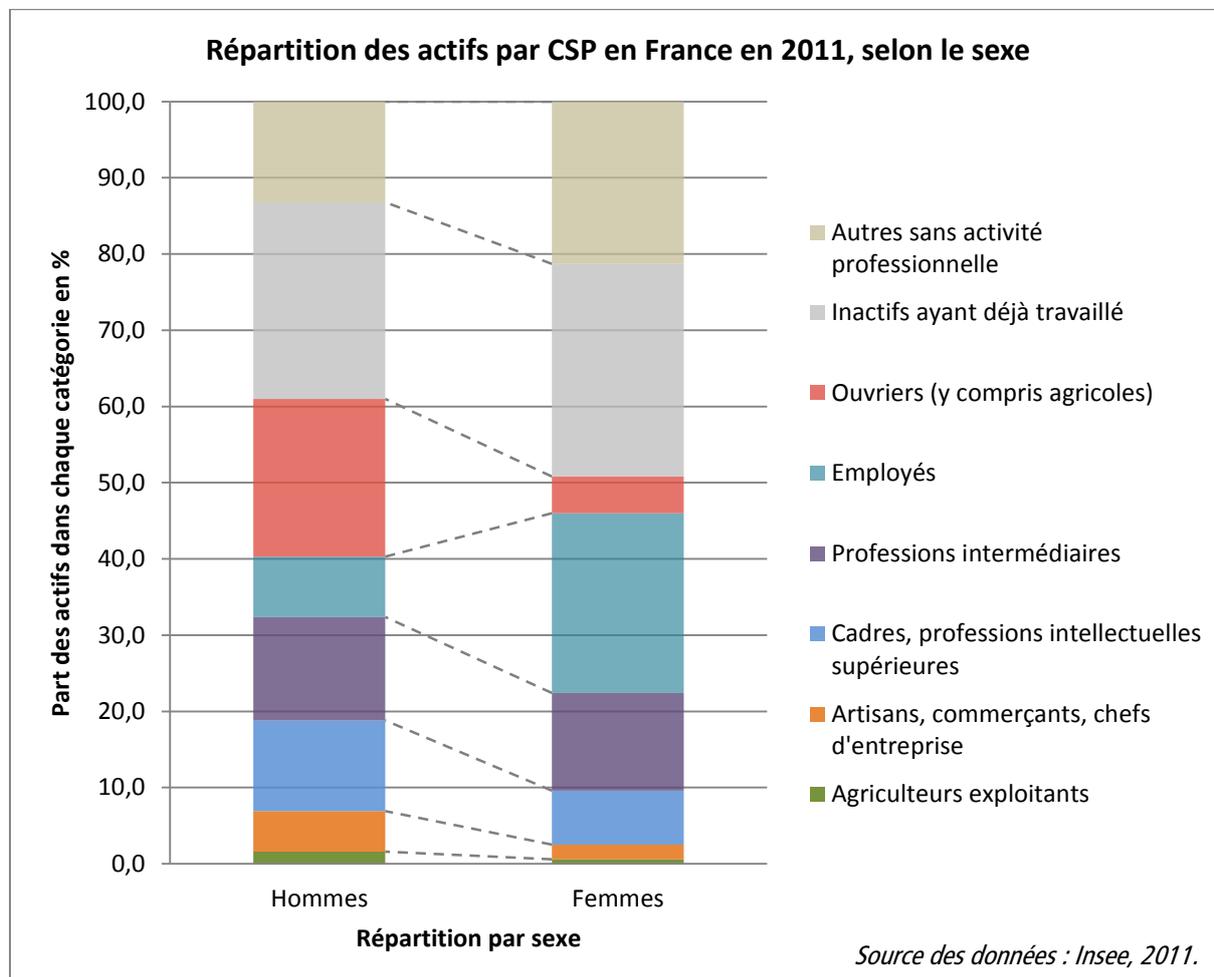
## Barres empilées

Les barres empilées, en effectifs absolus ou en base 100 (*i.e.* les effectifs bruts sont convertis en pourcentages) selon les objectifs de la représentation, permettent une meilleure visualisation puisque la perception de la hauteur de barre est plus immédiate pour l'œil que l'interprétation d'un secteur angulaire. Ici, l'exemple proposé est un exemple en base 100 qui compare la répartition en CSP de la population active, déployée en deux sous-ensembles selon le sexe. Le choix ici a été de représenter cette répartition en base 100, afin de comparer spécifiquement les poids de chaque CSP

pour chacune des deux sous-populations considérées. Ce graphique combine donc deux objectifs : représenter à la fois une décomposition (répartition par CSP) et effectuer une comparaison (entre la répartition en CSP des actifs féminins et masculins).

I Figure 26 I

Représentation d'une comparaison par barres empilées

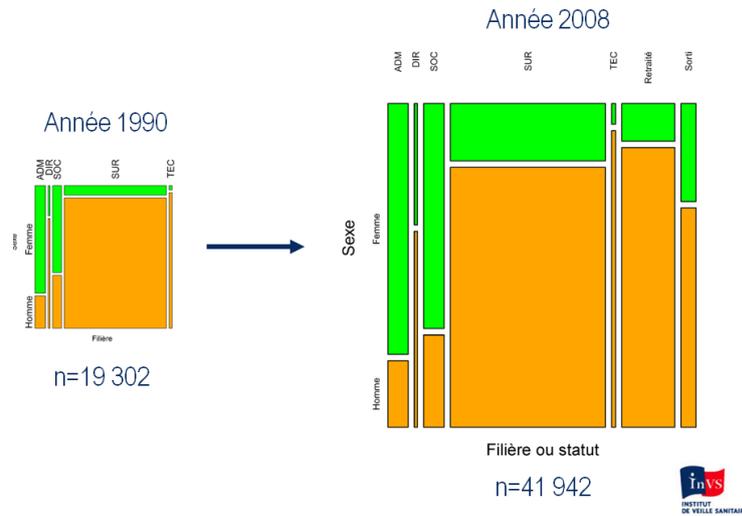


### Mosaic plot

L'exemple présenté ci-après est extrait d'une présentation réalisée par le Département santé travail (DST) en 2013, portant sur la description de la mortalité des agents et ex-agents de l'Administration pénitentiaire (AP) en France. L'InVS a été questionné à propos de l'espérance de vie réduite de ces agents de l'AP. Les agents de cette administration ont été suivis dans le cadre d'une cohorte historique entre 1990 et 2008. Les deux graphiques présentés ci-dessus illustrent la répartition des agents, par sexe, et par secteur d'activité au sein de l'administration (le plus important secteur étant la surveillance des détenus, « SUR », qui représentent trois quarts des effectifs). L'apparition de retraités et de « sortis » (personnes ayant quitté l'AP) en 2008 s'explique par le vieillissement d'une part et les mouvements professionnels d'autre part (cessation, radiation de cadres, licenciements, démissions, etc.) des agents dont le suivi avait démarré en 1990. La surface des rectangles est proportionnelle aux effectifs de chacune des catégories représentées et aux effectifs totaux, ce qui rend possible la comparaison entre les deux dates de début et de fin de suivi de la cohorte. Cette forme de représentation d'une décomposition d'une population en plusieurs sous-ensembles est particulièrement riche. Ce graphique a été réalisé dans R (annexe 1).

## I Figure 27 I

### Exemple de graphique en mosaïc Plot

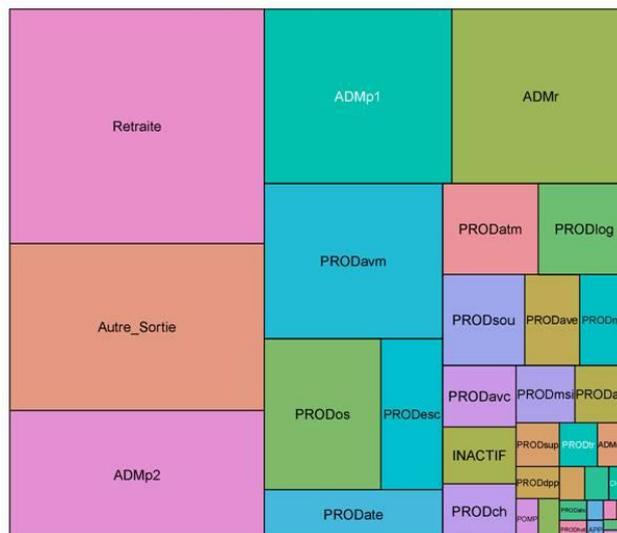


### Tree map

Dans un esprit assez proche de celui des mosaic plot, il est possible de réaliser avec R des graphiques en tree map. Dans ce graphique, l'objectif est de représenter le poids respectif des différents sous-ensembles identifiés dans une population d'intérêt donné. La surface de chaque rectangle est donc proportionnelle au poids de la catégorie qu'il illustre dans la population totale, qui elle, est figurée par la totalité du rectangle. Dans l'exemple proposé, issu également des travaux du DST, chaque rectangle représente un métier de l'AP, et pour chacun la surface est proportionnelle au nombre de personnes-années.

## I Figure 28 I

### Graphique en tree map réalisé dans R



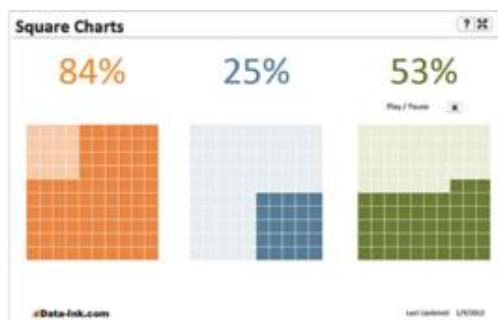
Une information complémentaire peut être apportée par le choix d'un code couleur, ce qui n'est cependant pas le cas dans l'exemple proposé ici (annexe 2).

### Camemberts carrés

Le camembert carré, proposé dans l'ouvrage de B. Lebellet [3] comme une alternative très avantageuse au diagramme circulaire classique est en quelque sorte une version un peu simplifiée des deux exemples de graphiques présentés précédemment. Un peu plus simple et adapté à des décompositions moins fines, il se rapproche cependant beaucoup du tree map présenté précédemment. Dans ce mode de représentation, il n'y a pas de découpage en secteurs, ce n'est donc pas à l'estimation d'un angle que le lecteur doit associer une proportion (B. Lebellet propose par ailleurs une méthodologie pour recréer ce type de graphique en « camemberts carrés » dans Excel<sup>8</sup>).

#### I Figure 29 I

Exemple de camemberts carrés<sup>9</sup>



### 2.1.5 Représenter une évolution

Dans le schéma d'Abéla (figure 2), la représentation d'une évolution est regroupée avec celle de la comparaison ou encore de la décomposition. Cependant, l'importance du facteur temps dans nombre des travaux conduits dans le domaine de la santé, qu'il soit un temps court ou un temps long, induit que l'on accorde ici une place spécifique à la représentation des évolutions.

#### Évolution pour un facteur

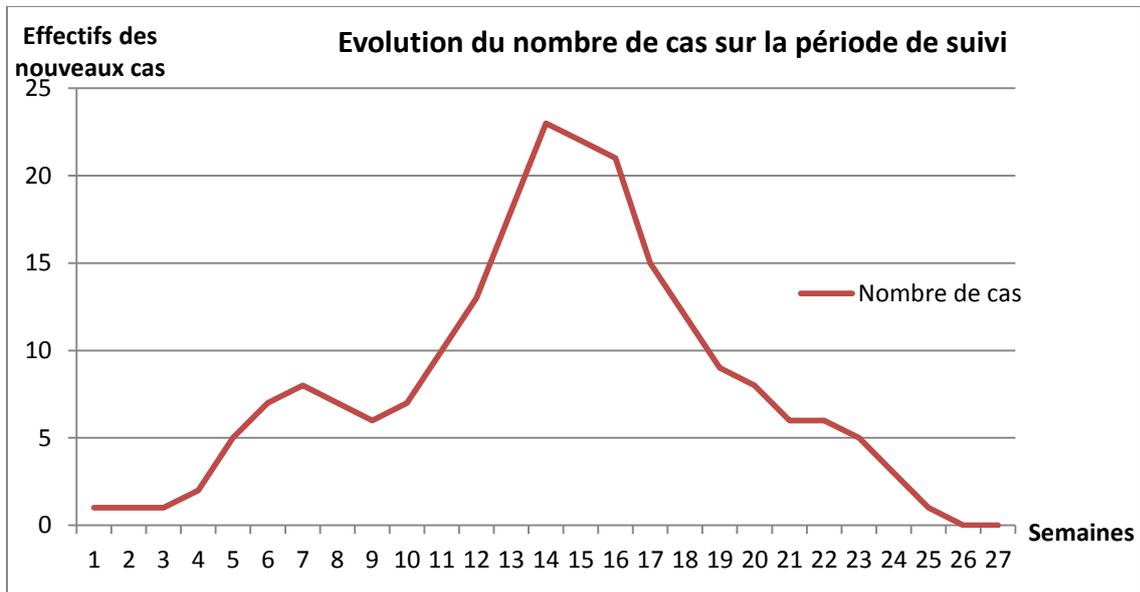
La solution graphique la plus immédiate et la plus simple est la courbe temporelle (figure 30). Ce mode de représentation nécessite d'être vigilant quant à l'affichage de l'axe temporel (bonne lisibilité du pas de temps). Il est particulièrement adapté à la visualisation d'un phénomène continu dans le temps (absence de rupture d'une date à une autre).

<sup>8</sup> <http://www.impactvisuel.net/2010/01/21/creer-un-camembert-carre-sous-excel-square-pie-tutorial.html#comment-18100>

<sup>9</sup> <http://www.datavisualization.fr/blog/hlinks/page/2/>

### I Figure 30 I

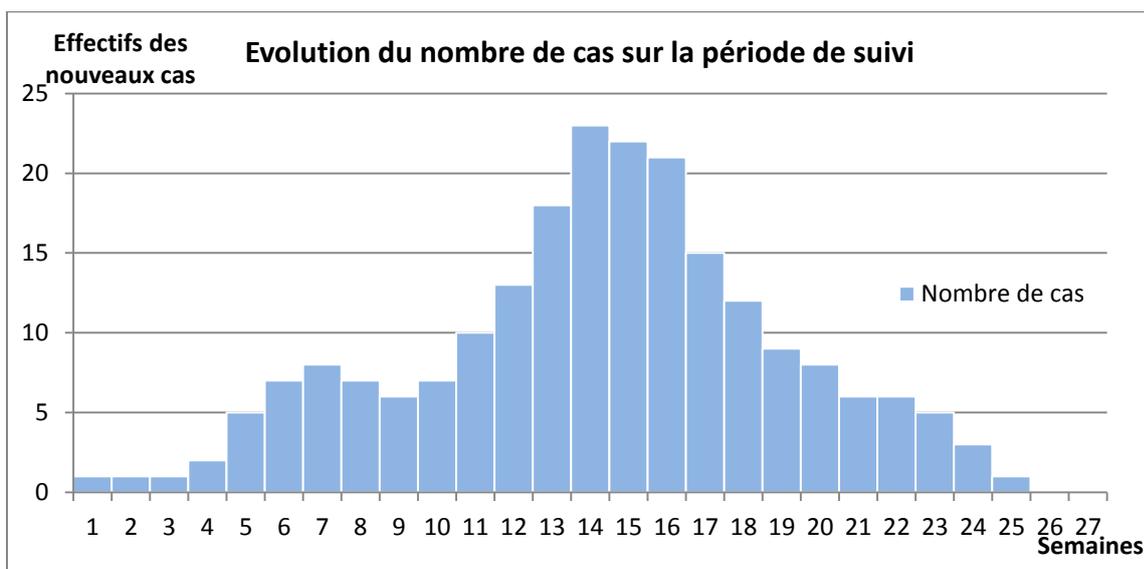
Courbe épidémique réalisée dans Excel



Le graphique en barres peut également être utilisé pour représenter une évolution. On le réservera prioritairement aux phénomènes que l'on étudie en « découpant » le temps d'observation de la variable considérée en périodes bien délimitées (semaines, mois, par exemple). Le graphique en barres ci-dessous (figure 31) représente exactement les mêmes données que la courbe de suivi précédente.

### I Figure 31 I

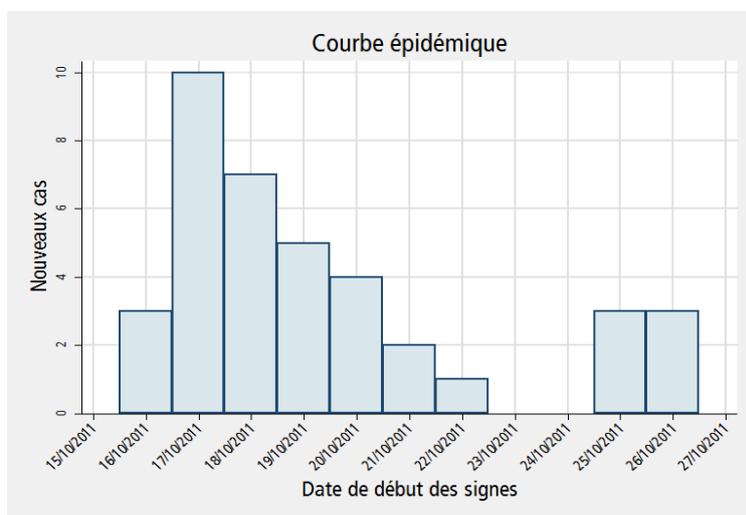
Autre exemple de courbe épidémique réalisé dans Excel



Voici un autre exemple de courbe épidémique, réalisée sous STATA.

I Figure 32 I

Courbe épidémique réalisée sous STATA

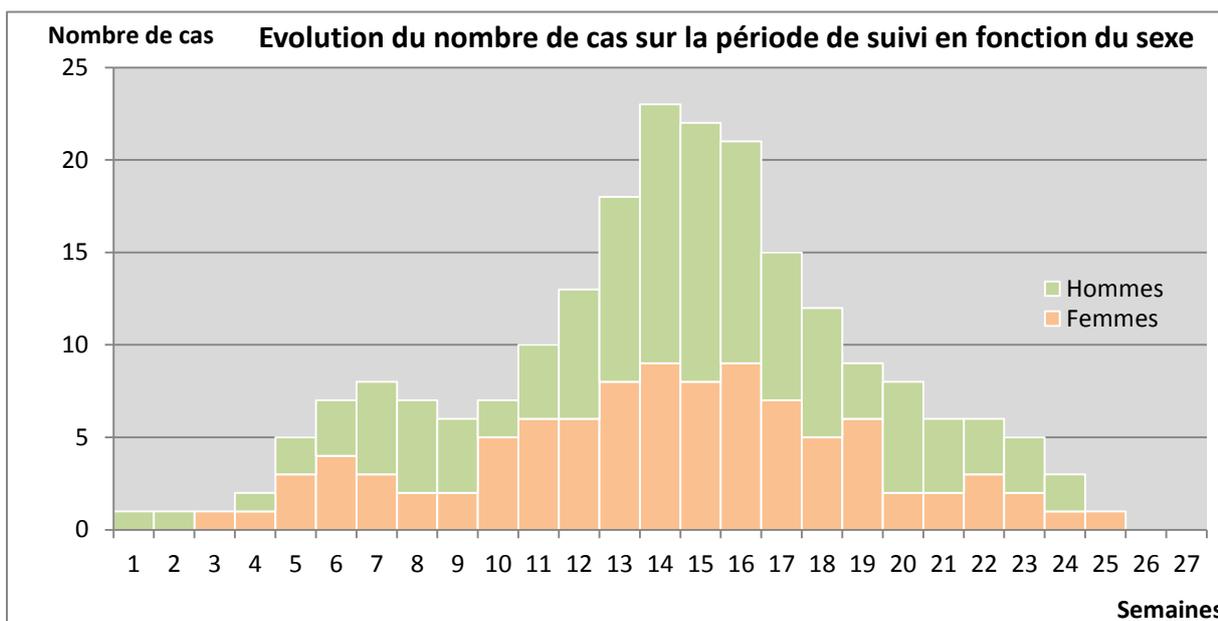


Évolution pour un facteur, d'un effectif décomposé en plusieurs classes

Les barres empilées (figure 33) pourront être utilisées : l'axe des abscisses figurera là aussi le temps, décomposé en périodes bien délimitées (mois, semaines par exemple), et chaque barre aura une hauteur signifiant l'effectif total pour chaque période, et sera décomposée en plusieurs barres (empilées) figurant les variations du poids des différentes sous-populations constitutives de l'effectif total considéré.

I Figure 33 I

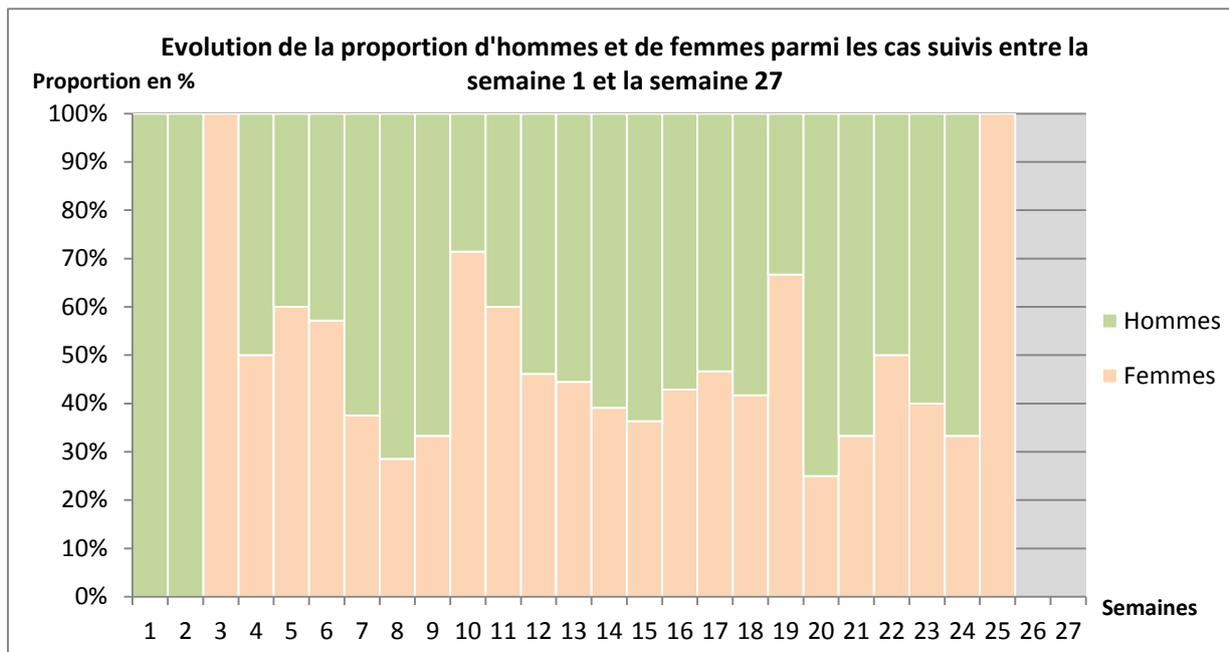
Représentation de l'évolution du nombre de cas d'une épidémie et de son sexe-ratio par un diagramme en barres en conservant les effectifs



Deux variantes possibles : les différences absolues sont importantes, et alors la hauteur totale des barres doit être représentée sur le graphique ; ou alors c'est la différence relative que l'on veut montrer, et alors les barres auront toutes la même hauteur (passage en base 100, figure 34).

### I Figure 34 I

Représentation de l'évolution des parts respectives d'hommes et de femmes parmi les cas recensés chaque semaine sur la période de suivi sans tenir compte des effectifs totaux de cas

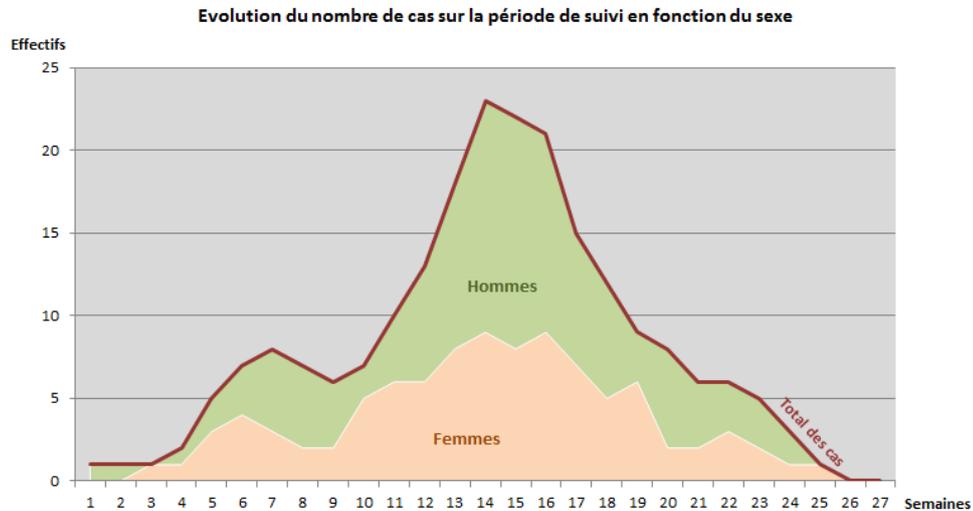


Ces modes de représentation en barres empilées sont adaptés à une période relativement courte, ou à un nombre de sous-périodes découpant la période d'étude relativement limitée. Pour des périodes plus longues, des diagrammes en aires empilées (figure 35) seront plus pertinents.

Il sera alors possible de faire le choix, en fonction de ses objectifs, entre un diagramme en aires empilées où les effectifs absolus sont conservés, et un diagramme en aires empilées avec un passage en base 100, où seules sont lisibles les variations relatives des différentes catégories considérées.

## I Figure 35 I

Représentation par un graphique en aires empilées de l'évolution du nombre de cas d'une épidémie et de son sexe ratio

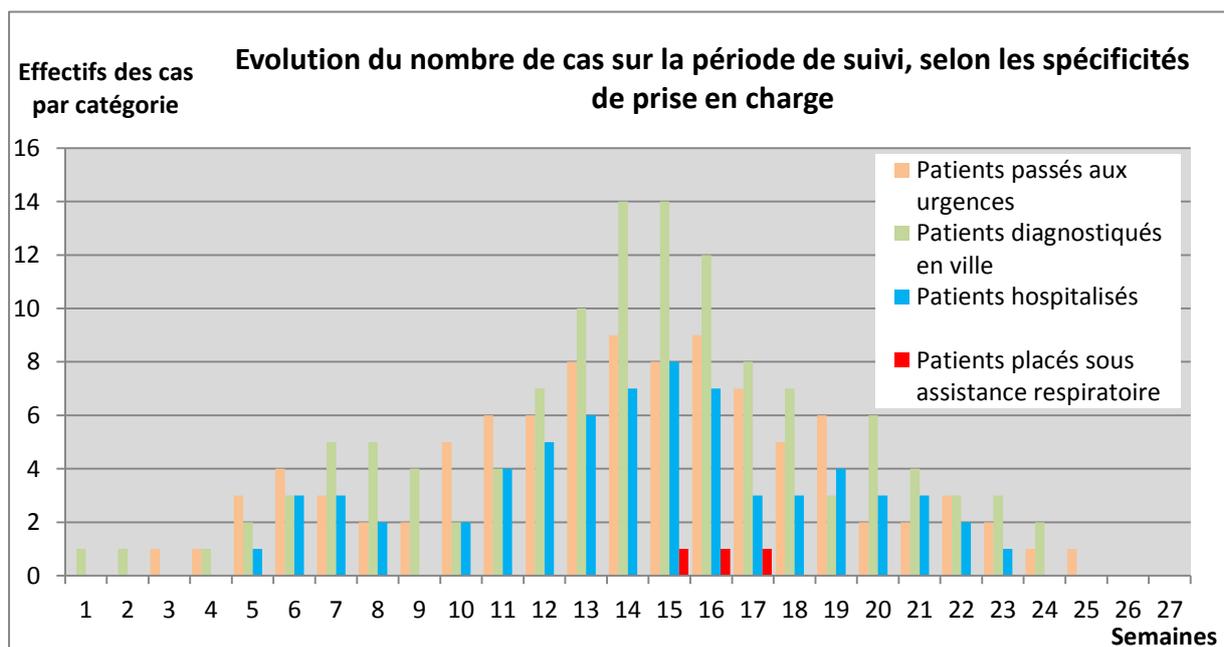


### Évolutions comparatives

Sur un petit nombre de facteurs considérés, le diagramme en barres semble bien adapté. Les barres, différenciées par exemple par une couleur différente, sont juxtaposées et représentent les différentes variables étudiées. Au-delà de quatre variables, la lisibilité devient plus complexe. Avec l'exemple ci-dessous (figure 36), sur une période de durée moyenne mais avec la représentation de quatre catégories on voit bien déjà que la lisibilité n'est pas excellente.

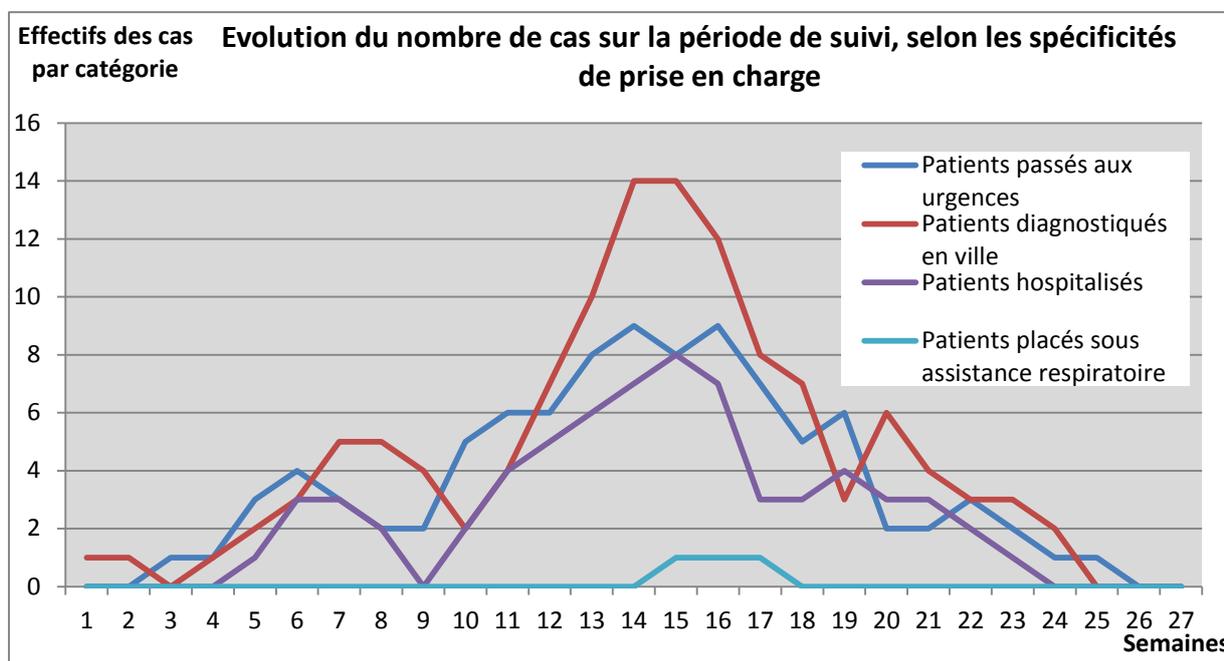
## I Figure 36 I

Représentation par un diagramme en barres de l'évolution d'une épidémie en fonction d'une autre variable



I Figure 37 I

Représentation de l'évolution d'une épidémie selon une autre variable par des courbes



Sur un temps plus long, il deviendrait plus pertinent de figurer les évolutions, pour chaque facteur, par une courbe différente. Les évolutions peuvent être comparées en plaçant ces différentes courbes dans un même repère. Cependant à l'usage, la représentation simultanée de très nombreuses courbes sur un même graphique entraîne toujours des difficultés de lecture. L'attention doit alors être portée sur les limites en termes de lisibilité de la co-représentation de trop nombreux facteurs d'intérêt (figure 37).

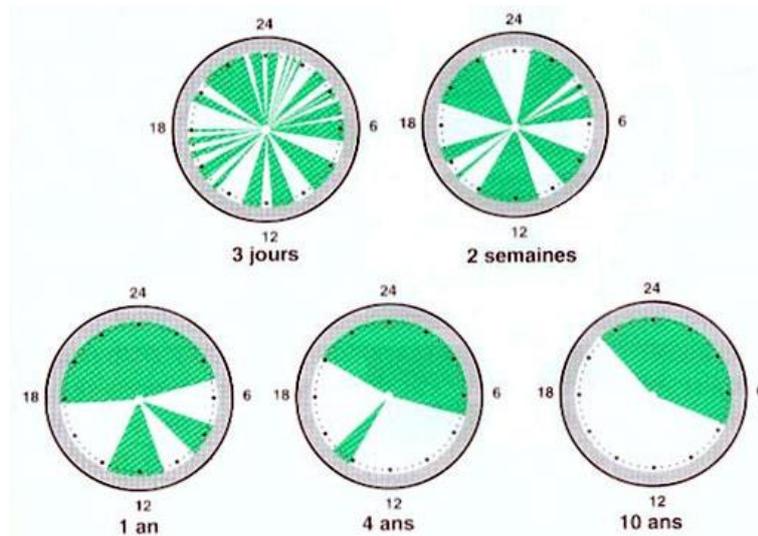
Sans mise en exergue d'une information par rapport à une autre, ce type de graphique est assez vite peu lisible : les lignes s'entrecroisent, se chevauchent, se dissimulent éventuellement entre elles, aucune ne se détache par rapport à l'autre. Pour peu que le choix des couleurs n'ait pas fait l'objet d'une quelconque attention (comme c'est le cas, à dessein, ici), la lisibilité de ce type de graphique est assez médiocre, et le temps qu'il faudra pour l'interpréter risque de rebuter le lecteur qui ne le lira peut-être pas. On étudiera un peu plus loin comment des techniques simples peuvent être mises en œuvre afin de hiérarchiser et classer et éventuellement regrouper l'information dans un graphique, afin d'en simplifier la lecture.

#### Cas particulier des données cycliques

Le facteur temps, dans le cas de certaines analyses, se révèle parfois plus intéressant à considérer en termes cycliques qu'en termes linéaires. En effet, dans certains phénomènes où la saisonnalité est particulièrement marquée par exemple, la cyclicité du temps trouve tout son sens. Il existe des modes de représentation spécifiques à la prise en compte du temps dans sa dimension cyclique. C'est le cas par exemple des diagrammes de forme circulaire : diagramme de cycles horaires par exemple, ou graphiques en « quadrants ».

## I Figure 38 I

Représentation en cadrans des alternances veille/sommeil en fonction de l'âge chez les enfants de 3 jours à 10 ans



Ce graphique (figure 38) est extrait d'une publication sur les troubles du sommeil disponible sur le site de la faculté de médecine de Marseille<sup>10</sup>.

Il illustre l'évolution normale juste après la naissance, et dans les premières années de vie, des alternances veille/sommeil par cycle de 24h. Les phases de sommeil sont figurées en vert, et les phases de veille, en blanc. La comparaison entre les différents cadrans permet de bien s'approprier l'évolution au fil des années.

Dans d'autres contextes, la représentation de la dimension cyclique du temps pourra également être figurée par une spirale.

### Rose de Nightingale

Il faut noter le caractère particulièrement innovant pour l'époque de cette représentation graphique (figure 39) proposée par l'infirmière Florence Nightingale sur les causes des décès pendant la guerre de Crimée (1853-1856). Son objectif était de montrer le poids relatif très élevé des décès des soldats pour cause d'infections, éventuellement évitables, par rapport aux décès liés directement aux suites de blessures reçues au combat<sup>11</sup>.

Une conversion en diagramme en barre du type de celui réalisé en figure 36 (évolution du nombre de cas sur la période de suivi selon les spécificités de prise en charge) aurait très bien pu convenir à la représentation des informations présentées.

Mais la dimension cyclique du temps et la mise en regard des mêmes mois d'une année à l'autre n'auraient pas forcément été aussi visuellement parlantes que dans la représentation qui a été finalement réalisée.

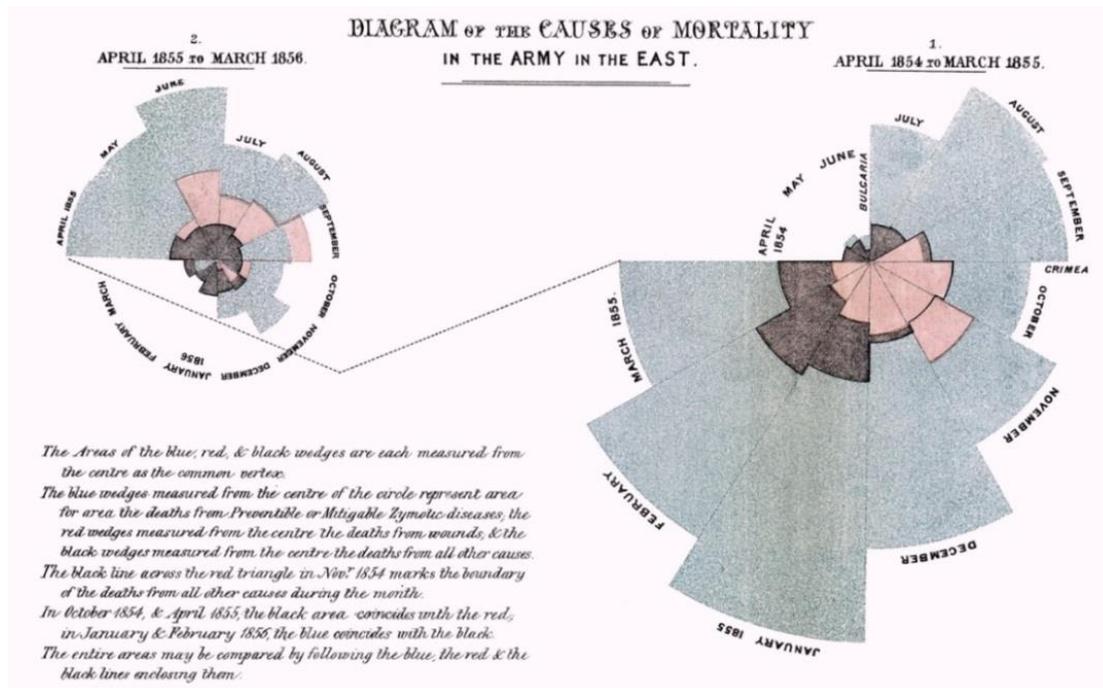
<sup>10</sup> Voir :

<http://medidacte.timone.univ-mrs.fr/webcours/Comite-etudes/ItemsENC/sitelocal/disciplines/niveaudiscipline/niveaumodule/Item43/Item43.htm>

<sup>11</sup> Ce graphique existe aussi sous forme d'animation : <http://understandinguncertainty.org/nightingale>

I Figure 39 I

Les roses de Nightingale : représentation des causes de mortalité dans les armées du front de l'Est pendant la guerre de Crimée



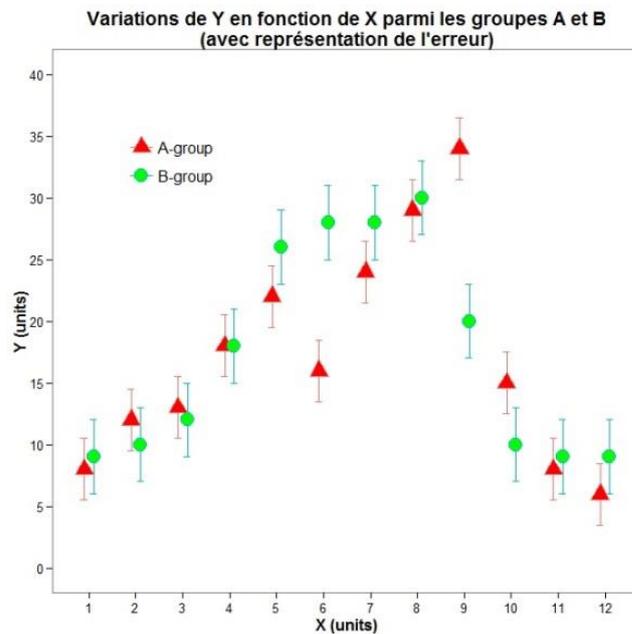
(Source : Wikipédia - [https://en.wikipedia.org/?title=Florence\\_Nightingale#/media/File:Nightingale-mortality.jpg](https://en.wikipedia.org/?title=Florence_Nightingale#/media/File:Nightingale-mortality.jpg))

## 2.1.6 Représenter l'incertitude

L'incertitude est un élément indissociable des résultats de nombres des analyses statistiques qui accompagnent les études épidémiologiques. Elle trouve son origine dans de très nombreux paramètres : faiblesse des échantillons, qualité des données, erreurs de mesures, valeurs manquantes sont quelques-uns des éléments qui caractérisent l'incertitude qui entoure les résultats de ces travaux. Cette incertitude peut souvent être quantifiée, ce qui en permet une représentation visuelle au même titre que toute autre information quantitative. La mention d'une « erreur », d'un intervalle de confiance, par exemple, sont autant de modalités de représentation de l'incertitude sur un graphique, chaque type de graphique disposant d'un mode plus approprié pour l'ajout de ce type d'information. Le graphique ci-après (figure 40) qui fait figurer une erreur pour chaque valeur considérée des deux variables étudiées est un exemple d'une manière parmi d'autres de représenter cette incertitude.

## I Figure 40 I

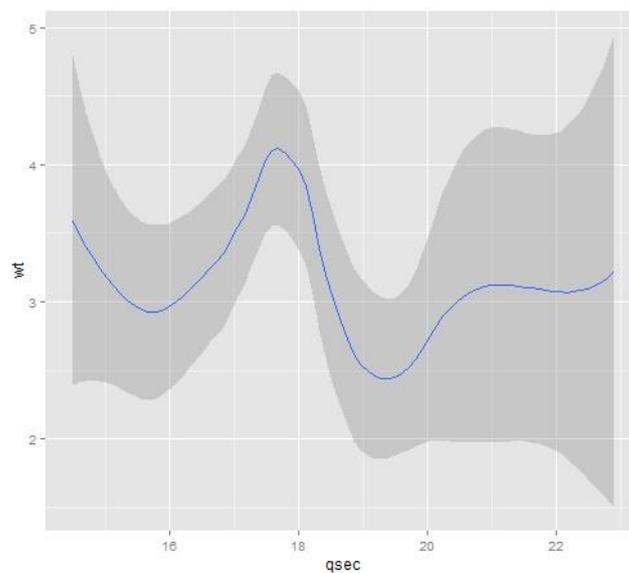
Représentation d'une erreur ou d'une incertitude sur un graphique en nuage de points



Pour un graphique en courbe, il y aura d'autres manières plus appropriées de faire apparaître cette incertitude, comme ici (figure 41), par une aire qui se déploie de part et d'autre de la courbe :

## I Figure 41 I

Représentation de l'incertitude sur un graphique en courbe

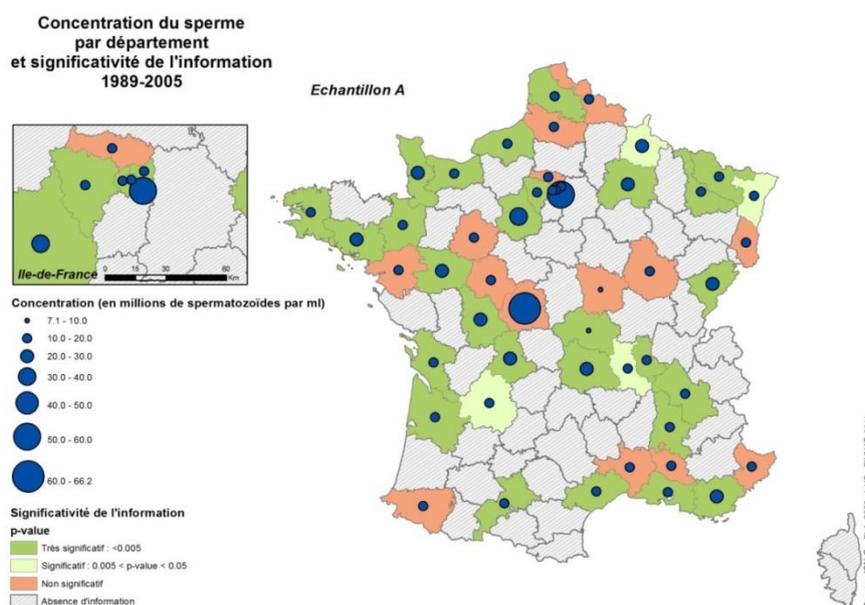


## Représentation cartographique de l'incertitude

La représentation de l'incertitude sur des cartes a fait l'objet de nombreux travaux de recherche. On retiendra quelques possibilités assez simples, comme le floutage, le tramage ou la transparence surajoutée aux informations cartographiées pour isoler les zones caractérisées par des seuils d'incertitude élevés ou par une significativité faible. L'utilisation de ces nouvelles variables visuelles, autorisées par l'évolution des outils mis à la disposition des cartographes, permet de combiner la représentation d'une variable d'intérêt et de l'incertitude associée.

### I Figure 42 I

#### Représentation cartographique de la p-value combinée à l'information d'intérêt



Cependant, il faut garder à l'esprit que de telles modalités visuelles ne sont pas déclinables en autant de niveaux que les trames de couleurs ou de valeurs couramment mobilisées pour la représentation des variables d'intérêt, et là encore, la lisibilité et la rapidité de l'accès à l'information demeurent des paramètres indispensables auxquelles l'auteur se doit de veiller. On peut dire que si une trame de valeur ou de couleur est déclinable en 5 à 7 classes qui peuvent être distinguées sans peine, autant les modalités de floutage, transparence ou tramage surajouté pour représenter l'incertitude ne pourront être déclinés en plus que 2, exceptionnellement 3 niveaux différents.

La significativité d'une information peut également faire l'objet d'une représentation combinée avec l'information d'intérêt, elle-même, donnant ainsi au lecteur de la carte la possibilité de relativiser l'importance et le poids d'un résultat en fonction de sa significativité. La carte précédente donne un exemple de prise en compte de la significativité dans la représentation d'un résultat (figure 42).

## 2.2 Mise en œuvre technique

### 2.2.1 Les composantes du graphique

Si le mode de représentation graphique est grandement conditionné par les objectifs visés, par la nature des données que l'on traite et par le public auquel il est destiné, il demeure cependant, une fois ce choix effectué, une certaine « marge de manœuvre » à l'auteur pour l'adapter, au plus près, au sujet qu'il traite. Les travaux de J. Bertin, géographe, ont joué un rôle précurseur dans l'émergence de la représentation graphique et de la cartographie [2]. En jouant sur les composantes du graphique et sur les éléments d'habillage, l'auteur du graphique a la possibilité d'optimiser l'adaptation de sa représentation à ses objectifs propres et d'améliorer la lisibilité au maximum. Les composantes du graphique peuvent être énumérées et définies comme suit :

#### **La ou les série(s) de données :**

Les données sont évidemment la « matière première » du graphique. Elles en sont la composante fondamentale, et également la seule sur laquelle l'auteur ne peut jouer : l'auteur doit « faire avec » la structure et la qualité des données qu'il souhaite représenter. Ces notions de structure et de qualité des données ne font d'ailleurs pas partie des sujets évoqués ici, même s'il est nécessaire, lors de la réalisation d'un graphique, de garder en permanence en tête les informations qui ont trait à ces questions. En effet, elles se posent naturellement très en amont de la représentation des résultats puisqu'elles interviennent au plus tard au moment de la collecte, mais doivent théoriquement même avoir été déjà abordées lors de l'élaboration du protocole d'étude. Les données sont la matière première incontournable du graphique, mais, éventuellement aussi, une manière de contrainte pour la représentation graphique : une distribution particulière peut par exemple nécessiter un choix de représentation spécifique.

#### **L'aire graphique :**

C'est la « surface » sur laquelle le graphique va prendre place, accompagné parfois, mais sans obligation, de certains éléments complémentaires comme la légende, le titre, la source, par exemple. La taille de cette aire graphique a son importance dans la lisibilité du graphique, elle peut également faire l'objet, dans certaines situations, de contraintes (contraintes de place dans un contexte éditorial par exemple). Mais en l'absence de contraintes fortes l'auteur a la possibilité de jouer sur cette aire graphique pour améliorer la lisibilité : augmentation de taille dans le cas de très nombreuses données par exemple, modification de la couleur du fond, modifications des proportions (rapport hauteur/largeur), etc. En effet par exemple, il est démontré qu'un graphique est plus aisé à analyser pour l'œil lorsqu'il s'inscrit dans un rectangle horizontal que lorsque l'aire graphique prend la forme d'un carré ou d'un rectangle d'orientation verticale.

#### **Le choix de l'échelle ; les axes quantitatifs, qualitatifs, temporels ; les quadrillages intermédiaires ; le coefficient de mensonge :**

Dans le cas de la représentation d'une information quantitative, l'échelle retenue est le plus souvent une échelle linéaire, c'est en effet *a priori* la plus simple à lire et à interpréter. Elle consiste à matérialiser directement et sans transformation l'étendue sur laquelle est mesurée la série de données que l'on étudie. Elle se traduit, dans l'aire graphique, par la figuration d'un ou plusieurs axes. Dans certains cas ou dans certaines configurations de distribution, une transformation mathématique des données (transformation logarithmique, par exemple ou encore choix d'une échelle de puissance) peut permettre une représentation plus lisible des valeurs prises par la série dans le graphique. Cette transformation est alors appliquée sur l'ensemble des valeurs de la série. Il est important d'avoir en tête qu'en fonction du choix de l'échelle, le résultat visuel sera complètement différent. Par ailleurs, dans le cas des échelles impliquant une transformation mathématique des données, la lecture du

graphique est inévitablement un peu plus complexe et moins immédiate, il est nécessaire de ne faire ce choix que lorsque le message du graphique en est nettement plus lisible, et quand, par ailleurs, le public destinataire est en mesure de le comprendre. Enfin, la matérialisation par les axes du graphique doit être d'autant plus explicite que l'échelle retenue est complexe.

C'est par les axes du graphique que l'échelle retenue pour la représentation des données prend corps au sein de l'illustration. C'est dans l'aire graphique présentée précédemment que s'inscrivent les axes du graphique. Par convention, les axes s'intersectent dans le coin bas/gauche de l'aire graphique, à moins que des valeurs négatives n'imposent de modifier ce positionnement. Si un placement alternatif de ces axes est préféré (coin haut gauche, ou bord droit de l'aire graphique par exemple), il doit toujours être fait uniquement dans l'objectif d'améliorer la lisibilité du graphique, et rejeté sinon, car ces placements ne sont pas intuitifs pour l'œil et auront tendance à compliquer la lecture.

Sur ces axes doivent être positionnées des graduations. Celles-ci sont dépendantes des données représentées. La lisibilité est davantage préservée avec une graduation régulière des axes. Les valeurs absentes de la série de données (dans le cas par exemple de données discrètes) ne doivent pas être pour autant oubliées sur les axes, elles doivent y apparaître avec la même régularité que les valeurs effectivement présentes dans la série : les discontinuités dans la série de données sont une information en soi quant à la distribution des données et ne doivent pas être dissimulées.

De ces axes, et du positionnement qu'ils prennent dans l'aire graphique découlent assez naturellement d'éventuels quadrillages complémentaires destinés à faciliter le repérage au sein de l'aire graphique des valeurs prises par les données. Ceux-ci ne doivent être envisagés que dans la mesure où ils facilitent la lecture. Une attention particulière doit être portée à leur mise en forme : l'épaisseur des traits, la couleur, la fréquence des axes intermédiaires, etc. doivent être toujours choisis avec un objectif de lisibilité accrue. Ces quadrillages peuvent en effet assez rapidement venir alourdir le graphique.

Enfin, ces axes doivent être au minimum légendés (mention du titre des axes en fonction de la variable qui y est reportée, ainsi que l'unité de mesure s'il y a lieu).

La notion de coefficient de mensonge a été définie par Edward Tufte [1] comme la mesure de l'écart entre la réalité des données représentées et l'impression donnée par la représentation graphique qui en est faite. Il l'exprime sous une forme mathématique comme suit :

$$\text{Coefficient de mensonge} = \text{importance de l'effet visuel} / \text{importance de l'écart quantitatif.}$$

Si le ratio est proche de 1, alors la représentation visuelle s'efforce d'être fidèle à la réalité des données sous-jacentes. En cherchant à atteindre un coefficient de mensonge égal à 1, l'auteur d'une représentation graphique vise une moindre distorsion de la réalité des données par le graphique.

### **Le conteneur visuel du graphique :**

C'est en quelque sorte le graphique lui-même : la courbe elle-même dans un graphique en courbe, les barres d'un histogramme, les points d'un nuage, etc. Il est lui aussi très conditionné par la série de données elle-même, et chaque type de graphique possède son jeu de règles (surfaces proportionnelles aux fréquences de classes pour les barres d'un histogramme par exemple), mais il existe cependant des possibilités de jouer sur ce conteneur, toujours avec l'objectif d'améliorer la lisibilité du graphique. Pour jouer sur ce conteneur, l'auteur du graphique dispose de différentes variables visuelles, comme l'épaisseur des traits, leur forme (trait plein, tireté de largeur variable, etc.), leur couleur, ou la trame de remplissage pour les conteneurs de nature surfacique (secteurs d'un graphique en « camembert », barres d'un histogramme ou d'un diagramme en barres, etc.) par exemple. Des choix sur les couleurs, l'épaisseur, peuvent permettre de faciliter la lecture en la rendant plus intuitive par exemple : choix de couleurs figuratives associées dans l'inconscient collectif au

phénomène que l'on étudie, par exemple ; l'épaisseur, la forme d'un trait peuvent également être mobilisées pour mettre en exergue une information par rapport à une autre, par exemple (cf. la couleur, et les techniques de mise en exergue).

### **Les textes d'explicitation (titre, sous-titre, source, etc.) :**

Ces textes sont un élément majeur pour la bonne compréhension du graphique. Dans l'absolu, si le graphique a bien été conçu comme un document à part entière, ce sont les textes qui, en complément du graphique lui-même et de sa légende, doivent permettre une pleine compréhension du phénomène représenté. Une bonne optique de construction pour ces textes d'accompagnement (titre et sous-titre éventuel) est de les concevoir comme une forme de conclusion-résumé du graphique.

Ils peuvent être, soit positionnés à l'intérieur de l'aire graphique, soit reportés à l'extérieur, en fonction des modalités de mise en page et de l'intégration du graphique au texte qu'il illustre. Cependant, si le titre et l'éventuel sous-titre sont placés à l'extérieur de l'aire graphique ils devront en demeurer très proches afin de ne pas altérer la compréhension. Ils doivent être relativement concis, et ne pas omettre pour autant quelques informations essentielles (lieu, date, éventuellement choix des unités à moins qu'il ne figure dans la légende, par exemple).

La source quant à elle doit mentionner de manière brève mais complète le nom du fournisseur de la donnée, la date et, s'il y a lieu, l'auteur et la date du ou des traitements qui a conduit au graphique (exemple : *Sniir-AM, 2012 ; InVS, 2013* : les données sont issues du Système national d'information inter-régimes de l'Assurance maladie (Sniir-AM) et datent de 2012 ; elles ont ensuite été exploitées par l'InVS pour aboutir à des résultats en 2013).

### **La couleur :**

La couleur représente un outil pertinent pour la représentation graphique des données car elle a un très fort pouvoir visuel. Il convient de l'utiliser en tenant cependant compte d'un certain nombre de paramètres qui sont susceptibles d'avoir un impact assez fort sur la lisibilité et la compréhension du graphique. En effet, la couleur a un fort pouvoir discriminant, mais les humains ne sont pas tous égaux face à la visualisation de la couleur. Certaines anomalies entraînent une perception de la couleur moins efficace, voire ne permet pas la distinction de certaines nuances (cas du daltonisme). Il existe des outils disponibles en ligne très aisément qui permettent de s'assurer que les couleurs que l'on a choisies pour une représentation visuelle ne constitueront pas un obstacle à la lecture pour les daltoniens<sup>12</sup>. Les trames colorées proposées sur le site de Cynthia Brewer<sup>13</sup> sont également classées en fonction de leur lisibilité par ce public particulier. Pour ce qui est spécifiquement des couleurs utilisées en cartographie, un travail de thèse a été réalisé sur ce sujet et soutenu en 2011 [13;14].

La couleur est également largement tributaire de l'impression du document. Il ne faut éventuellement pas négliger dès la conception du graphique le fait que la lecture de ce dernier ne se fera peut-être pas dans les conditions optimales dans lequel il a été créé. Une vérification d'une bonne transformation en niveaux de gris à l'impression des couleurs retenues peut s'avérer utile.

Enfin, la couleur doit être utilisée moyennant quelques précautions pour conserver sa pleine efficacité visuelle. Ainsi, il existe des liens assez forts dans l'inconscient entre certaines couleurs et des phénomènes associés. Les avoir en tête lors de la conception, permettra une utilisation à bon escient des couleurs connotées et, du même coup, une lecture plus immédiate. Il est tout de même intéressant de constater que ces couleurs et les phénomènes associés peuvent différer en fonction des cultures. Dans les cultures occidentales, le rouge est associé au chaud, au sang, au danger, à

---

<sup>12</sup> <http://www.vischeck.com/vischeck/>

<sup>13</sup> <http://colorbrewer2.org/>

l'interdit, etc. alors qu'en Chine par exemple il est plutôt associé à des événements heureux, au mariage, à la chance. Là encore, le public destinataire du graphique ne doit pas être oublié.

### **La forme :**

La forme est, avec la taille et la couleur, la variable que l'on identifie le plus facilement et celle dont on perçoit le plus aisément les variations. La forme d'un élément graphique est définie par les contours de cet élément. Cette variable peut s'appliquer à des éléments graphiques de type ponctuel (marqueur sur un graphique, par exemple, les points d'un nuage de points ; sur une carte il s'agira des éléments ponctuels de localisation, tels que les villes sur une carte de France), ou à des éléments graphiques de type linéaire (tracé d'une courbe, représentation linéaire sur une carte : réseau routier ou hydrographique, par exemple). Elle n'est pas utilisée pour des éléments surfaciques sauf dans des cas très précis : pour les cartes en anamorphoses<sup>14</sup>, par exemple, la forme des entités surfaciques est déformée à dessein afin de représenter une information donnée, c'est la forme prise par l'entité qui est porteuse de sens. Dans un graphique, si l'on souhaite utiliser la forme pour discriminer des éléments surfaciques, cette variation de forme ne se limitera qu'aux contours des surfaces (de ce fait ce sera plutôt une variation de forme sur le linéaire du contour).

Pour des objets ponctuels, la déclinaison est quasi infinie, des éléments géométriques simples aux symboles figuratifs les plus élaborés, qu'en pratique on tâchera d'éviter tant ils peuvent éventuellement compliquer la lecture. Pour les objets linéaires, les variations de forme passent par les tiretés de tout type, de toute longueur et se déclinent là encore quasiment à l'infini, si ce n'est la limite de différenciation entre deux symboles utilisés dans une même représentation.

### **La taille :**

En cartographie, la variation de taille est une variable majeure de la représentation des variations d'effectifs. Pour ce qui est de la représentation dans un graphique, la taille est souvent largement conditionnée par la série de données et par les dimensions de l'aire graphique sur laquelle le graphique se déploie. De ce fait, elle n'est pas un paramètre directement maîtrisable par l'auteur du graphique, elle peut, à la marge, servir à mettre en exergue telle ou telle information (épaisseur d'un trait, ou grossissement d'un marqueur visuel pour mettre en avant la donnée correspondant à ce point, par exemple).

### **La légende :**

La légende tient lieu de décodeur pour la lecture et l'interprétation du message porté par le graphique. En ceci elle joue un rôle indispensable dans la compréhension et elle constitue un élément qui doit être tout particulièrement soigné lors de la réalisation du graphique. Les couleurs, les variations éventuelles de forme du conteneur graphique (variation de formes des points, ou d'un linéaire – tireté ou non par exemple) doivent y être explicitées. Des mentions particulières concernant les axes du graphique et leur légendage ont été proposées plus haut. La légende peut prendre place dans l'aire graphique, si son importance le permet : elle ne doit pas entraver la lecture ou encombrer de manière trop importante le graphique au risque d'en compromettre la lisibilité. Dans certaines circonstances, il sera opportun de privilégier un regroupement dans un caisson unique (une seule zone de texte par exemple) des éléments de légende, dans certains autres cas, la répartition sur le graphique de ces éléments permettra de gagner en rapidité de lecture, tout dépend du contexte.

---

<sup>14</sup> Une carte en anamorphose est une déformation de la réalité géographique d'un territoire en fonction de l'importance d'un phénomène donné. On peut par exemple figurer un territoire découpé en communes en augmentant la superficie des communes pour lesquelles le caractère d'intérêt est le plus fort, on peut également déformer les localisations relatives afin de donner une représentation de la perception des temps de parcours, par exemple. Difficiles à lire parce qu'elles sortent de l'image que l'on se fait de la réalité géographique d'un territoire, elles restent peu utilisées.



## 2.2.2 Choix des outils

Les outils disponibles au DSE en interne pour réaliser les représentations graphiques proposées précédemment sont divers et n'offrent pas tous les mêmes possibilités et variantes. De manière générale, la simplicité d'accès des outils bureautiques standards n'est pas forcément un gage de rigueur et Excel doit, s'il est relativement abordable et simple d'accès, être néanmoins utilisé avec précaution, en ayant en tête certaines de ses limites. Pour les utilisateurs qui souhaitent privilégier Excel, il existe des ressources<sup>15</sup> qui permettent d'accéder à des fichiers d'exemples pour la réalisation guidée d'un certain nombre de graphiques dans Excel. Le choix du graphique est orienté, comme cela a été fait dans ce guide, en fonction de ce que l'auteur souhaite montrer de ses données à son lecteur.

Les autres outils disponibles en interne (SAS, Stata, R) nécessitent tous une prise en main impliquant un investissement un peu plus important. Ils ne présentent pas tous les mêmes facilités de paramétrage (choix des couleurs, positionnement des textes sur le graphique, choix des symboles et conteneurs, etc.) et il faut parfois un peu de temps avant de parvenir à faire un graphique à la fois juste et esthétiquement satisfaisant. On trouve cependant aujourd'hui sur Internet un grand nombre de forums ou de liste de diffusion de bouts de code permettant de prendre en mains ces outils et de se familiariser avec leur utilisation. Cela dit, il est important d'avoir en tête que l'utilisation de ces différents outils peut nécessiter des mises en formes des données différentes, donc un travail de préparation des données non-négligeable qui fait que l'utilisation des ressources glanées sur Internet comme celles de collègues ou de partenaires ne dispense pas d'un minimum de familiarisation préalable avec le logiciel que l'on utilisera.

En lien avec la mouvance de l'open data, la mission Etalab<sup>16</sup> fait un tour d'horizon des possibilités simples de « datavisualisation » disponibles actuellement pour la diffusion de ses données<sup>17</sup>.

## 2.2.3 Précautions à prendre, écueils à éviter

Il est important de ne pas négliger les **erreurs** susceptibles d'être introduites par certains outils ou lors de leur utilisation. On peut citer l'exemple d'Excel qui fonctionne par la définition manuelle par l'utilisateur des plages de données à retenir pour la création d'un graphique, et non pas par un fonctionnement en mode « base de données ». L'ajout d'une donnée supplémentaire dans une série dans Excel oblige à redéfinir la plage des cellules à afficher. L'obtention du graphique nécessite donc une intervention, source potentielle d'erreur. Une approche vectorielle, comme dans la plupart des logiciels statistiques, effectue cette manipulation automatiquement puisque le vecteur, contenant les données, est pris dans son intégralité lors de la création du graphique.

Les **difficultés de lecture** peuvent être introduites par les échelles logarithmiques. La manipulation des échelles logarithmiques peut s'avérer nécessaire à la compréhension d'un phénomène ou sa représentation du fait d'une distribution des modalités particulière. Il faut néanmoins être conscient qu'à la lecture, la compréhension s'avère un peu plus délicate. Une attention spécifique du lecteur doit être appelée sur ce choix d'une échelle particulière, par un texte d'accompagnement ou une mention sur le graphique lui-même, et une justification dans le texte dans lequel le graphique s'insère.

---

<sup>15</sup> <http://labs.juiceanalytics.com/chartchooser/index.html>

<sup>16</sup> La mission Etalab est chargée du pilotage, sous l'autorité du Premier ministre, de la politique d'ouverture et de partage des données publiques.

<sup>17</sup> [http://wiki.data.gouv.fr/wiki/Outillage\\_pour\\_les\\_datavisualisations](http://wiki.data.gouv.fr/wiki/Outillage_pour_les_datavisualisations)

Le **choix des couleurs** nécessite lui aussi une certaine vigilance. Un certain nombre de couleurs sont associées culturellement à certains phénomènes comme cela a été précisé précédemment. La prise en compte de ces associations ne doit pas se faire totalement à la légère sous peine de fausser l'interprétation du graphique. Par ailleurs, l'utilisation de certaines couleurs ou tonalité peut dans certains cas s'avérer maladroite (utilisation du rouge un peu stigmatisant dans le cas de la représentation d'une information un peu sensible, par exemple, oppositions trop systématiques rouge/vert = positif/négatif).

Des **échelles fausses** notamment lorsque l'on a pour objectif de comparer des informations entre elles, sont une source d'erreur importante. Avec certains outils, la possibilité d'utiliser une échelle fausse n'est pas totalement absente. Ainsi, dans un graphique figurant une courbe épidémique avec des données journalières, si, pour un jour, il n'y a pas de cas, l'absence de ce jour sur l'axe fausse l'information. Les échelles doivent intégrer l'ensemble de l'étendue, et il faut veiller à éviter, autant que faire se peut, la « compression » des échelles avec suppression de certaines valeurs, même pour gagner de l'espace car les repères visuels et les proportions sont alors faussés et l'interprétation du lecteur erronée.

Les **écueils visuels** sont nombreux : sans forcément fausser l'interprétation ou la compréhension d'un graphique, la surcharge en texte ou en information d'un graphique peut conduire à une perte de lisibilité importante. On retiendra que, même si la tendance actuelle va souvent vers un fleurissement des infographies et autres représentations parfois très utilisatrices de symbolisations en tout genre, la sobriété, la sélection réfléchie des informations à mettre en exergue sont très souvent des gages solides d'une bonne compréhension d'un graphique.

Enfin, dans des situations de représentation complexe, (information complexe en soi, ou bien combinaison d'informations, nombre d'individus/d'entités à distinguer, période d'étude longue ou divisée en plusieurs sous-périodes, etc.), une piste à explorer pourrait consister par exemple à tester plusieurs méthodes de représentation puis de les soumettre à diverses personnes pour savoir ce qu'elles comprennent et, éventuellement, choisir une méthode plutôt qu'une autre. Dans la plupart des situations de représentation graphique d'une ou plusieurs informations, il n'y a sans doute pas une seule et unique manière de les représenter et l'exercice pourrait être intéressant de tester la compréhension par un panel de différentes personnes afin d'évaluer et éventuellement remettre en question les choix de représentation que l'on a envisagés. Dans tous les cas, procéder par itération semble une démarche assez adaptée à la recherche du meilleur mode de représentation graphique en fonction d'objectifs préalablement fixés.

## 2.3 S'adapter à son public : un enjeu de taille

La question du public destinataire du graphique doit être posée en amont de la réalisation des graphiques eux-mêmes, car elle conditionne souvent le choix du mode de représentation à privilégier. Par ailleurs, on s'aperçoit également que la question de la représentation graphique intervient à différentes phases d'un projet, d'un travail et qu'en fonction de ces phases, le public concerné sera différent.

En effet, on peut être amené à concevoir une représentation graphique au moment de la description des données, lors d'une phase d'appropriation pendant laquelle les représentations vont essentiellement être destinées aux membres d'une équipe projet, d'un groupe de travail voire éventuellement d'un comité scientifique, soit un groupe constitué de pairs, avec lesquels le partage d'un vocabulaire commun, de notions communes, et de clefs de lecture graphique va *a priori* de soi. Le graphique vise alors à décrire les données, en amont de l'analyse. La représentation graphique qui est réalisée à cette étape du travail n'est donc à proprement parler pas celle de résultats, mais elle permet à un ensemble de personnes travaillant sur un sujet commun de s'approprier le matériel de

façon commune. Les graphiques produits pourront éventuellement, à cette phase, se présenter sous la forme de documents de travail, moins finalisés qu'aux étapes ultérieures.

À un second niveau, la représentation graphique s'adresse à un public scientifique averti pour qui la compréhension ne pose *a priori* pas de difficultés. Mais bien souvent cette communication des résultats est soumise à des contraintes assez fortes. La publication dans de nombreuses revues scientifiques impose un nombre de graphiques limités, ce qui implique d'avoir priorisé en amont les résultats à présenter sous cette forme. Par ailleurs des exigences de mise en forme pourront éventuellement dans certains cas être imposées. À ce niveau de communication, ce n'est pas tant la compréhension du lectorat qui est susceptible de poser problème que les contraintes et les priorités de représentation, qui touchent à la mise en forme sans exiger pour autant d'adaptation ou de remise en cause des graphiques en eux-mêmes.

Enfin, pour un certain nombre de travaux, un troisième niveau de communication s'ajoute puisque, dans un certain nombre de situations, les résultats doivent faire l'objet d'une présentation à un public non averti. La représentation graphique doit alors tenir compte du fait que le lectorat risque de ne pas maîtriser l'ensemble du vocabulaire et les notions qui sont manipulés. Un travail d'adaptation est alors nécessaire. À ce niveau de communication c'est bien la bonne compréhension du public qui est en jeu et, de ce fait, un travail d'adaptation voire de vulgarisation pourra dans certains cas s'avérer nécessaire.

Un assez bon exemple en la matière est celui des travaux d'Évaluation quantitative des risques sanitaires (EQRS) qui peuvent être réalisés suite à des sollicitations locales. En effet, les méthodes et les concepts manipulés en EQRS sont loin d'être simples à saisir pour un public non averti. Or il s'avère que souvent, ces travaux doivent faire l'objet de communication à un public local très demandeur d'informations qui doit être en mesure de comprendre le message. Plusieurs exemples de travaux conduits en EQRS sont présentés plus loin dans ce travail, pour illustrer cette complexité.

## 3. État de l'art au DSE : les modes de représentations graphiques couramment utilisés

Après un aperçu rapide des « bonnes pratiques » et des écueils à éviter, il sera question, ici, de faire un tour d'horizon des représentations graphiques couramment utilisées dans les travaux du DSE, afin de mieux en cerner la variété, mais aussi d'envisager d'éventuelles améliorations. De ce fait, il a été proposé de présenter une sélection de travaux de chacune des unités du département dans ce domaine, en détaillant, à chaque fois :

- les objectifs du graphique ;
- le public destinataire ;
- les difficultés rencontrées ;
- les éléments et les questions éventuelles en vue d'une amélioration.

Cette présentation distinguera deux grands groupes : les graphiques d'une part, et les cartographies d'autre part. On traitera également dans un troisième temps de quelques exemples de schémas, même s'ils s'éloignent un peu de la problématique précise de ce travail : en effet, les schémas visent généralement à présenter une démarche, au-delà des seules données ou des seuls résultats. Mais ils constituent un outil de communication très intéressant au pouvoir visuel non négligeable.

### 3.1 Représentations graphiques

#### 3.1.1 Représentations graphiques des résultats dans le cadre de l'investigation de syndromes collectifs inexplicables

En septembre 2010, le personnel de différents services administratifs du Conseil général de l'Eure-et-Loir (CG28) déménage dans les locaux d'un ancien laboratoire d'analyses médicales. Assez rapidement, l'apparition d'odeurs nauséabondes provoque, chez certains membres du personnel, le développement de différents symptômes, assez peu spécifiques (céphalées, picotements de la gorge, toux, douleurs abdominales, etc.). Mi-octobre 2010 les locaux sont temporairement fermés, des travaux de nettoyage des sols, des interventions sur la ventilation ainsi qu'un certain nombre d'analyses sont réalisés. Mais fin janvier 2011, les symptômes réapparaissent : l'Agence régionale de santé (ARS) est alors saisie de cette question.

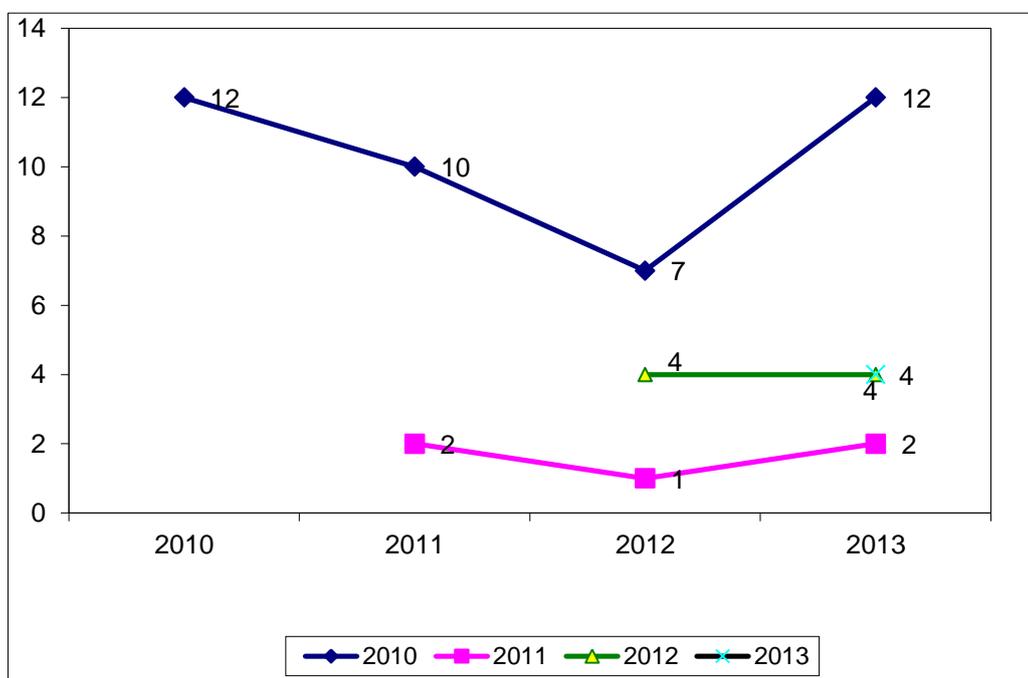
Les exemples ci-après sont extraits de la présentation réalisée lors d'une réunion de « cellule de coordination », instance que l'on recommande de réunir en cas de syndromes collectifs inexplicables et qui rassemble toutes les personnes concernées par le problème : le directeur de la structure, les représentants des salariés, le gestionnaire du bâtiment, les divers experts impliqués (médecins, toxicologues, météorologistes, épidémiologistes, etc.) et les autorités (inspection du travail, ARS, etc.). L'objectif de cette présentation est de détailler les caractéristiques de l'épisode, afin de mettre en œuvre les investigations nécessaires et éventuellement d'envisager des actions. Ces caractéristiques portent notamment : sur la durée, la répétabilité, le nombre de personnes concernées, elles portent également sur les lieux de survenue des symptômes, et enfin, sur les symptômes ressentis et leur fréquence. Trois graphiques ont été extraits de la présentation afin d'en détailler les objectifs et les particularités, d'en relever les limites et éventuellement de donner quelques pistes d'amélioration.

## Représenter la persistance des symptômes au cours du temps

L'épisode en exemple comporte plusieurs spécificités : il a duré dans le temps, et certaines personnes ayant été concernées une première fois, ont été de nouveau concernées à la seconde, mais pas toutes ; d'autres n'ont ressenti les symptômes que plus tard, certaines personnes ont ressenti des symptômes sur l'ensemble de la période, d'autres, de façon plus ponctuelle. Le souhait de pouvoir représenter l'ensemble de ces informations sur un seul et même graphique pose des difficultés dans les choix de représentation et le choix a été initialement fait de réaliser ce graphique (figure 44) simple en apparence mais dont la lecture s'avère en fait assez délicate. 2013, année en cours au moment de la réalisation du graphique, est peu lisible et se superpose aux nouveaux cas de 2012 : en effet, les effectifs de nouveaux cas en 2012 et 2013 sont semblables (4 cas).

I Figure 44 I

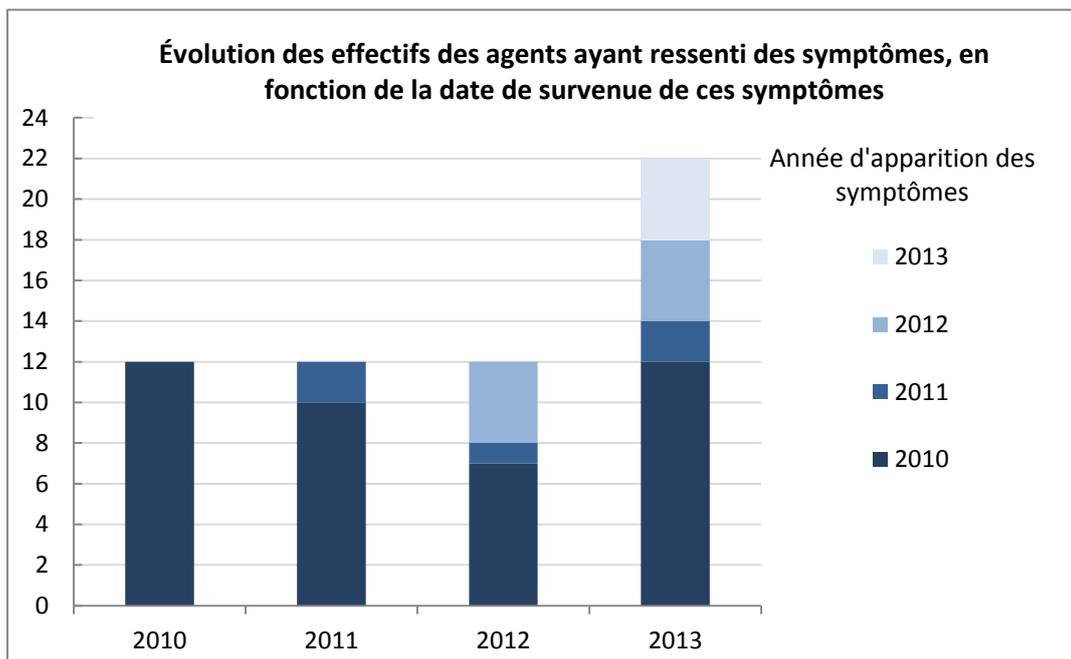
Représentation initiale des variations des effectifs de cas selon la date d'apparition des symptômes



À partir des mêmes données, en utilisant le même outil (Excel), et en visant les mêmes objectifs, un second graphique est ici proposé (figure 45).

Les objectifs qui étaient de représenter l'évolution des effectifs dans le temps tout en distinguant les groupes de cas en fonction de la date de survenue des symptômes, sont bien atteints. De plus ce graphique parvient à faire apparaître une information qui n'était pas directement lisible dans la version initiale : en effet, ainsi représentés, les effectifs des différents groupes peuvent être sommés selon l'année. On peut donc lire à la fois l'évolution annuelle du nombre de cas dans chacun des groupes, mais aussi de l'évolution de l'effectif total des agents ayant ressenti des symptômes par an.

Proposition de refonte du graphique précédent dans un diagramme en barres empilées



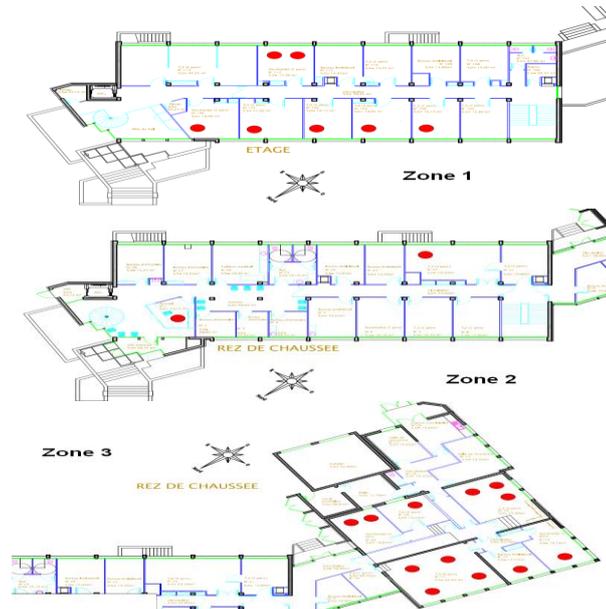
#### Localiser la survenue des symptômes

Dans l'investigation de ce type de syndromes, il existe une autre difficulté liée à une contrainte technique : lors de la survenue d'un épisode de syndromes inexpliqués dans un bâtiment, les investigations nécessitent de s'interroger sur une possible explication liée au bâtiment lui-même (*i.e.* explication environnementale). Dans le cas présenté ici, il a été envisagé que le système de ventilation des locaux puisse être en cause. Ces investigations nécessitent de localiser les lieux de survenue des symptômes (figure 46). Cependant, à la déontologie implique un strict anonymat dans cette description, anonymat qui ne peut être considéré comme respecté si l'on localise les cas sur les plans en les pointant dans les bureaux eux-mêmes, puisque la répartition du personnel dans les bureaux est connue. D'autre part, les plans qui sont fournis aux épidémiologistes dans ce genre de situation sont souvent transmis dans des formats numériques image ou PDF figés, sur lesquels aucune modification de fond n'est possible. La représentation des lieux de survenue sous la forme de points figurant en fait des cas comme ici (les cas sont figurés par formes ovales rouges) n'est pas satisfaisante.

Il pourrait être envisagé, sans que cela ne nécessite de pouvoir modifier les plans tels qu'ils étaient fournis, de représenter non pas les cas par des points, mais par des zones de survenue sous forme de rectangles (figure 47) ou d'ovales, agrégeant de cette façon plusieurs cas et évitant de rompre l'anonymat.

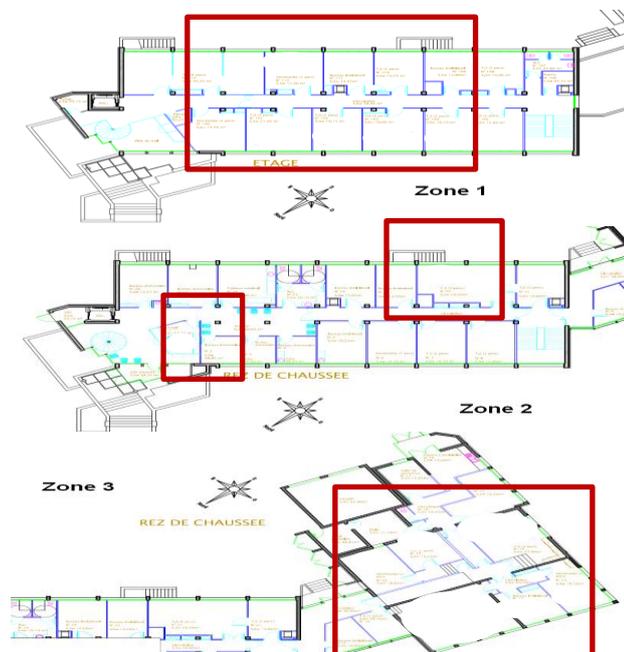
I Figure 46 I

Localisation sur un plan des locaux des cas ayant présenté des symptômes, avec les problèmes d'anonymat que pose la précision de ce schéma



I Figure 47 I

Proposition d'alternative au schéma précédent pour localiser les victimes de façon suffisamment imprécise pour ne pas compromettre leur anonymat

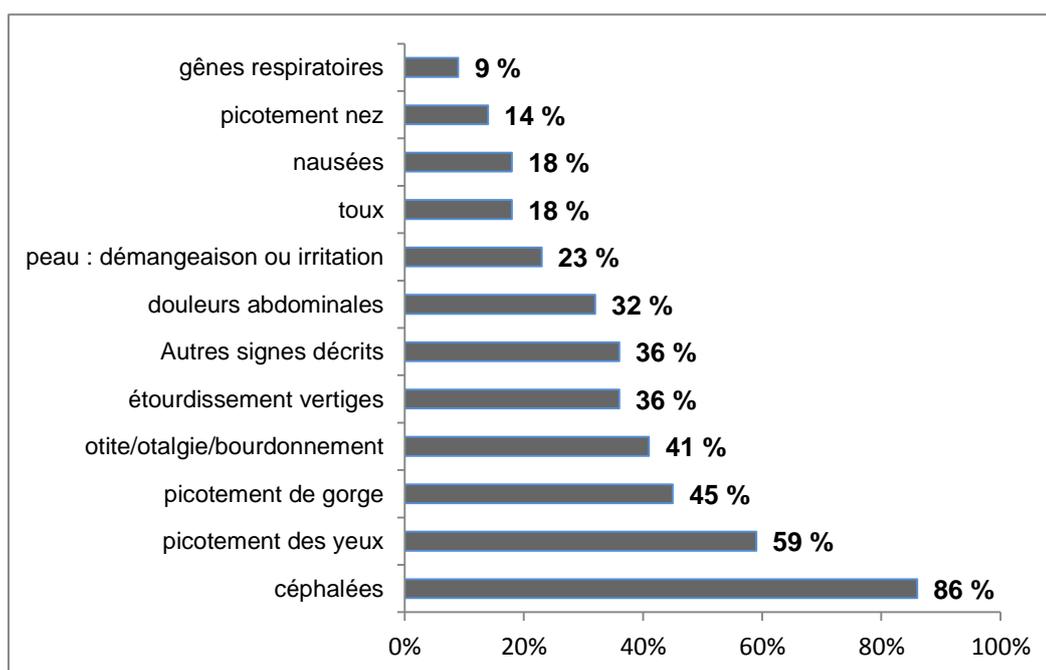


### Représenter la fréquence de ressenti des différents symptômes décrits

Enfin, plus classiquement lors de ces épisodes, les symptômes ressentis sont assez divers et assez peu spécifiques, et un graphique représentant la fréquence de survenue des différents symptômes sur l'ensemble de la population touchée est souvent réalisé, comme celui-ci (figure 48), dont la lecture et l'interprétation ne posent *a priori* pas de problèmes.

I Figure 48 I

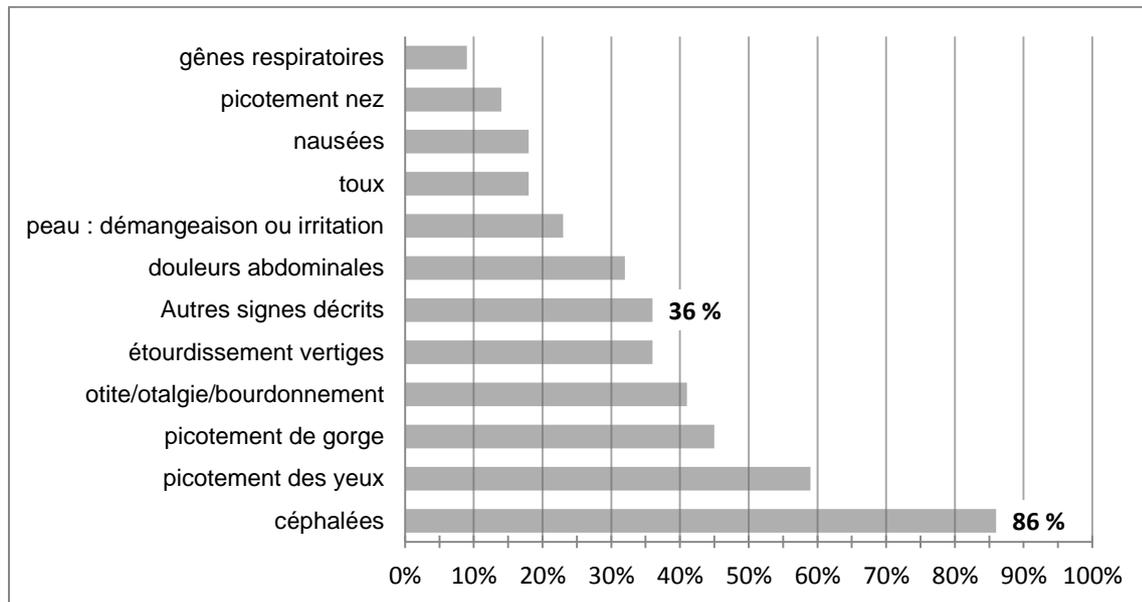
Représentation des fréquences de survenue des différents symptômes par un diagramme en barres



On pourrait faire le choix d'alléger le graphique en évitant de faire figurer systématiquement les valeurs prises par chaque classe de symptômes et choisissant plutôt de faciliter le repérage par rapport aux axes en faisant apparaître des lignes de graduations dans le fond du graphique. Seules quelques valeurs sont alors mentionnées sur le graphique lui-même ce qui permet éventuellement de mettre en avant des informations qui sont par ailleurs explicitées dans le texte d'accompagnement. Sur la figure 49 on ne fait plus figurer que le symptôme le plus fréquemment ressenti ainsi que de la proportion des cas ayant signalé d'autres symptômes que ceux qui sont détaillés ici. Des intitulés plus courts dans la description des symptômes sur le graphique auraient également joué en faveur de l'allègement.

## I Figure 49 I

Proposition d'amélioration pour le graphique précédent (fréquence de survenue des symptômes)



### 3.1.2 Représentations graphiques dans le champ des études d'impact sanitaire dans les domaines de l'air et du climat

Les exemples présentés ici sont extraits de travaux réalisés au sein de l'unité Air-Eau-Climat du DSE, et visaient principalement à illustrer des projets d'Évaluation d'impact sanitaire (EIS) dans des posters scientifiques présentés au congrès de l'International Society for Environmental Epidemiology (Isee). Ils s'adressent donc à un public relativement averti, mais dans le cadre d'un support de communication relativement contraignant, où la dimension esthétique a son importance pour attirer l'œil, dans la mesure où le nombre de posters présentés dans la session est extrêmement élevé ce qui ne permet évidemment pas de tous les voir. Nécessité est donc, dans un tel contexte, de jouer sur l'aspect visuel et esthétique. Cet objectif d'esthétique a son importance ici, et il doit être rappelé car il joue sans doute un rôle prépondérant, de manière plus générale, dans la lisibilité d'un graphique quel qu'il soit.

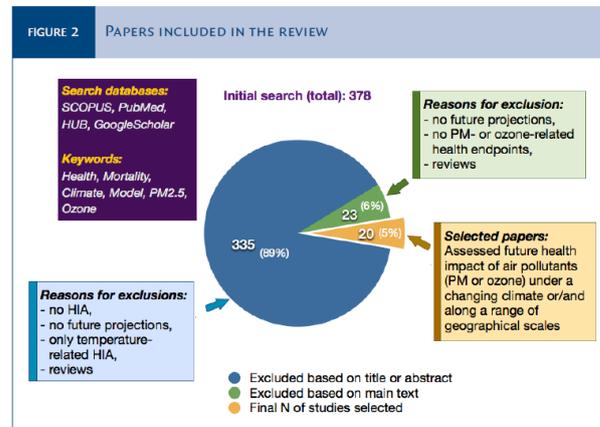
#### Illustrer la répartition des références exclues ou retenues dans un travail de revue de la littérature

L'objectif de ce graphique (figure 50) était de représenter une répartition entre les références retenues et les références exclues de la revue de littérature réalisée dans le cadre du projet Air Climate Health Impact Assessment (ACHIA)<sup>18</sup> ainsi que de catégoriser les références exclues en fonction des raisons de leur exclusion. Le choix d'un graphique en secteurs, ou camembert, est un choix classique pour représenter une répartition entre un nombre limité de catégories. La lecture de ce type de graphique et des légendes qui y sont apportées ne pose pas de difficulté particulière *a priori*, étant donné le faible nombre de catégories.

<sup>18</sup> <http://www.gisclimat.fr/projet/achia>

## I Figure 50 I

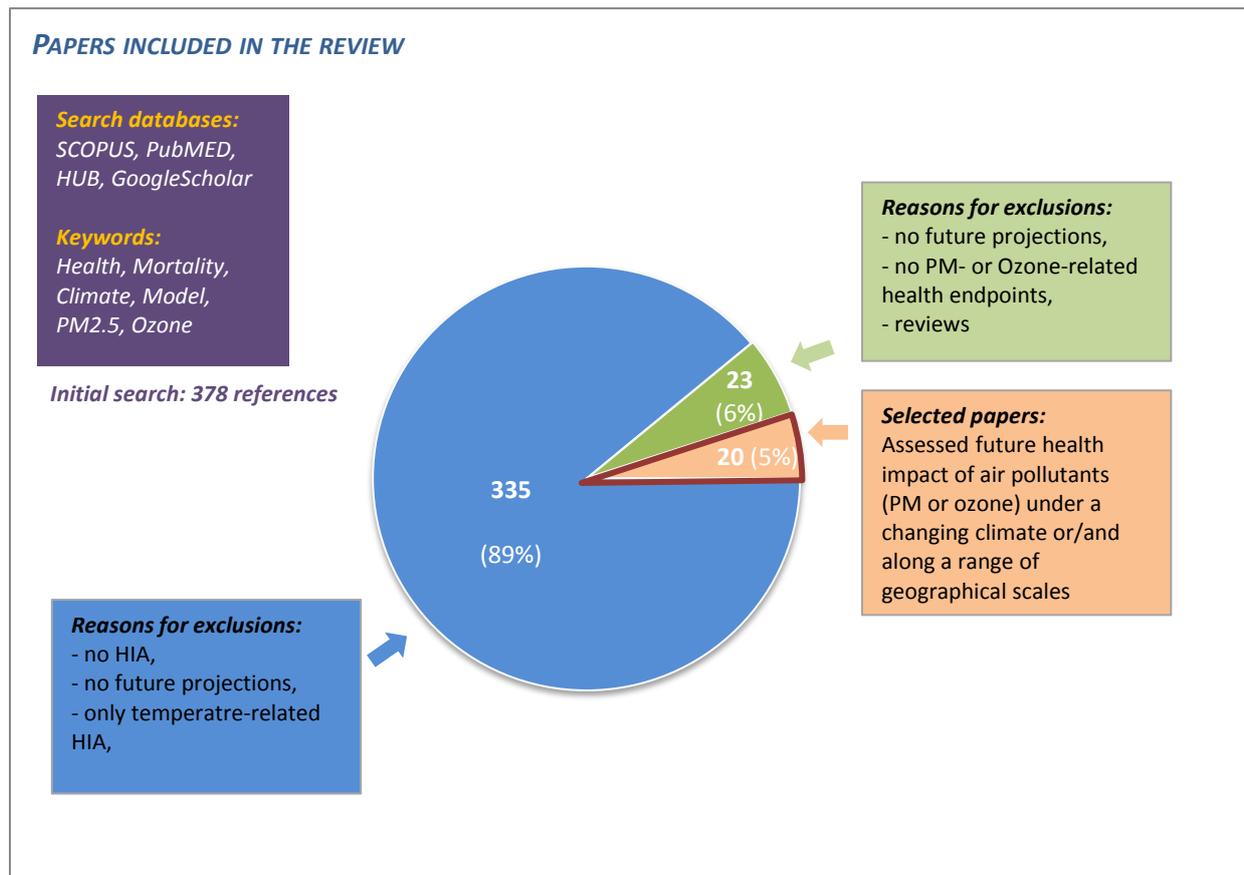
Graphique extrait du poster “A review of quantitative health impact assessments of ozone and particulate matter under a changing climate” – Victoria Likhvar, ISEE, 2013



Cependant, le choix de mettre en avant la catégorie des articles retenus, qui se trouve être la plus faible en termes d'effectif, par rapport aux autres catégories des articles exclus sur la base du titre seul ou sur la base du texte complet, en agrandissant le secteur correspondant semble mal approprié : en effet, lui donner une surface plus grande que celle qui lui revient, certes, lui donne effectivement davantage de poids, mais elle fausse aussi l'interprétation que l'on a de la part qu'elle représente dans le total. On a en effet l'impression que cette catégorie des articles retenus comprend un nombre de références équivalent voire supérieur au nombre des articles rejetés sur la base du texte complet. Le mode de mise en exergue de l'information ne doit pas introduire de détournement de la quantité représentée, et doit sans doute se faire autrement : choix d'une couleur plus vive, chiffres plus gros si l'espace le permet, détournement du secteur, etc.

Ci-après (figure 51), une proposition restant le plus fidèle possible au graphique d'origine, visant à mettre en exergue l'information d'intérêt – ici, la part des références retenues dans le travail de revue de la littérature – sans déformer l'information et fausser les données.

Proposition d'amélioration pour le graphique précédent



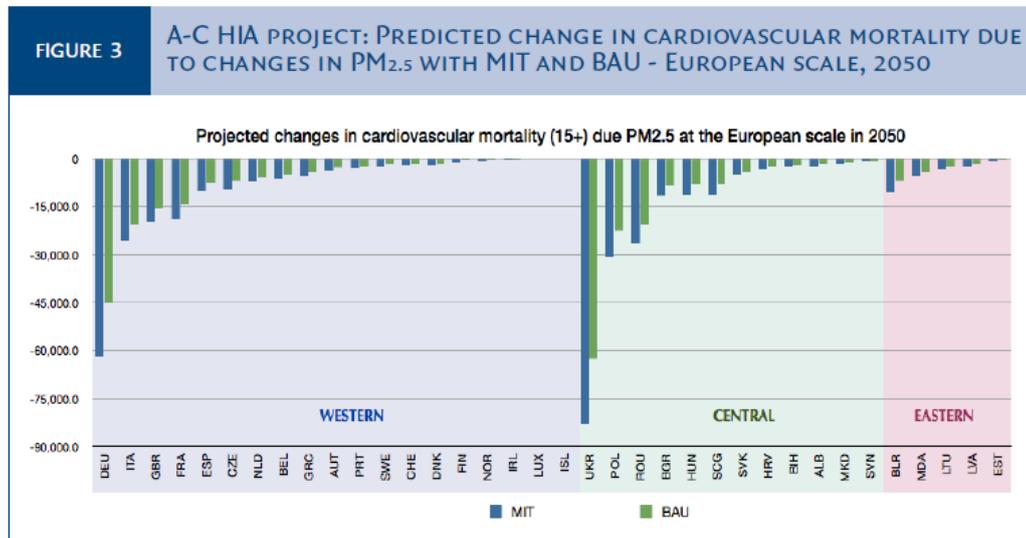
Représenter dans un graphique les résultats d'une EIS en fonction de différents scénarii sur un nombre assez élevé d'individus (ici, des entités géographiques)

Les deux graphiques ci-après visent à représenter les changements en termes de mortalité cardiovasculaire en Europe à l'échéance 2050 – dans les différents pays de l'Union Européenne mais aussi à une échelle plus agrégée (l'Europe occidentale, la région centrale de l'Europe, l'Europe orientale) – en fonction de deux scénarii : un scénario dit de « mitigation » où des actions seraient menées en vue de faire diminuer les niveaux de concentration en PM<sub>2,5</sub>, et un scénario dit « business as usual », pour lequel aucune action ne serait menée.

Pour le premier graphique (figure 52), on identifie bien les deux scénarii (les résultats pour l'un sont figurés par une barre bleue – mitigation, pour l'autre par une barre verte – business as usual).

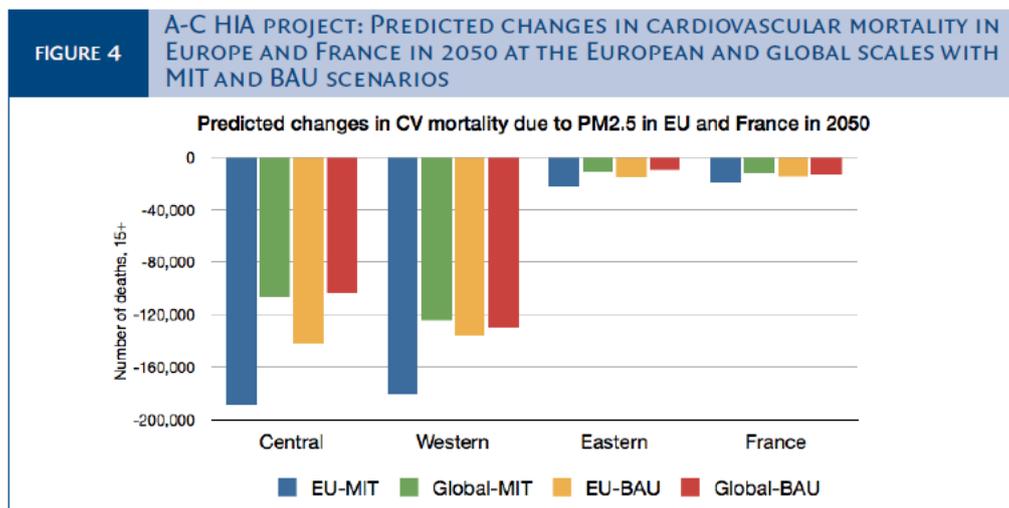
## I Figure 52 I

Projet ACHIA : évolutions prévisibles de la mortalité cardiovasculaire selon les deux scénarii (MIT et BAU) à l'échelle de l'Europe



## I Figure 53 I

Graphique extrait du poster “Quantifying health impacts of air pollution under a changing climate: from a global to the regional and local scales” – Victoria Likhvar, ISEE, 2013



Pour chaque pays, le graphique présente les deux résultats par deux barres, et les pays sont regroupés par zone de l'Europe. De gauche à droite, zone par zone, apparaissent à gauche les pays concernés par une diminution des décès par cause cardio-vasculaire la plus importante, et en allant vers la droite, les pays pour lesquels la diminution du nombre de décès est moins importante. La lecture de ce premier graphique ainsi construit ne pose pas de difficulté particulière. Le passage au second graphique (figure 53) cependant ne va pas de soi après le premier, sans doute essentiellement du fait d'une incohérence entre les couleurs utilisées dans le premier graphique (deux

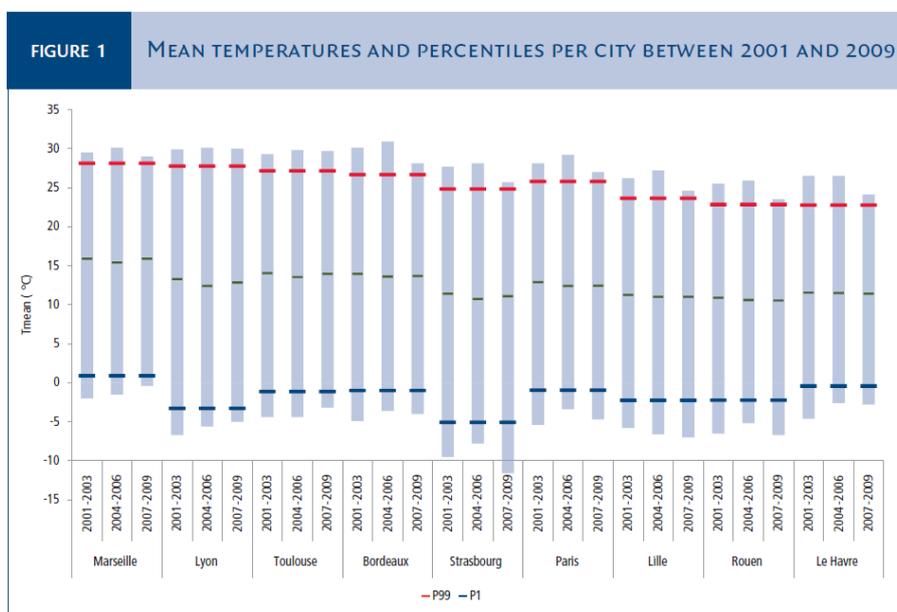
couleurs pour deux scénarii) et le second graphique : ces mêmes couleurs sont utilisées pour catégoriser autre chose dans le second graphique, le vert qui caractérisait un scénario « business as usual » dans le premier graphique est associé à un scénario de mitigation dans le second. Du coup, dans un contexte d'espace très contraint comme celui d'un poster, on ne perçoit pas réellement de prime abord ce qu'apporte le second graphique par rapport au premier si ce n'est qu'il compare les trois grandes zones (après agrégation des résultats par pays) entre elles et avec la France en particulier. L'interprétation en est un peu malaisée, et il aurait peut-être été plus simple de comparer les deux scénarii en conservant les codifications couleur utilisées dans le graphique précédent.

Représenter les variations de températures moyennes dans un nombre de villes élevé sur un découpage de la période d'étude en trois sous-périodes

L'objet de ce graphique (figure 54) est de représenter l'amplitude de variation de la température moyenne journalière dans les neuf villes du Programme de surveillance air et santé (PSAS) sur la période 2001 à 2009, en distinguant trois sous-périodes : 2001-2003, 2004-2006, et 2007-2009. Le fait de découper la période d'étude en trois alors que l'on dénombre neuf villes sur lesquelles la représentation graphique doit porter, fait que le nombre de barres à faire figurer est élevé, et c'est là que se trouve l'essentielle difficulté dans la compréhension de ce graphique. La lecture en est malaisée, et ce d'autant que les couleurs sont constantes : l'attribution de chacune des barres aux neuf villes et à la période correspondante n'est pas immédiate pour l'œil.

**I Figure 54 I**

Représentation des températures moyennes et percentiles par ville entre 2001 et 2009



L'utilisation de couleurs aurait pu sans doute pallier en partie cette difficulté : il était possible par exemple de faire alterner des barres gris clair pour la première ville, gris foncé pour la deuxième, de nouveau gris clair pour la troisième, et ainsi de suite. Les villes s'en seraient trouvées plus faciles à distinguer visuellement. Par ailleurs, l'indication des percentiles n'est pas cohérente avec l'indication de la valeur moyenne : les moyennes des températures moyennes sont calculées pour chacune des subdivisions de la période d'étude et pour chacune des villes, alors que les percentiles P1 et P99 sont

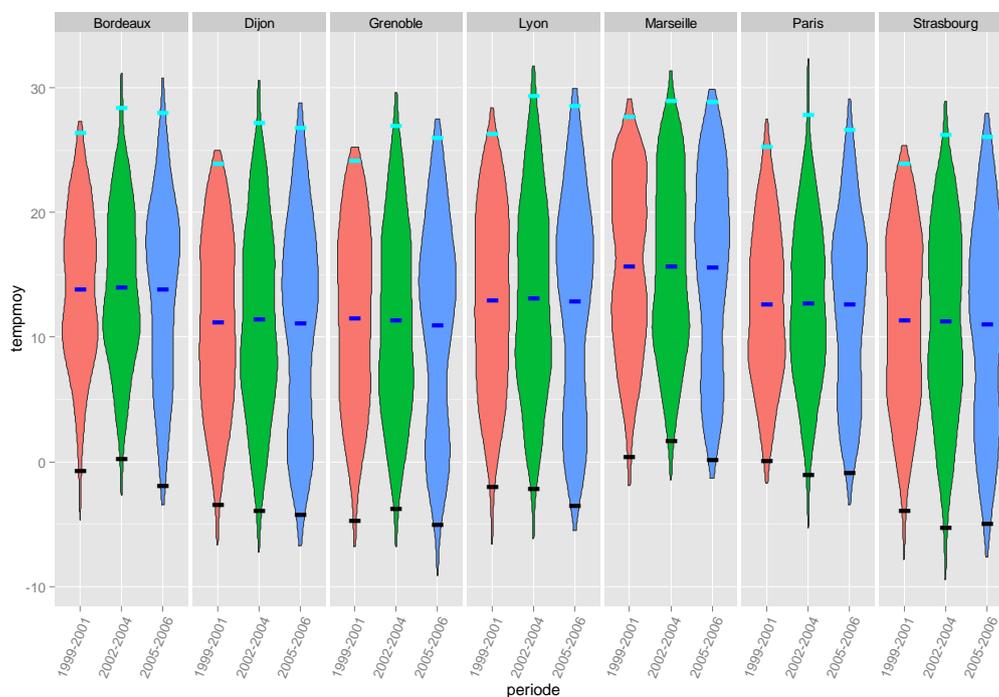
calculés, pour chaque ville, sur l'ensemble de la période d'étude (ce qui entraîne qu'ils ont la même valeur pour chacune des trois sous-périodes).

Enfin, les barres du graphique ne devraient pas empiéter sur les mentions de légende, ici la barre de Strasbourg recouvre en partie la mention de la période 2007-2009, il aurait été préférable de laisser davantage d'espace à la zone graphique pour que cela ne se produise pas.

Dans la mesure où ce graphique donne une mauvaise perception de la distribution des valeurs pour chacune des catégories, en laissant supposer que la distribution de la température est uniforme, il aurait été plus pertinent de présenter un box plot ou, mieux, un violin plot<sup>19</sup> (cf. une proposition en figure 55). Ici, pour correspondre au graphique initial, les traits représentent le P1, la moyenne et le P99, calculés cette fois-ci par période.

I Figure 55 I

Proposition d'améliorations pour la représentation des distributions des températures par ville



### Représenter l'augmentation de la mortalité en fonction de la température

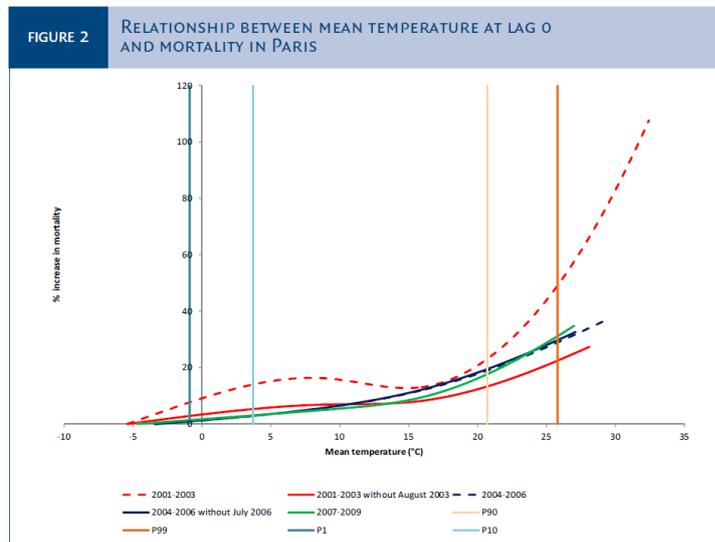
L'objet de ce graphique est de faire figurer la relation entre la température moyenne au « lag 0 » (le jour-même), avec la mortalité à Paris sur différentes périodes, en incluant ou en excluant les périodes de températures extrêmes comme l'épisode de canicule de 2003 ou dans une moindre mesure, celui de 2006.

La légende du graphique (figure 56) est relativement explicite et l'utilisation du code couleur est assez immédiate, on relie bien visuellement les deux périodes incluant des épisodes de températures extrêmes aux mêmes périodes lorsque l'on exclut les phases caniculaires. Cela en fait un graphique relativement bien compréhensible alors même que ce qu'il représente est une relation plutôt complexe.

<sup>19</sup> Voir ce graphique partie 2.1.1 Représenter la distribution statistique d'une ou plusieurs variables

## I Figure 56 I

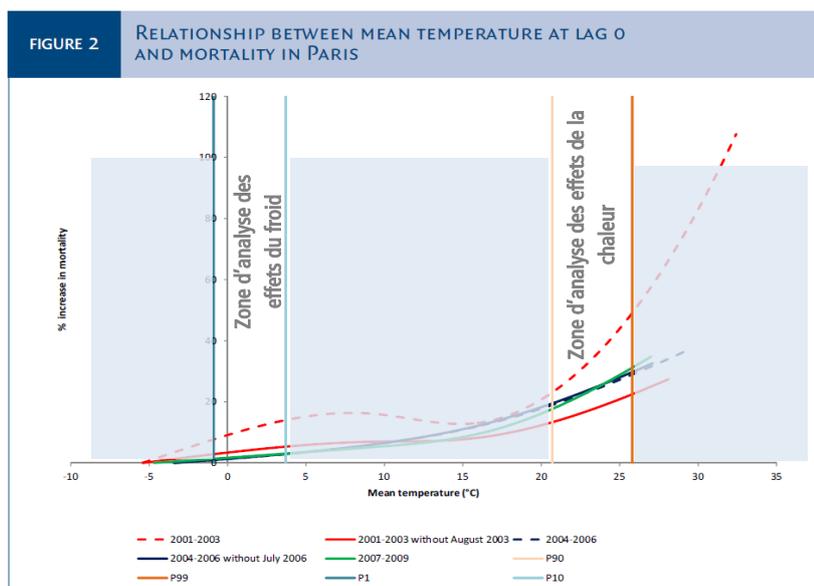
Graphique extrait du poster “Impact of temperature on mortality in nine french cities between 2001 and 2009”, Mathilde Pascal, ISEE, 2013



Cependant, rajouter des codes couleurs pour faire figurer les percentiles des températures moyennes ne facilite pas la lecture. Le graphique aurait peut-être gagné en lisibilité en représentant, comme le suggère la figure 57, par une zone gris clair, les deux intervalles P1 à P10 des températures sur lequel on analyse les effets du froid sur la mortalité, et P90 à P99 des températures sur lequel on analyse les effets de la chaleur. De plus, cette solution aurait permis d'éviter que la droite figurant la valeur P1 des températures ne vienne se superposer avec les graduations de l'axe des ordonnées.

## I Figure 57 I

Propositions pour améliorer la compréhension du graphique de la figure 56



Représenter les écarts entre les mesures fournies par deux systèmes de mesure d'un même polluant, en fonction de la saison

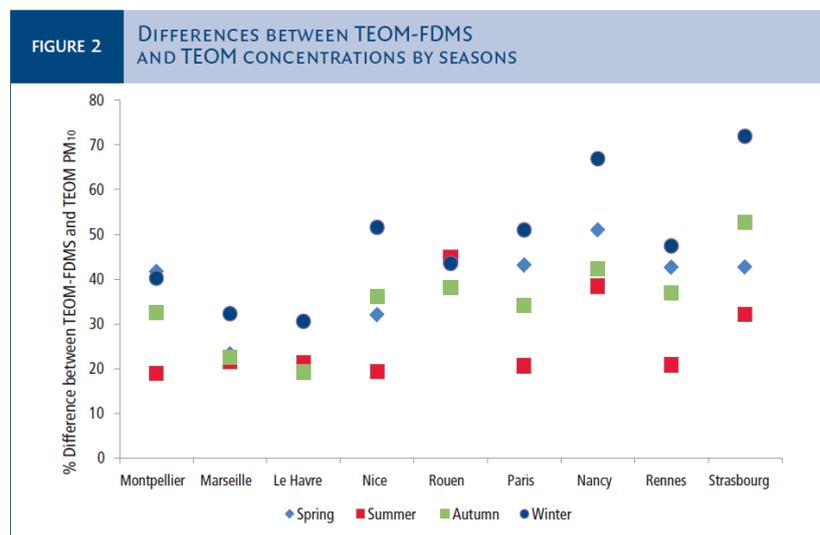
Le graphique ci-après (figure 58) a pour objectif de montrer les variations des écarts de mesures d'un même polluant (les PM<sub>10</sub>) entre deux systèmes utilisés dans différentes villes, en fonction de la saison. Il s'adresse à un public scientifique averti et vise à montrer que ces écarts varient de façon importante à la fois en fonction de la saison mais aussi de la ville dans lequel il est étudié et que de ce fait, il n'est pas possible de se contenter d'appliquer un simple facteur correctif sur les mesures effectuées par un système pour obtenir les mesures de l'autre système. La lecture n'en reste pas moins relativement délicate, mais sans doute en grande partie du fait que la notion que le graphique cherche à présenter n'est en elle-même déjà pas très simple à appréhender. Des améliorations graphiques pourraient sans doute être explorées, notamment pour rendre plus lisibles les situations de superposition des points. Mais la complexité du sujet du graphique ne peut être totalement résolue uniquement par des stratégies d'améliorations graphiques.

D'autres solutions alternatives auraient pu être envisagées : représenter l'étendue de l'écart entre la saison où la différence entre les deux systèmes est la plus faible et celle où elle est la plus forte par une barre, par exemple.

De plus les ponctuels sont gros ici, certains se chevauchent et se masquent, et la valeur que l'on y associe est de ce fait assez imprécise. Par ailleurs, le fait d'associer deux variables (forme et couleur) n'aide pas forcément à la lecture : des points plus petits, de forme identique, avec seulement une variation de couleur auraient peut-être été plus adaptés pour signifier les écarts de mesure par saison.

I Figure 58 I

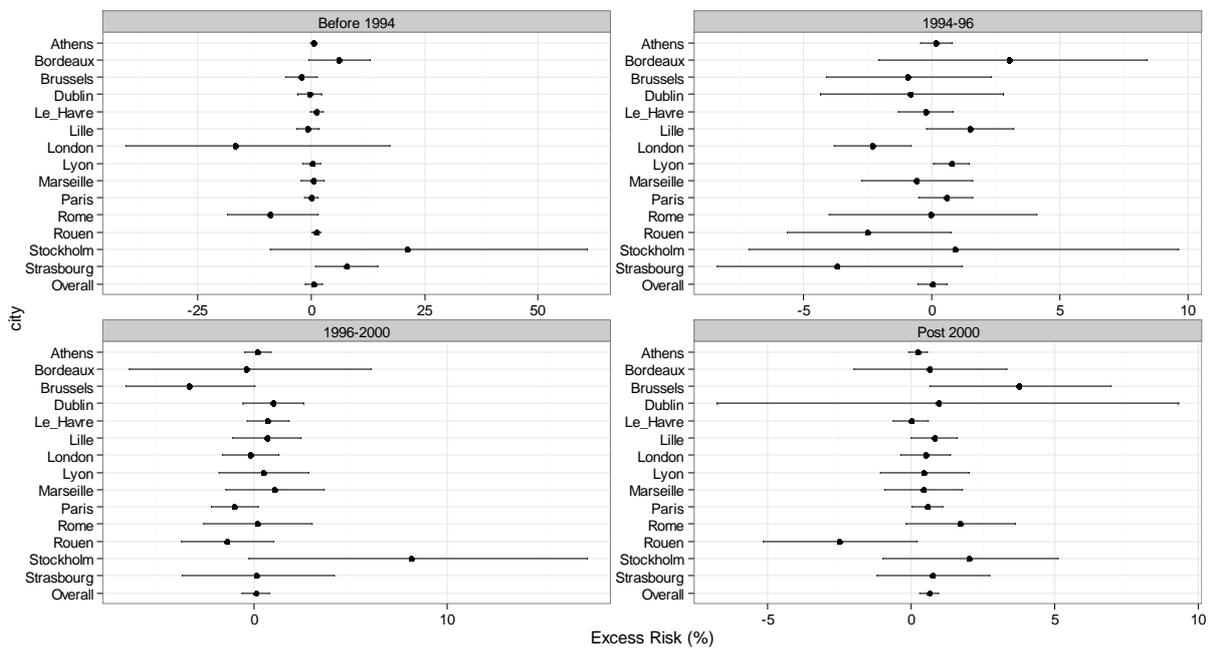
Graphique extrait du poster "Influence of the PM<sub>10</sub> measurement techniques on the assessment of the short-term mortality impacts of PM<sub>10</sub>", Mathilde Pascal, ISEE, 2013



Enfin, il manque une information essentielle, l'incertitude autour de ces points. Selon le but recherché, on aurait pu montrer les variations entre villes ou les variations entre saisons, en réalisant quatre graphiques avec leur estimateur central et IC95 % par période (figure 59), ou un seul et unique graphique présentant les 4 estimateurs centraux avec leurs IC95 %, mais légèrement décalés afin qu'ils ne se chevauchent pas, comme ci-après (figure 60).

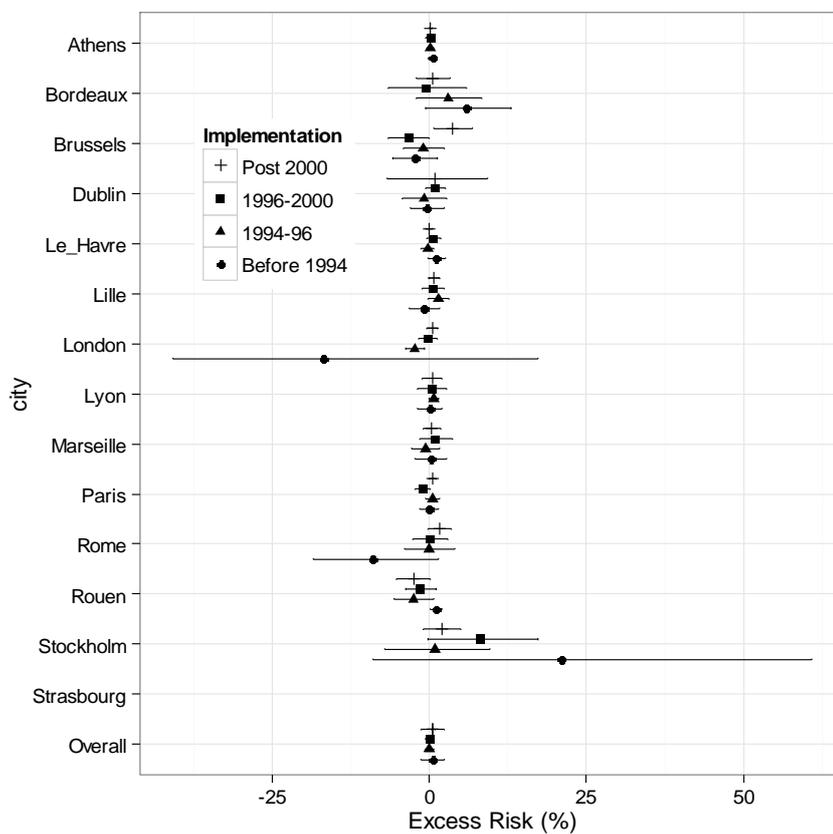
I Figure 59 I

Représentation sur 4 graphiques



I Figure 60 I

Alternative à la proposition d'une représentation en 4 graphiques



### 3.1.3 La représentation graphique des résultats dans le champ de l'évaluation quantitative des risques sanitaires (EQRS)

Les exemples présentés ci-après sont extraits de divers travaux d'EQRS menés dans le cadre de sollicitations récentes traitées au DSE, ou avec une participation importante du DSE sur les aspects méthodologiques notamment.

Présenter des résultats complexes à un public de décideurs et à la population générale malgré la complexité des concepts et des méthodes utilisés : exemple de la Sollicitation concernant Chenôve – Rapport « Évaluation des risques sanitaires des composés du traitement du bois » [InVS, Anses, 2009]

Lors de la détection en 2007 de pesticides dans l'air intérieur d'une école de Chenôve (Côte-d'Or), une alerte est lancée par l'Agence de sécurité sanitaire, de l'alimentation, de l'environnement et du travail (Anses) ; la fermeture de l'école est décidée en septembre ainsi que le déplacement des élèves dans trois autres établissements. Un comité scientifique et technique (CST) et une Commission locale d'orientation et d'information (CLOI) sont mis en place, puis maintenus jusqu'en juin 2008. Un rapport a été rédigé par le CST dont la présidence était assurée par l'InVS et le secrétariat scientifique par l'Anses. Les résultats de l'EQRS ont été rendus en juin 2008 et le rapport final a été publié en avril 2009. L'objectif du rapport était de présenter de manière exhaustive les résultats de l'étude aux décideurs et à la population générale, c'est-à-dire un public non spécialiste des EQRS.

Les données mises en œuvre sont issues d'une campagne métrologique, d'une recherche bibliographique pour la toxicologie et d'avis d'experts. Les scénarii de l'EQRS ont été décidés en concertation avec les représentants de la population. Les résultats présentés ci-après concernent les indicateurs de risques, ici les quotients de danger (QD). Ils sont formalisés dans une série de tableaux, commentés, suivis d'une rapide synthèse en encadré sans doute plus abordable pour un public non averti : cf. ci-après un exemple d'extrait de tableau présenté, ainsi que l'encadré de synthèse des résultats.

Ce premier tableau (tableau 1) présente les QD moyens pour les différentes catégories de personnes exposées (enfants en fonction de l'âge, et personnels selon le métier).

## I Tableau 1 I

### QD moyens pour les différentes catégories de personnes exposées à Chenôve

Composés	Quotient de danger	3 ans	4 ans	5 ans	6 ans	7 ans	8 ans	9 ans	10 ans	11 ans	12 ans	enseignant maternelle	enseignant élémentaire	ATSEM	agent d'entretien maternelle	agent d'entretien élémentaire
<b>α-HCH</b>	Quotient de danger total	1,9E-03	1,7E-03	1,5E-03	3,7E-03	2,9E-03	2,6E-03	2,3E-03	2,2E-03	2,0E-03	1,8E-03	6,9E-04	1,4E-03	7,9E-04	6,0E-04	1,1E-03
	Quotient de danger école	1,5E-03	1,3E-03	1,2E-03	3,3E-03	2,5E-03	2,2E-03	2,0E-03	1,8E-03	1,6E-03	1,5E-03	3,5E-04	1,1E-03	4,5E-04	2,6E-04	7,8E-04
<b>β-HCH</b>	Quotient de danger total	1,3E-02	1,2E-02	1,1E-02	2,3E-02	1,8E-02	1,6E-02	1,5E-02	1,4E-02	1,3E-02	1,2E-02	6,1E-03	1,0E-02	6,6E-03	5,7E-03	8,5E-03
	Quotient de danger école	8,1E-03	7,1E-03	6,2E-03	1,9E-02	1,3E-02	1,2E-02	1,1E-02	9,7E-03	8,4E-03	7,5E-03	1,6E-03	5,6E-03	2,1E-03	1,2E-03	4,0E-03
<b>γ-HCH</b>	Quotient de danger total	<b>1,1</b>	<b>1,0</b>	0,9	<b>1,6</b>	<b>1,2</b>	<b>1,1</b>	<b>1,0</b>	<b>1,0</b>	0,9	0,8	0,5	0,7	0,5	0,4	0,6
	Quotient de danger école	0,6	0,6	0,5	<b>1,2</b>	0,8	0,7	0,6	0,6	0,5	0,5	0,1	0,4	0,2	0,1	0,3
<b>tri-tetrachlorophénol</b>	Quotient de danger total	1,1E-04	9,3E-05	8,2E-05	5,0E-05	4,0E-05	3,5E-05	3,2E-05	2,9E-05	2,7E-05	2,4E-05	2,5E-05	1,6E-05	3,2E-05	1,9E-05	1,2E-05
	Quotient de danger école	1,1E-04	9,2E-05	8,1E-05	4,9E-05	3,9E-05	3,4E-05	3,1E-05	2,9E-05	2,6E-05	2,3E-05	2,5E-05	1,6E-05	3,2E-05	1,9E-05	1,1E-05
<b>pentachlorophénol</b>	Quotient de danger total	2,0E-02	2,0E-02	1,9E-02	1,8E-02	1,7E-02	1,7E-02	1,7E-02	1,7E-02	1,6E-02	1,6E-02	1,5E-02	1,5E-02	1,6E-02	1,5E-02	1,5E-02
	Quotient de danger école	2,6E-03	2,3E-03	2,0E-03	1,5E-03	1,2E-03	1,1E-03	9,8E-04	9,0E-04	7,9E-04	7,1E-04	5,7E-04	4,6E-04	7,2E-04	4,2E-04	3,3E-04
<b>aldrine</b>	Quotient de danger total	2,5E-02	2,2E-02	1,9E-02	2,5E-02	2,0E-02	1,7E-02	1,5E-02	1,4E-02	1,3E-02	1,2E-02	5,0E-03	7,4E-03	5,7E-03	4,3E-03	6,0E-03
	Quotient de danger école	1,7E-02	1,4E-02	1,3E-02	1,9E-02	1,4E-02	1,3E-02	1,1E-02	1,1E-02	9,4E-03	8,4E-03	3,1E-03	5,5E-03	3,9E-03	2,3E-03	4,0E-03
<b>dieldrine</b>	Quotient de danger total	2,2E-01	2,2E-01	2,1E-01	2,2E-01	2,2E-01	2,1E-01	2,1E-01	2,1E-01	2,1E-01	2,1E-01	2,0E-01	2,1E-01	2,0E-01	2,0E-01	2,1E-01
	Quotient de danger école	1,3E-02	1,1E-02	9,7E-03	1,8E-02	1,4E-02	1,2E-02	1,1E-02	1,0E-02	9,1E-03	8,1E-03	2,1E-03	5,4E-03	2,7E-03	1,6E-03	3,9E-03
<b>dioxines et furanes</b>	Quotient de danger total	6,1E-01	5,8E-01	5,5E-01	5,6E-01	5,0E-01	4,8E-01	4,7E-01	4,6E-01	4,5E-01	4,4E-01	4,0E-01	4,1E-01	4,0E-01	3,9E-01	4,0E-01
	Quotient de danger école	1,4E-01	1,2E-01	1,1E-01	1,3E-01	7,8E-02	6,9E-02	6,2E-02	5,7E-02	5,0E-02	4,5E-02	2,3E-02	3,3E-02	2,9E-02	1,7E-02	2,4E-02
<b>α-HCH+β-HCH+tri-tetrachlorophénols+aldrine+dieldrine</b>	Quotient de danger total	2,6E-01	2,5E-01	2,4E-01	2,7E-01	2,6E-01	2,5E-01	2,5E-01	2,4E-01	2,4E-01	2,4E-01	2,2E-01	2,3E-01	2,2E-01	2,1E-01	2,2E-01
	Quotient de danger école	3,9E-02	3,4E-02	3,0E-02	5,9E-02	4,4E-02	3,9E-02	3,5E-02	3,2E-02	2,9E-02	2,5E-02	7,2E-03	1,8E-02	9,2E-03	5,3E-03	1,3E-02

Les discussions méthodologiques associées à la présentation des résultats font également l'objet de rapides synthèses encadrées (exemple, ci-dessous, d'une synthèse expliquant le faible apport de la médiane par rapport à la moyenne pour ces résultats) :

« Finalement, le choix du paramètre statistique « médiane » au lieu du paramètre « moyenne » pour traiter les résultats de la campagne de mesures a peu d'influence sur les résultats de la caractérisation des risques sanitaires. »

Enfin, l'ensemble des conclusions sur ces résultats font l'objet d'une courte synthèse elle aussi figurée par un encadré dans le texte du rapport :

« Finalement, seul l' $\alpha$ -HCH a présenté, dans toutes les campagnes de mesures à l'intérieur du groupe scolaire des valeurs supérieures à celles retrouvées dans la littérature (Bouvier, 2005).

Mais, même avec ces valeurs, les risques sanitaires liés à l'exposition à l' $\alpha$ -HCH ont été écartés.

Pour le  $\gamma$ -HCH, les concentrations mesurées sont au niveau du 95<sup>e</sup> percentile de la distribution des mesures dans la seule autre étude française disponible portant sur des logements (Bouvier, 2005), mais elles restent cependant toujours incluses dans l'intervalle (minimum-maximum) de ces données. »

Les résultats des campagnes de mesures effectuées dans l'air du groupe scolaire et dans d'autres milieux intérieurs référencés sont quant à eux présentés sous forme d'un tableau (tableau 2).

## I Tableau 2 I

### Comparaison des concentrations mesurées dans l'air du groupe scolaire avec des valeurs de référence

Afssset • InVS RAPPORT « Chenôve »

**Tableau 47 : Comparaison des concentrations dans l'air du groupe scolaire des Bourdenières dans différentes campagnes et dans d'autres milieux intérieurs référencés (médiane [minimum-maximum])**

Concentrations dans l'air intérieur (ng.m <sup>-3</sup> )		$\alpha$ -HCH	$\beta$ -HCH	$\gamma$ -HCH	tri- et tetra-chlorophénols	penta-chlorophénol	aldrine	dieldrine	dioxines et furanes	
Mesures réalisées à Chenôve dans l'école	2008	Maternelle	42,7 [23,4-60,8]	1,3 [0,4-2,8]	10,7 [5,6-14,5]	6,3 [4,8-34,9]	1,8 [0,4-7,9]	0,8 [0,8-2,5]	0,8 [0,8-2,5]	0,5.10 <sup>-4</sup> [0,05.10 <sup>-4</sup> -0,5.10 <sup>-4</sup> ]
		Elémentaire	75,9 [30-92,5]	1,6 [0,8-2,5]	27,3 [14,4-49,5]	4 [3,6-11,2]	1,3 [0,5-3,4]	1,3 [0,4-3,7]	1,3 [0,8-6,2]	2,3.10 <sup>-4</sup> [0,9.10 <sup>-4</sup> -3,8.10 <sup>-4</sup> ]
	2007	Elémentaire	222,9 [185,7-316,1]	8,7 [5,6-12,8]	97,4 [65,8-130,8]	-	-	0,1 [0,1-0,8]	2,2 [1,1-3,2]	-
	2006	Elémentaire 1 mesure ponctuelle	38,3	20,2	7,7	-	-	-	2,9	-
Données de la littérature	Valeur	0,2 [<LQ-5,8]	-	2,9 [<LQ-209,3]	-	1,9 [<0,15-6,1]	<LQ [<LQ-5,4]	<LQ [<LQ-9,7]	4,2.10 <sup>-4</sup> [0,1.10 <sup>-4</sup> -17,4.10 <sup>-4</sup> ]	
	Source de données	France : Bouvier, 2005	-	France : Bouvier, 2005	-	France : Desmetres P., 2006	France : Bouvier, 2005	France : Bouvier, 2005	Von Manikowsky S. et al., 1998	

- : pas de donnée

En choisissant une présentation des éléments à retenir par encadrés aisément identifiables dans le corps du texte il devient plus facile pour un public non averti de repérer l'information directement abordable à son niveau de connaissance. Pour l'EQRS sans doute encore plus que pour d'autres travaux du DSE, la question du public se pose tout particulièrement dans la mesure où les méthodes et les concepts manipulés demeurent assez complexes. Ici, le tableau a ceci de spécifique qu'il confronte les données mesurées et les données de la littérature, ce qui permet de fournir au lecteur

des valeurs référencées avec lesquelles envisager des comparaisons (données de la littérature française, européenne et, à défaut des autres, de la littérature étrangère).

Présenter les résultats d'une EQRS au décideur : les résultats du rapport : avis de la Cellule de l'InVS en région (Cire) sur l'étude d'HPC Envirotec pour AFTRP concernant les jardins familiaux du Fort d'Aubervilliers

La commune d'Aubervilliers (93) dispose depuis 1907, attenants au Fort d'Aubervilliers, de jardins familiaux (183 parcelles) couvrant une surface d'environ 7,2 hectares. Ces jardins sont très largement exploités comme potagers pour la production de fruits et de légumes destinés à l'autoconsommation. Depuis le début des années 1990, et du fait des anciennes activités militaires qui se sont exercées sur ces terrains, diverses investigations avaient fait apparaître une contamination des sols par des hydrocarbures, des éléments de traces métalliques (ETM) et des hydrocarbures aromatiques polycycliques (HAP). C'est dans ce contexte qu'à la fin des années 2000 la Cire Ile-de-France est sollicitée pour la réalisation d'une EQRS et la rédaction de recommandations concernant l'usage de ces jardins. Au vu des connaissances sur le sujet, ce sont les risques liés à une exposition à faible dose mais sur des temps longs qui ont été analysés et non les risques d'intoxication aiguë. Le DSE a été sollicité pour apporter son appui méthodologique sur ce dossier. Les résultats de cette évaluation des risques ont fait l'objet, d'une part, d'un rapport assez détaillé contenant un avis destiné aux décideurs, et d'une présentation en réunion publique, donc destinée à un public *a priori* totalement profane dans le domaine des EQRS.

Il est intéressant de noter que le rapport, destiné aux décideurs, s'ouvre sur un résumé de l'avis dans lequel les éléments importants sont retenus et présentés de façon très accessible. Les choix qui ont été faits visent donc clairement à optimiser l'efficacité en termes de communication d'un rapport qui, plus loin dans le corps du texte, rentre davantage dans le détail des méthodes et des concepts plus complexes de l'EQRS.

Ainsi, le rapport détaille de nombreuses informations concernant le risque d'exposition par ingestion des enfants qui consomment les légumes et les fruits du potager. Le chapitre 5.4 du rapport y est entièrement consacré. La distribution du QD selon divers scénarii dans un histogramme y est présentée, ainsi qu'une analyse de sensibilité sur différentes variables.

Dans un premier temps, les résultats des analyses dans les sols et les végétaux sont rapportés dans des tableaux (tableau 3).

En comparant ces doses aux valeurs toxicologiques de référence (VTR) disponibles auprès de divers organismes (OMS en particulier), on peut qualifier le risque.

I Tableau 3 I

Présentation détaillée des résultats pour le plomb, le cuivre et le cadmium

	légumes feuilles			légumes fruits				légumes racines					pommes de terre	sol
	salades	choux	blettes	tomates	courgettes	aubergines	citrouilles	poireaux	carottes	navets	radis	betteraves		
<b>nombre de prélèvements (URS +HPC Envirotec)</b>	1	5	3	6	4	1	1	4	3	3	1	2	3	11 ou 6 selon composé
somme	9			12				13						
parcelle	HPC (74)	URS (2/16/20 /22/74)	HPC (34/48/ 160)	HPC (30/33/34/4 8/74/160)	HPC (30/33/34/ 48)	HPC (74)	HPC 160	URS (2/16/20/ 74)	URS (16/20) et HPC 30	URS (20/22/74)	HPC 33	HPC (74/160)	HPC (30/34/48)	URS et HPC
<b>plomb (mg/kg)</b>														
résultats par légume ou sol	1,2	<0,05	0,63	<0,05	<0,05	<0,05	<0,05	0,1	0,12	0,10	0,07	0,70	0,13	258,18
étendue	-	-	0,07 à 1,5	-	-	-	-	<0,05 à 0,25	0,07 à 0,17	<0,05 à 0,21	-	0,29 à 1,1	0,05 à 0,22	130 à 540
maximum par légume ou sol	1,2	<0,05	1,5	<0,05	<0,05	<0,05	<0,05	0,25	0,17	0,21	0,07	1,1	0,22	540,00
moyenne pour groupe	0,37			<0,05				0,20						
maximum pour le groupe			1,5									1,1		
<b>cadmium (mg/kg)</b>														
résultats par légume ou sol	0,02	0,01	0,04	0,01	<0,01	0,01	<0,01	0,01	0,67	0,01	0,03	0,05	0,07	1,08
étendue	-	-	<0,01 à 0,08	<0,01 à 0,02	-	-	-	0,01 à 0,02	<0,01 à 0,02	<0,01 à 0,02	-	0,01 à 0,08	0,01 à 0,1	0,4 à 2,4
maximum par légume ou sol	0,02	0,01	0,08	0,02	<0,01	0,01	<0,01	0,01	0,67	0,02	0,03	0,08	0,10	2,40
moyenne pour groupe	0,02			0,01				0,02						
maximum pour le groupe			0,08	0,02								0,08		

	légumes feuilles			légumes fruits				légumes racines					pommes de terre	sol
	salades	choux	blettes	tomates	courgettes	aubergines	citrouilles	poireaux	carottes	navets	radis	betteraves		
<b>nombre de prélèvements (URS +HPC Envirotec)</b>	1	5	3	6	4	1	1	4	3	3	1	2	3	11 ou 6 selon composé
somme	9			12				13						
parcelle	HPC (74)	URS (2/16/20 /22/74)	HPC (34/48/ 160)	HPC (30/33/34/4 8/74/160)	HPC (30/33/34/ 48)	HPC (74)	HPC 160	URS (2/16/20/ 74)	URS (16/20) et HPC 30	URS (20/22/74)	HPC 33	HPC (74/160)	HPC (30/34/48)	URS et HPC
<b>cuivre (mg/kg)</b>	remarque : URS n'a pas analysé cet élément dans sa campagne d'où l'absence de résultats pour certains légumes													
résultats par légume ou sol	1,20	-	1,45	0,82	1,05	0,90	0,65	-	0,17	-	1,90	2,95	4,33	105,83
étendue	-	-	0,65 à 2,1	0,56 à 1,2	0,59 à 1,5	-	-	-	-	-	-	2,9 à 3	1,8 à 8,7	78 à 140
maximum par légume ou sol	1,20	-	2,10	1,20	1,50	0,90	0,65		0,17		1,90		8,70	140,00
moyenne pour groupe	1,29			0,77				1,99						
maximum pour le groupe			2,10		1,50							3,00		

Ces données permettent d'estimer par la suite la dose reçue par les enfants qui consomment les produits du potager.

Revenons sur le tableau des résultats d'analyse : il est assez long et fastidieux à lire. De ce fait, une proposition de sélection et de synthèse des informations essentielles à la compréhension est fournie ci-après (tableau 4).

Les valeurs mesurées sont reprises du tableau précédent, mais au lieu d'avoir pour chaque polluant une ligne pour la valeur moyenne, une ligne pour le minimum, une ligne pour le maximum, l'information a été présentée autrement pour être davantage groupée. Par ailleurs, les différents légumes sur lesquels des mesures ont été réalisées sont regroupés et présentés sous une forme hiérarchisée. Enfin, les valeurs réglementaires sont mises en avant graphiquement par une écriture différente, ce qui permet une comparaison plus immédiate avec les valeurs mesurées.

I Tableau 4 I

Proposition de simplification pour le tableau 3

	Nb de prélèvements	Parcelles	Données mesurées (par rapport aux valeurs réglementaires)			Concentrations attendues (selon la base BAPPET)	
			Pb (mg/kg) Moyenne (min : max)	Cd (mg/kg) Moyenne (min : max)	Cu (mg/kg) Moyenne (min : max)	Pb (mg/kg) En matière sèche	Cd (mg/kg) En matière sèche
<b>Légumes feuilles</b>	9	URS (2/16/20/22/74) HPC (74/34/48/160)	<b>0,4</b> (<0,05 : 1,5)	<b>0,02</b> (<0,01 : 0,08)	<b>1,3</b> (0,7 : 2,1)		
<i>dont</i> <i>Valeur réglementaire</i>			<i>0,30</i>	<i>0,20</i>			
<b>Salade</b>	1	HPC (74)	<b>1,2</b>	<b>0,02</b>	<b>1,2</b>	<b>16,7</b>	<b>0,49</b>
<b>Choux</b>	5	URS (2/16/20/22/74)	<b>&lt;0,05</b>	<b>0,01</b>	<b>-</b>	<b>4,9</b>	<b>-</b>
<b>Blettes</b>	3	HPC (34/48/160)	<b>0,6</b> (0,1 : 1,5)	<b>0,04</b> (<0,01 : 0,08)	<b>1,5</b> (0,7 : 2,1)	<b>12,7</b>	<b>0,8</b>
<b>Légumes fruits</b>	12	HPC (30/34/48/74/160)	<b>&lt;0,05</b>	<b>0,01</b> (<0,01 : 0,02)	<b>0,8</b> (0,6 : 1,5)		
<i>dont</i> <i>Valeur réglementaire</i>			<i>0,10</i>	<i>0,05</i>			
<b>Tomates</b>	6	HPC (30/33/34/48/74/160)	<b>&lt;0,05</b>	<b>0,01</b> (<0,01 : 0,02)	<b>0,8</b> (0,6 : 1,2)	<b>-</b>	<b>0,4</b>
<b>Courgettes</b>	4	HPC (30/33/34/48)	<b>&lt;0,05</b>	<b>&lt;0,01</b>	<b>1,1</b> (0,6 : 1,5)	<b>-</b>	<b>-</b>
<b>Aubergines</b>	1	HPC(74)	<b>&lt;0,05</b>	<b>0,01</b>	<b>0,90</b>	<b>-</b>	<b>0,6</b>
<b>Citrouilles</b>	1	HPC(160)	<b>&lt;0,05</b>	<b>&lt;0,01</b>	<b>0,7</b>	<b>-</b>	<b>-</b>
<b>Légumes racines</b>	13	HPC(30/33/74/160) URS(16/20/22/74)	<b>0,2</b> (<0,05 : 1,1)	<b>0,02</b> (<0,01 : 0,08)	<b>2,0</b> (0,2 : 3,0)		
<i>dont</i> <i>Valeur réglementaire</i>			<i>0,10</i>	<i>0,10</i>			
<b>Poireaux</b>	4	URS (2/16/20/74)	<b>0,1</b> (<0,05 : 0,2)	<b>0,01</b> (0,01 : 0,02)	<b>-</b>	<b>-</b>	<b>-</b>
<b>Carottes</b>	3	URS(16/20) HPC(30)	<b>0,1</b> (0,1 : 0,2)	<b>0,67</b> (<0,01 : 0,67)	<b>0,2</b>	<b>0,7</b>	<b>0,155</b>
<b>Navets</b>	3	URS(20/22/74)	<b>0,1</b> (<0,05 : 0,2)	<b>0,01</b> (<0,01 : 0,02)	<b>-</b>	<b>-</b>	<b>-</b>
<b>Radis</b>	1	HPC(33)	<b>0,1</b>	<b>0,03</b>	<b>1,9</b>	<b>-</b>	<b>0,127</b>
<b>Betteraves</b>	2	HPC(74/160)	<b>0,7</b> (0,3 : 1,1)	<b>0,05</b> (0,01 : 0,08)	<b>3,0</b> (2,9 : 3,0)	<b>2</b>	<b>-</b>
<b>Pommes de terre</b>	3	HPC(30/34/48)	<b>0,1</b> (0,05 : 0,2)	<b>0,07</b> (0,01 : 0,1)	<b>4,3</b> (1,8 : 8,7)	<b>0,2</b>	<b>0,159</b>
<i>Valeur réglementaire</i>			<i>0,10</i>	<i>0,10</i>			
<b>Sols</b>	11 ou 6 selon le composé	HPC URS	<b>258</b> (130 : 540)	<b>1,2</b> (0,4 : 2,4)	<b>151</b> (78 : 140)	<b>259,5 : 273</b>	<b>1,16 : 1,33</b>

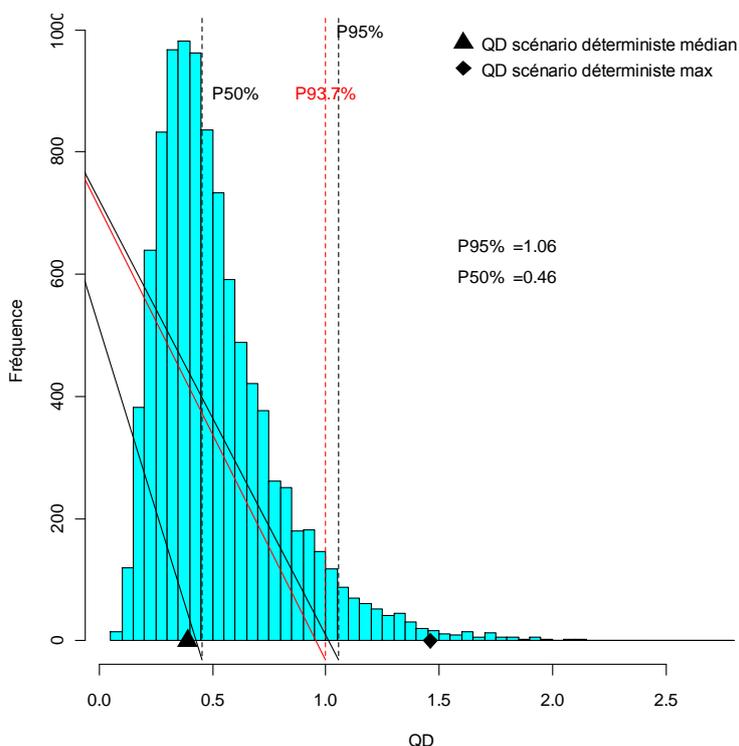
Les données des valeurs réglementaires<sup>20</sup> sont fournies en matière fraîche pour les aliments, en masse sèche pour les sols ; les données de BAPPET<sup>21</sup> sont fournies sur la matière sèche.

Les données BAPPET sont sélectionnées pour des teneurs des sols proches de 250 mg/kg de sol pour le plomb (valeurs fournies pour des teneurs allant de 259,5 à 273 mg/kg de sol), et de 1,2 mg/kg de sol pour le cadmium (données fournies pour des teneurs allant de 1,16 à 1,33 mg/kg de sol).

L'histogramme proposé ici (figure 61) présente les résultats de distribution des QD. La gamme d'étendue des valeurs des QD s'étale de 0,06 à 2,78 ; la probabilité que le QD soit supérieur à 1 est de 6,3 % ; et donc celui d'être inférieur ou égal à 1 est de 93,7 %. Il propose en fait d'affiner le résultat de l'EQRS en essayant d'utiliser une distribution des valeurs possibles pour chacune des variables à renseigner dans le calcul de risque. Le résultat n'est donc pas un QD unique mais une distribution des QD à laquelle une probabilité d'occurrence est associée.

I Figure 61 I

### Distribution des QD pour les enfants par ingestion de fruits et légumes autoproduits



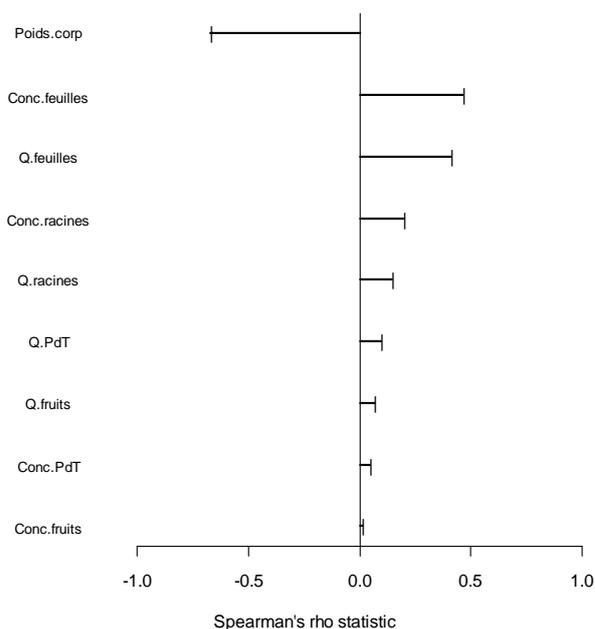
<sup>20</sup> Source des valeurs réglementaires pour l'Union Européenne : [http://www.assemblee-nationale.fr/europe/pdf/doc\\_e/e5966.pdf](http://www.assemblee-nationale.fr/europe/pdf/doc_e/e5966.pdf)

<sup>21</sup> BAPPET : Base de données sur les teneurs en éléments traces métalliques de plantes potagères : <http://www.developpement-durable.gouv.fr/BAPPET-BAse-de-donnees-sur-les.html>

Sur cet histogramme, ce sont les valeurs situées au-delà des valeurs de référence qui font l'objet d'une attention particulière. L'analyse de sensibilité est présentée dans le diagramme de Tornado<sup>22</sup> (figure 62) et définit les paramètres sur lesquels il est éventuellement possible d'agir pour diminuer le risque, à savoir un faible poids corporel, la quantité de légumes feuilles consommée, et la concentration de plomb dans les légumes feuilles. Le résultat statistique représenté est le coefficient de corrélation de Spearman (ou rho de Spearman) qui étudie les corrélations entre les rangs des valeurs prises par les deux variables dont on étudie la corrélation (et non pas entre les valeurs elles-mêmes). Cette corrélation est utilisée lorsque les distributions des variables sont asymétriques.

I Figure 62 I

Résultat de l'analyse de sensibilité (diagramme de Tornado) permettant d'identifier les paramètres les plus sensibles



Ce diagramme permet d'identifier les paramètres les plus sensibles : la masse corporelle (plus la masse corporelle d'un individu est faible, plus le QD est élevé) ; les concentrations de plomb dans les légumes feuilles ; les quantités de légumes feuilles consommées.

Les analyses présentées ici pour l'exemple des potagers d'Aubervilliers amenaient, lors de la publication du rapport, à conclure sur l'absence d'effets sanitaires attendus via l'exposition des enfants au cadmium et au cuivre, et à un risque peu élevé, mais présent pour l'exposition liée au plomb (dans le cas d'une exposition maximaliste, le risque de dépassement du QD existe, mais reste peu élevé puisqu'inférieur à 10 %). Des recommandations ont pu être formulées (limiter la consommation des légumes feuilles autoproduits pour les jeunes enfants, laver minutieusement ces légumes afin de limiter au maximum la présence de particules de sols sur ces légumes).

<sup>22</sup> Un diagramme de Tornado est un graphique en barres dans lequel les données sont représentées par des barres verticales, les valeurs étant triées et les plus grandes étant représentées en haut du graphique, les plus petites en bas, ce qui lui donne l'apparence d'une tornade.

Préserver des résultats à un public de décideurs et à un public plus général le résultat d'un avis : exemple de la saisine concernant la présence de formamide dans des jouets en mousse : les tapis puzzle – juillet 2011. [12]

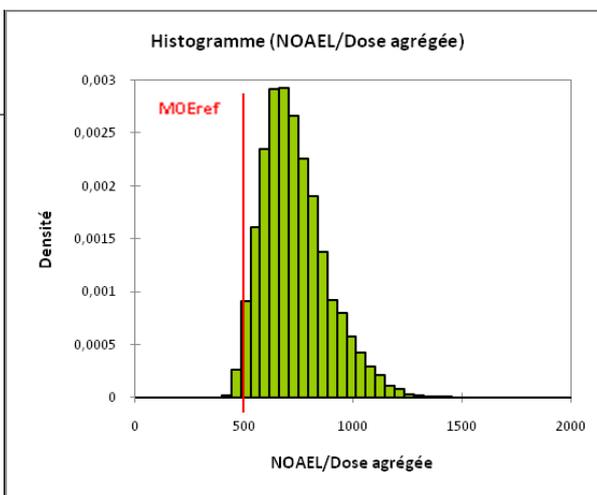
Cette saisine de l'Anses par la Direction générale de la concurrence, de la consommation et de la répression des fraudes (DGCCRF) en juillet 2011 fait suite à des tests réalisés en Belgique à partir de 2009 et mettant en évidence la présence dans les jouets « tapis puzzle » de formamide, une substance classée comme reprotoxique<sup>23</sup> avec un « Risque pendant la grossesse d'effets néfastes pour l'enfant ». Cette découverte a entraîné en France le retrait du marché de ces jouets pendant 3 mois à compter du 16 décembre 2010, durée censée permettre à la DGCCRF de poursuivre les prélèvements réalisés sur ces jouets commercialisés en France et, à l'Anses, de rendre un avis. Un avis et un rapport relatifs aux usages de formamide dans les produits de consommation et aux risques sanitaires liés au formamide dans les jouets en mousse « tapis puzzle » ont donc été conduits par l'Anses.

I Figure 63 I

Exemple de la présentation du risque sanitaire lié aux expositions long terme chez les enfants de 0 à 2 ans

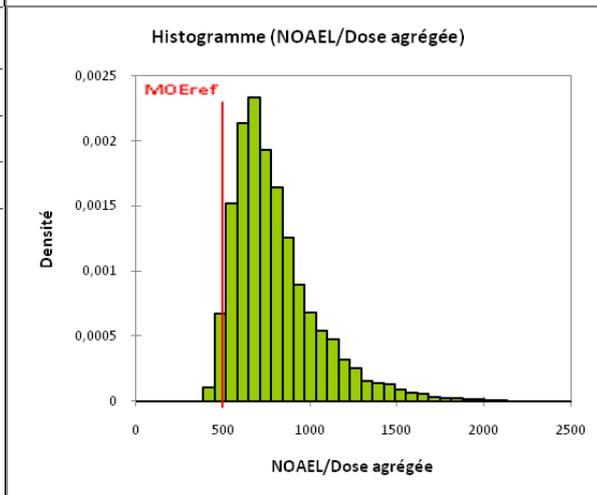
**Enfant de 0-1 an**

Observations	Moyenne	Écart-type
10000	734	149
Percentile	Valeur	
Maximum 100 %	1 438	
99 %	1 157	
95 %	1 019	
90 %	941	
3 <sup>e</sup> Quartile 75 %	819	
Médiane 50 %	712	
1 <sup>er</sup> Quartile 25 %	626	
10 %	562	
5 %	527	
1 %	481	
Minimum 0 %	407	



**Enfant de 1-2 ans**

Observations	Moyenne	Écart-type
10000	801	243
Percentile	Valeur	
Maximum 100 %	2 124	
99 %	1 614	
95 %	1 278	
90 %	1 128	
3 <sup>e</sup> Quartile 75 %	906	
Médiane 50 %	743	
1 <sup>er</sup> Quartile 25 %	631	
10 %	556	
5 %	518	
1 %	463	
Minimum 0 %	401	



<sup>23</sup> Classé en catégorie 1B par le règlement CLP 1272/2008, c'est-à-dire : reprotoxique de catégorie 2 au sens de la directive 67/548/CEE relative à la classification, à l'emballage et à l'étiquetage des substances et préparations dangereuses, avec la phrase de risque R61.

Cet avis est construit autour de la définition d'une marge of exposure (MoE) car il n'est pas possible ici de calculer un QD du fait de l'absence de VTR pour le polluant en question. On ne dispose *a priori* que des résultats d'études expérimentales toxicologiques (NOAEL pour No Observed Adverse Effect Level) pour le formamide qui pour une raison ou une autre (manque de connaissances pour une population donnée, manque de qualité des données ou d'intérêt pour cette substance, etc.) n'ont pas abouti à définir une VTR.

Par rapport à une VTR qui intègre des facteurs d'incertitudes en vue d'une application à la population générale, le NOAEL est un résultat brut d'analyse pour une espèce animale ; au final lorsqu'on calcule une marge of exposure, cela revient à obtenir un facteur qui peut être comparé aux facteurs d'incertitudes qu'on s'autoriserait lors de la construction d'une VTR. La moyenne figurée dans les tableaux ci-dessous est une moyenne NOAEL/dose agrégée, donc il n'y a pas d'unité puisque les deux grandeurs du rapport s'expriment dans la même unité.

Le reste de la démarche reste proche de celle d'une évaluation des risques (par détermination des niveaux d'exposition).

Dans l'absolu, pour interpréter le résultat, il faut se fixer alors une valeur en deçà de laquelle on considère que le facteur calculé est trop bas pour garantir la sécurité sanitaire de toute la population. Dans le graphique précédent (figure 63) cette valeur fixée est de 500. S'il existe moins d'un facteur 500 entre la NOAEL et la dose agrégée alors nous nous situons dans une zone où il est difficile de garantir une absence d'effets sanitaires.

Cette valeur représente une marge d'exposition minimale chez l'homme par rapport à une dose critique en condition expérimentale (*i.e.* chez l'animal). Cette marge minimale est jugée prédictive de l'absence d'effet néfaste chez l'homme si nous nous situons au-dessus (le facteur d'incertitude est suffisamment important entre la dose calculée chez l'homme et l'effet toxique observée chez l'animal pour être confiant sur l'absence d'un risque sanitaire). Elle correspond au produit des facteurs d'incertitude, qui eux, reflètent l'incertitude scientifique liée à la transposition inter-espèces, interindividuelle, ou à la transposition d'une situation d'exposition à une autre. Ils sont choisis après confrontation d'avis d'experts. Il est par ailleurs rappelé que la détermination d'une marge d'exposition de référence ne se substitue pas à la notion d'acceptabilité du risque, qui relève du gestionnaire du risque lui-même. C'est d'ailleurs ce qui conduit à la formulation de la conclusion qui suit.

Des résultats sous forme de graphiques sont donnés pour différentes catégories, en particulier les jeunes enfants qui sont les premiers utilisateurs de ce type de jouets, ainsi que pour les femmes enceintes puisque c'est sur une exposition pendant la grossesse de l'enfant à naître que le risque reprotoxique est clairement documenté.

Ce rapport de l'Anses constitue un premier rapport français présentant les résultats d'un avis en approche probabiliste. Par ailleurs, la présentation des résultats est relativement innovante. Les graphiques accompagnant les tableaux, par exemple, sont particulièrement novateurs. En effet, à la différence des graphiques de l'exemple d'Aubervilliers présenté précédemment, ces histogrammes décrivent une zone « à risque » située en deçà de la valeur seuil de 500, la valeur dite MOEref (pour « reference margin exposure »).

Ces graphiques et tableaux très détaillés – notamment pour les tableaux, la mention des très nombreuses valeurs de percentiles aurait éventuellement pu se limiter à 5 ou 6 valeurs – sont accompagnés d'une conclusion synthétique qui résume les résultats en des termes plus compréhensibles pour un public moins averti et peu rompu aux notions assez complexes et aux équations qui entrent en jeu dans une EQRS :

#### « Conclusion

Les résultats de l'évaluation réalisée à partir des concentrations d'émission (inhalation) et de migration (mise en bouche) mesurées sur 3 tapis puzzle montrent que les risques sanitaires ne peuvent être exclus chez des enfants de 0 à 2 ans exposés de manière chronique au formamide présent dans un tapis puzzle, ces risques présentant néanmoins une probabilité d'occurrence inférieure à 5 % . »

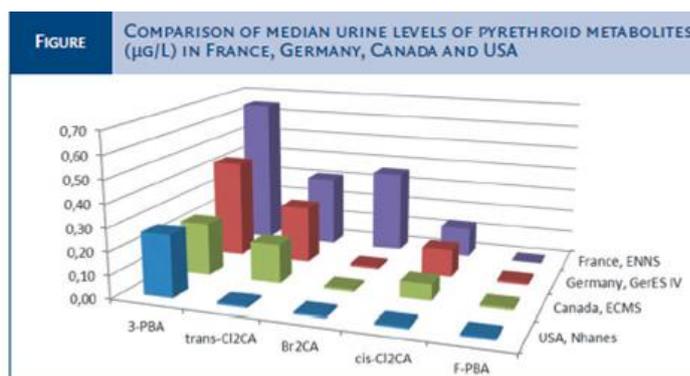
Si l'on s'en tenait aux tableaux et aux graphiques présentant les résultats de ce travail, on en resterait de manière assez évidente à des notions et des concepts qui ne sont pas partagés par tout un chacun, et ce d'autant qu'on se place ici dans un contexte d'approche méthodologique assez novatrice. Ces graphiques et ces tableaux sont le résultat de calculs complexes. Le texte d'accompagnement – et en particulier les conclusions synthétiques – est donc fondamental ici pour une bonne communication du message afférent aux résultats présentés.

### 3.1.4 Exemples de représentations graphiques dans le champ de la biosurveillance et des indicateurs de santé

Le premier exemple proposé dans le champ de la biosurveillance (figure 64) compare les niveaux d'imprégnation par différents métabolites urinaires de pesticides (pyréthrinoïdes) observés dans différents pays dans lesquels des données sont disponibles. Le graphique présente pour chaque pays les niveaux de concentration mesurés pour chaque métabolite analysé. La représentation adopte une visualisation en « 3D » difficile à lire, notamment pour les valeurs les plus faibles.

I Figure 64 I

Exemple peu lisible de représentation « 3D »

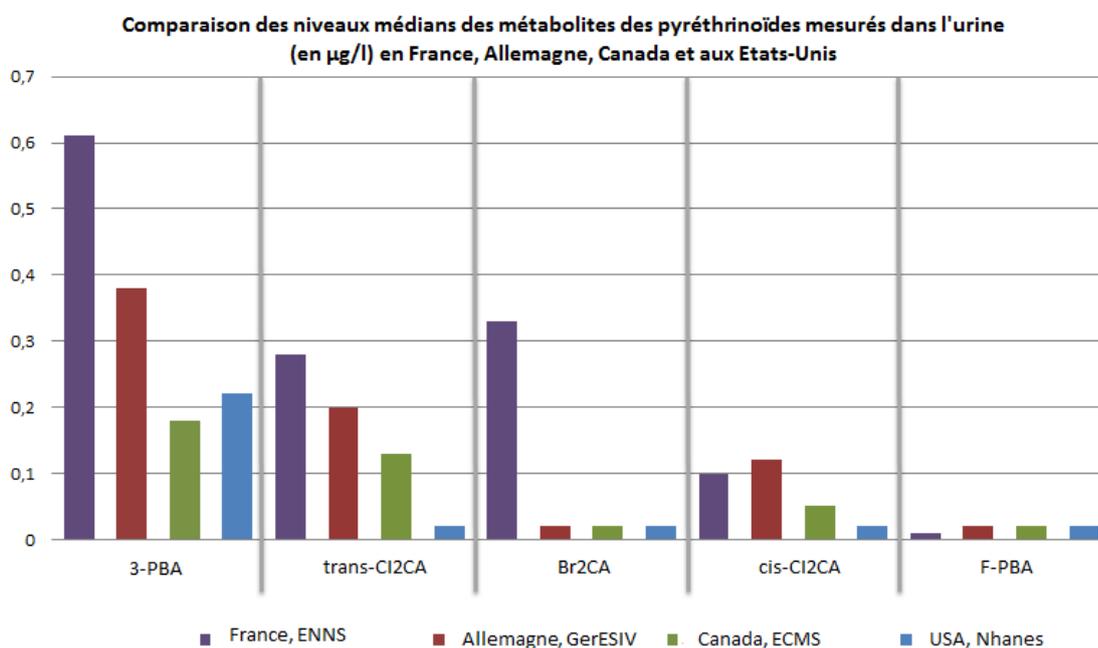


Une telle représentation permet à peu près de comparer visuellement les hauteurs des barres mais ne permet pas de quantifier de façon précise les résultats, en particulier pour les faibles niveaux. Ce graphique illustre la difficulté de représenter un nombre important d'informations dans un même visuel et le risque de perte de lisibilité que cela engendre.

La réalisation d'une collection de graphiques (un graphique par métabolite dosé) permettrait de s'affranchir de l'hétérogénéité des ordres de grandeur de mesure entre les métabolites, tout en garantissant l'objectif de comparaison des niveaux entre pays. Si la contrainte de tout faire figurer dans un même graphique est très forte (contraintes éditoriales des revues ou du support de publication par exemple) le choix peut être fait d'aligner l'ensemble des barres sur une seule et même représentation, en adaptant l'échelle des axes de manière à ne pas trop écraser les faibles niveaux, et en séparant chaque métabolite comme dans le graphique ci-après (figure 65).

## I Figure 65 I

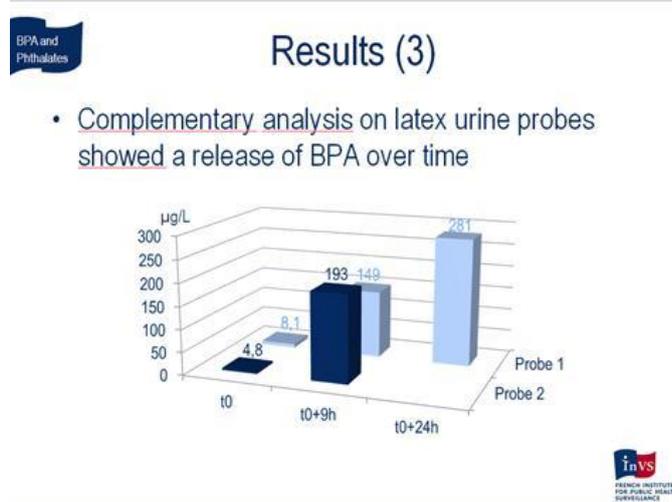
Proposition d'amélioration de la figure 64 par une représentation en deux dimensions



Un autre exemple de graphique « en 3D » présentant les mêmes difficultés de lecture que le précédent est présenté ci-dessous (figure 66). Il vise à montrer que, lors de la réalisation de l'étude pilote d'Elfe (Étude longitudinale française depuis l'enfance), la pose de sondes urinaires au moment de l'accouchement par césarienne conduisait à une contamination des échantillons d'urine recueillis, par relargage du Bisphénol-A (BPA). Dans la mesure où le nombre de valeurs représentées est plus faible que sur le graphique précédent, la solution pour le rendre plus lisible est similaire et simple à mettre en œuvre.

## I Figure 66 I

Autre exemple illustrant la difficulté de lecture liée aux représentations en « 3D »

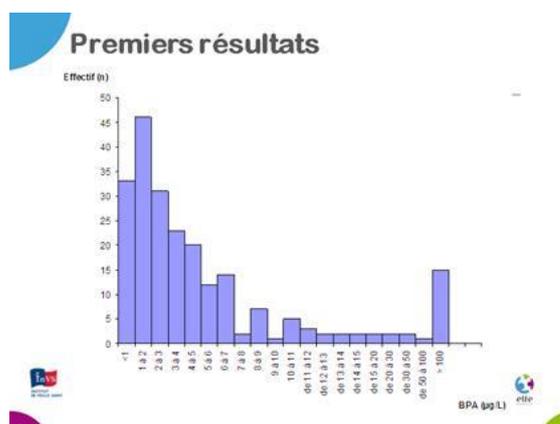


Le graphique suivant (figure 67) est un histogramme de la distribution des concentrations urinaires de BPA observées dans le cadre de l'étude pilote d'Elfe. Ce graphique a vraisemblablement été réalisé à l'aide d'Excel, et s'il ne pose pas de problème particulier de compréhension, il est néanmoins nécessaire de

rappeler ici les difficultés posées par Excel qui ne conserve pas la proportionnalité des surfaces dans la réalisation des histogrammes de distribution ce qui fausse le graphique lorsque l'amplitude des classes est différente, ce qui est le cas ici.

I Figure 67 I

Histogramme ne respectant pas la proportionnalité des surfaces du fait d'une amplitude très variable des classes figurant en abscisse

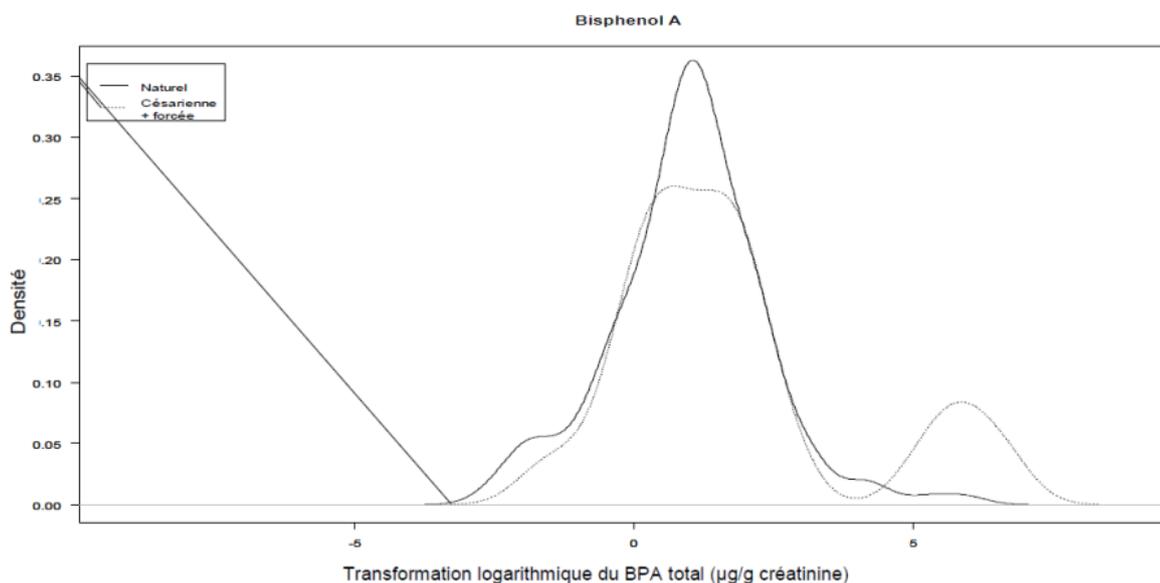


En effet, pour les faibles niveaux, les classes ont une amplitude de 1 µg/L, mais très vite, lorsque l'on atteint des niveaux plus élevés pour lesquels les effectifs sont plus faibles, les classes passent rapidement à 5, puis 10, 20, et 50 µg/L d'amplitude sans pour autant modifier la largeur des barres. De manière générale, il faut être très prudent avec l'utilisation d'Excel pour la réalisation des histogrammes de distribution, et lui préférer d'autres outils plus adaptés.

Enfin, un autre exemple est proposé ici (figure 68), illustrant les niveaux de BPA total en fonction du type d'accouchement (accouchement par voie basse *versus* accouchement par césarienne ou avec forceps). Ce graphique présente la difficulté de faire intervenir une transformation en logarithmique du niveau de BPA. On peut y lire que les faibles niveaux (se trouvant, une fois convertis en logarithme, autour de 0 et nettement inférieurs à 5) sont davantage représentés pour les accouchements par voie basse, alors que l'on observe un petit pic autour de 5 (valeur du logarithme de la concentration en BPA) et un peu au-delà pour les accouchements par césarienne ou forceps. Les représentations graphiques qui font appel à des conversions en logarithmes donnent lieu à des courbes que l'on peut aisément interpréter, mais dont les repères en termes de valeurs des données sont peu explicites. Ils s'adressent généralement à un public averti et rompu à la compréhension de la notion mathématique impliquée dans la transformation des valeurs. Par ailleurs, ce type de graphique est généralement utilisé pour démontrer une hypothèse, plus que pour lire explicitement des valeurs. L'hypothèse ici démontrée est que les femmes ayant accouché par césarienne ou par forceps avaient des concentrations en BPA supérieures aux femmes ayant accouché par voie basse, avec un pic de concentration démontrant l'existence d'une contamination au moment de l'accouchement.

## I Figure 68 I

Variation des niveaux de BPA total selon le type d'accouchement (avec transformation en log du niveau de BPA)

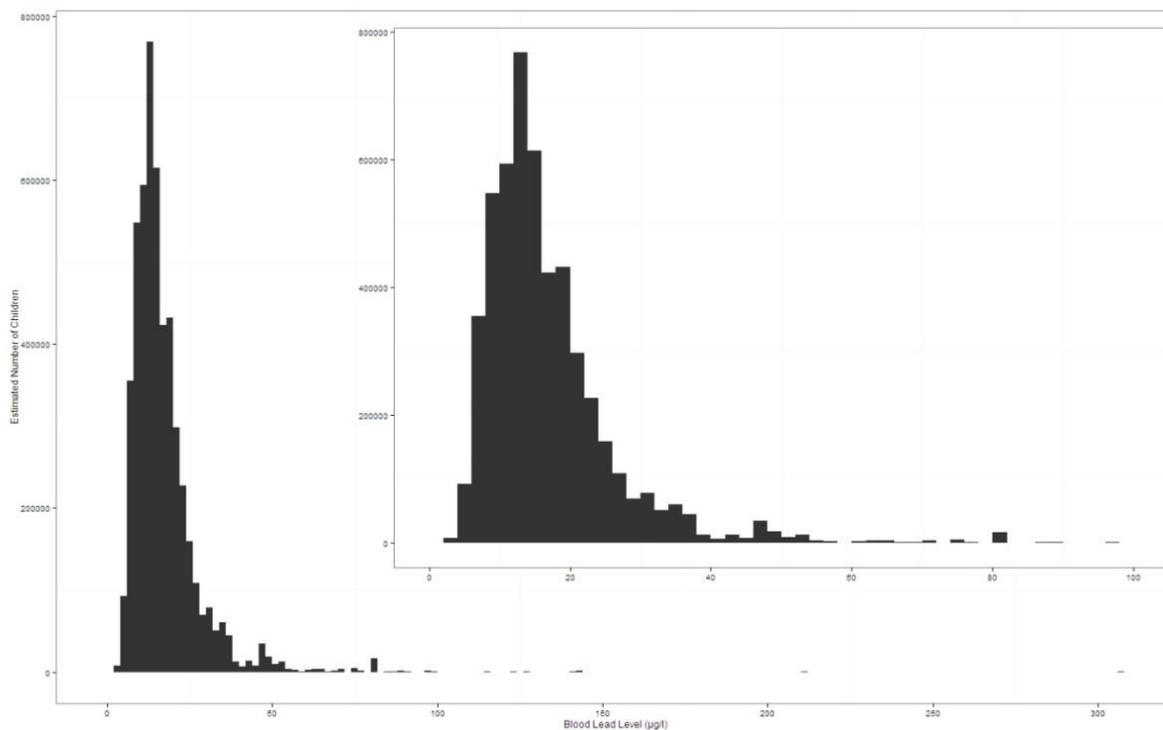


L'exemple proposé ci-après (figure 69) est un histogramme des distributions des niveaux de plomb dans le sang des enfants, avec « focus sur partie basse » : un « zoom » est fait sur la partie basse de la distribution, afin de mieux visualiser la répartition des effectifs sur les faibles niveaux. Les valeurs sont davantage détaillées sur l'axe des abscisses, mais les classes ne sont pas modifiées par rapport à l'histogramme qui représente l'intégralité de la distribution : l'histogramme est calculé sur l'intégralité de la distribution, mais les classes correspondant aux niveaux les plus élevés ne sont pas représentées sur le graphique « focus ».

L'histogramme peut également être utilisé pour comparer les distributions d'une même variable au sein de deux sous-populations, ici distinguées par le sexe des individus d'une population d'étude donnée (figure 70).

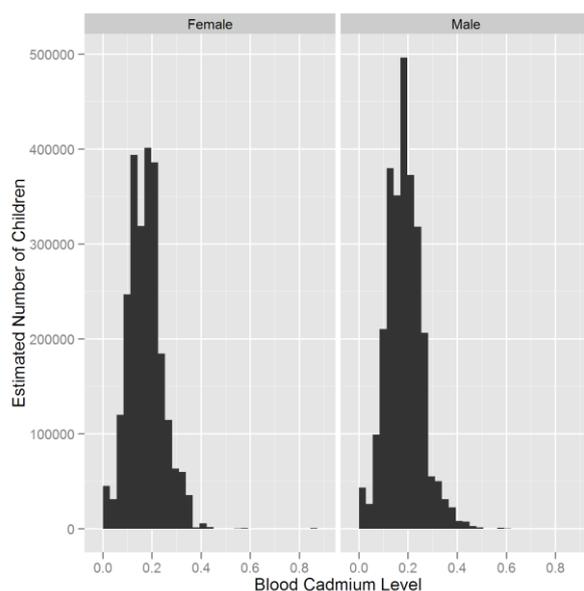
I Figure 69 I

Histogramme des distributions des niveaux de plomb dans le sang des enfants, avec focus sur partie basse



I Figure 70 I

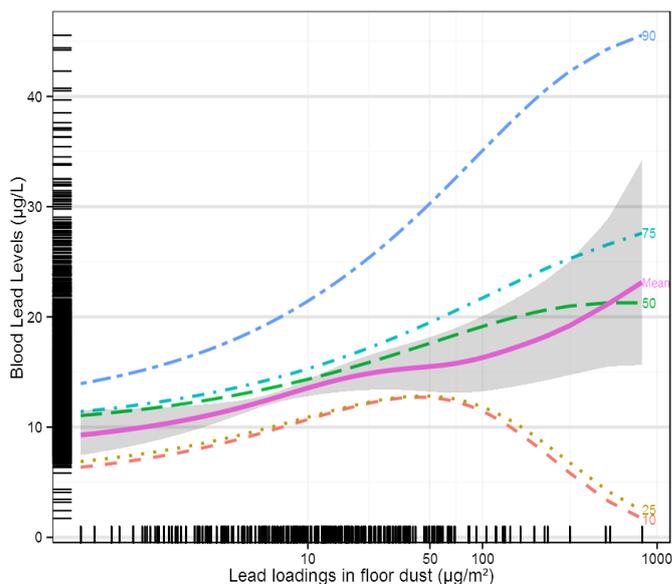
Utilisation de l'histogramme pour comparer les distributions d'une même variable sur deux sous-populations



Ci-dessous, un autre type de graphique ayant pour objectif la représentation de l'effet d'un même facteur de risque mais pour différents niveaux d'imprégnation des enfants (figure 71). La figure donne les niveaux attendus pour la moyenne et pour différents quantiles de la distribution de la plombémie (10, 25, 50, 75, 90) en fonction de la concentration moyenne en plomb dans les poussières des pièces du logement. La zone grisée représente l'intervalle à 95 % de l'estimation pour la moyenne.

I Figure 71 I

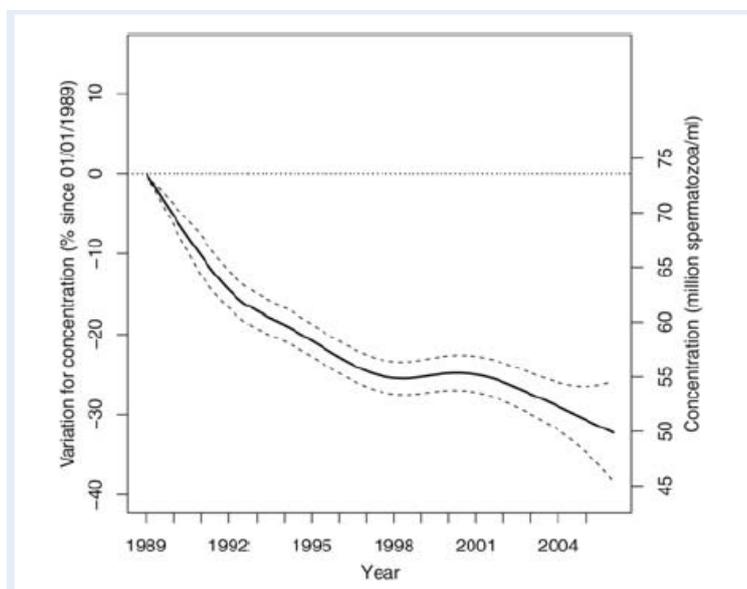
Effets de la présence de plomb dans la poussière des logements sur l'imprégnation des enfants



Dans le champ des indicateurs de santé, une sélection de représentation est présentée ci-après. Ces graphes sont extraits de travaux conduits dans le domaine de la santé reproductive, comme biomarqueur d'effets possible d'expositions environnementales. Le premier exemple est une courbe de la variation de la concentration spermatique chez les hommes de 35 ans, en France, entre 1989 et 2005. Ce graphique, illustrant un article paru en 2012 dans la revue scientifique Human Reproduction [8] s'adresse à un lectorat de spécialistes rompu à la lecture de représentations graphiques relativement complexes et à la compréhension de certaines des notions qui y sont fait mention, comme celles des intervalles de confiance.

## I Figure 72 I

Évolution (1989-2005) de la concentration spermatique des hommes de 35 ans en France métropolitaine (intervalle de confiance à 95 %)



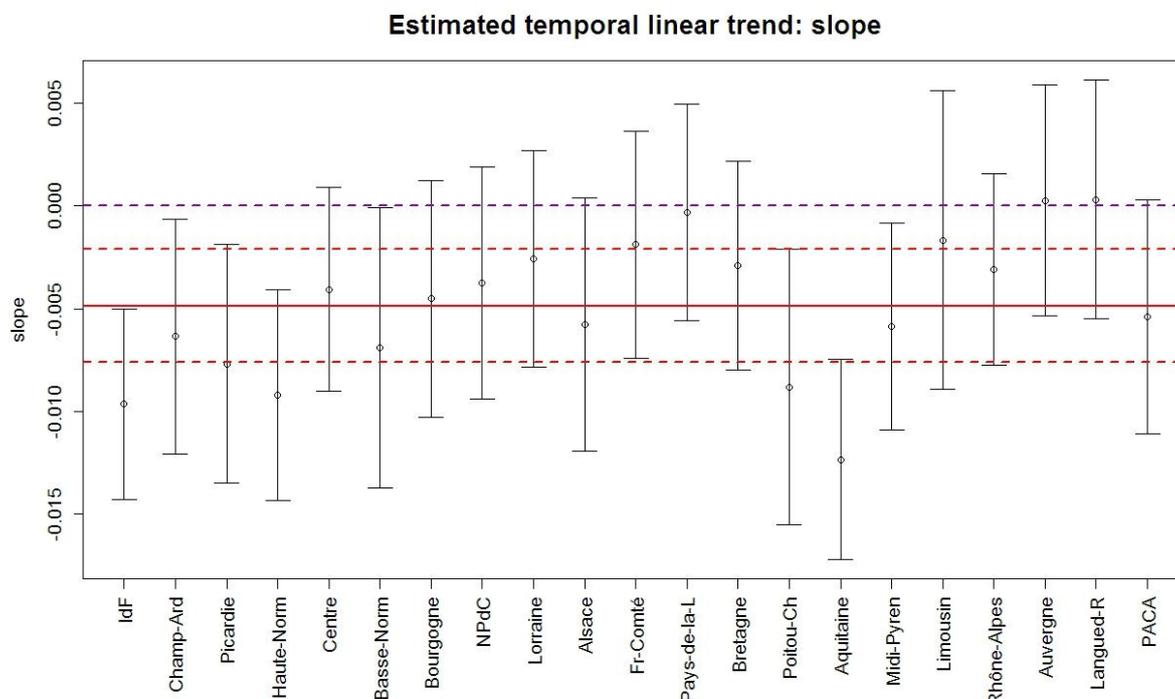
**Figure 2** Variation in percent for concentration since 1 January 1989 for a 35-year-old man in metropolitan France with 95% confidence intervals (left axis). Projected values in million spermatozoa per millilitre (right axis).

La courbe pleine (figure 72) représente la baisse de la concentration estimée, en pourcentage ; les courbes tiretées représentent les bornes de l'intervalle de confiance.

Le second graphique proposé (figure 73) est une représentation des tendances globales de l'évolution de la concentration spermatique entre 1989 et 2005, pour chacune des régions administratives françaises, par rapport à la tendance nationale qui, elle, est négative. Ce graphique vise à mettre en évidence les régions pour lesquelles la pente est à la baisse, ou à la hausse, de manière significative, ou non. Les régions pour lesquelles la valeur de la pente est statistiquement significative sont les régions pour lesquelles l'intervalle de confiance n'inclut pas la valeur caractéristique de l'effet nul (pente de 0, figurée en mauve sur le graphique). Ce graphique illustre un article soumis à la revue scientifique *Reproduction* [8] à l'automne 2013. Il s'adresse, comme le précédent, à un lectorat de spécialistes familiers de ces représentations complexes et des notions de statistiques qui y sont liées.

I Figure 73 I

Estimation des pentes de la concentration spermatique par région, entre 1989 et 2005, avec leur intervalle de crédibilité à 95 %, comparés à la pente 0 (en violet) et à la pente nationale (en rouge), avec son intervalle de crédibilité à 95 %



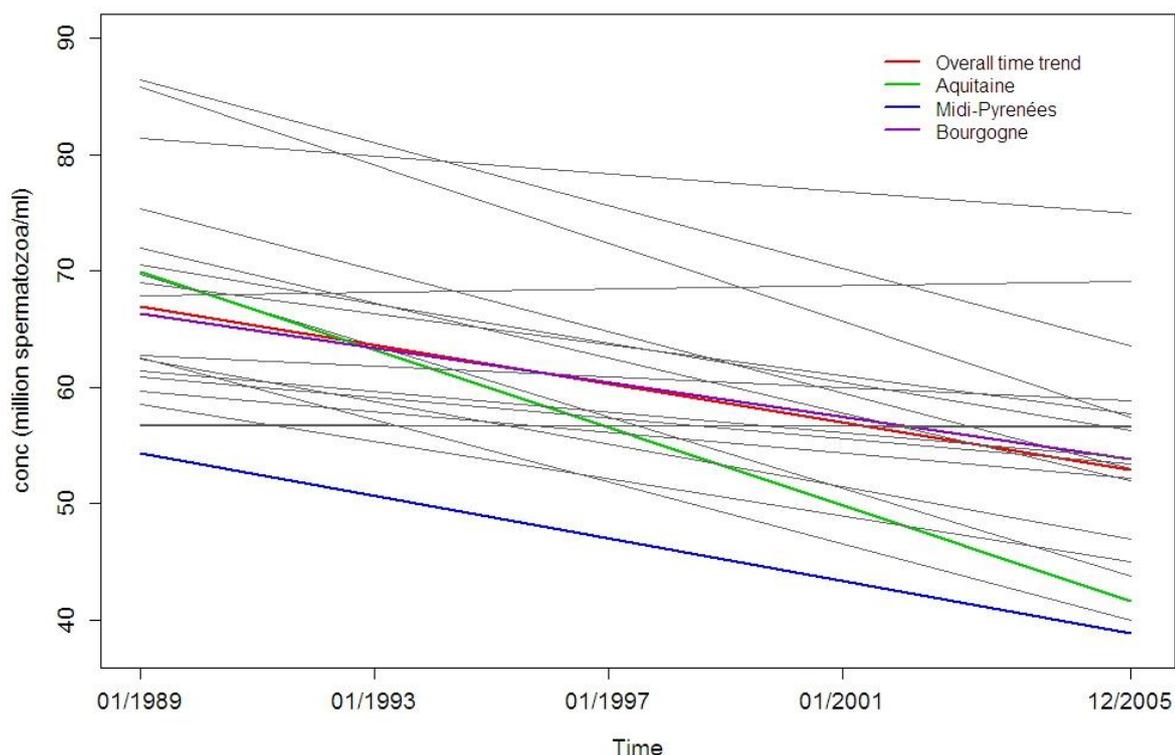
Le graphique ci-dessus (figure 73) fait apparaître, pour les 22 régions administratives françaises, les valeurs des pentes de la concentration spermatique avec leur intervalle de crédibilité. Les noms des régions sont, pour certains, raccourcis voire transformés en acronymes. Pour le lectorat anglophone et donc probablement international de la revue dans laquelle il doit être publié, les noms des régions françaises peuvent ne pas évoquer grand-chose, aussi faudrait-il veiller à ce que ces noms figurent dans la totalité de l'article de la même manière (cartographie de localisation, texte de l'article, autres graphiques, etc.). Par ailleurs, la pente 0 et la pente nationale sont représentées par des couleurs.

Or, elles se distinguent mal sur le graphique, hormis leur couleur, les tiretés sont semblables et, sur l'axe des ordonnées, les valeurs qui y sont liées ne sont pas mises en évidence même s'il y a là une volonté claire de ne pas charger le graphique. Pour ces résultats, c'est dans le libellé de figure que sont données toutes les clefs de lecture, qui gagneraient sans doute à figurer directement sur le graphique lui-même (valeur du 0 représenté en mauve, valeur pour la France entière figurée par le trait plein rouge, et bornes de l'intervalle de crédibilité pour la France entière figurées par les tiretés rouges). Par ailleurs, en fonction des valeurs régionales que l'on souhaite individualiser, un regroupement par catégories des régions aurait pu être envisagé : on aurait pu trier et regrouper sur le graphique, par exemple : les régions non significatives, les régions de pente négative et significative, les régions de pente positive (même si les deux régions pour lesquelles la pente est légèrement positive, ne le sont pas significativement comparativement à 0). Ces mises en exergue et regroupement d'informations comparables auraient l'intérêt de rendre la lecture et le décodage d'un tel graphique plus rapide, pour un public qui par ailleurs est familier des méthodes de représentation de ce type d'informations.

Le troisième exemple proposé est une illustration (figure 74) qui montre l'évolution linéaire des valeurs de la concentration spermatique (en millions de spermatozoïdes par millilitre) pour chaque région, sur la période 1989-2005. Le graphique n'est pas très lisible, et les tentatives d'individualisation de certaines régions d'intérêt ne semblent pas suffisantes. Le contraste entre l'ensemble des régions (en gris) avec les valeurs des régions mises en exergue (valeur pour l'Aquitaine, la région Midi-Pyrénées et la Bourgogne, et tendance nationale) n'est également pas suffisant. Pour renforcer le contraste, un jeu est possible sur le choix des couleurs, mais aussi sur l'épaisseur des traits, les droites des régions qui ne sont pas mises en avant pourraient éventuellement apparaître dans un gris plus clair, voire en blanc sur fond gris. Rajouter des axes secondaires sur le graphique ne ferait sans doute pas gagner en lisibilité vu la densité de lignes qui y figurent, par contre, afin de faciliter le repérage, l'axe des ordonnées aurait pu être doublé et figuré en bordure droite de la zone graphique. Enfin, autant que faire se peut, la mention des noms des régions directement à la limite des droites correspondantes plutôt que sur une légende distincte pourrait éventuellement faciliter la lecture du graphique.

I Figure 74 I

Estimation de l'évolution de la concentration spermatique (millions de spermatozoïdes/ml) entre 1989 et 2005, ajustée sur l'âge, pour chaque région



### 3.1.5 Exemples de représentations graphiques dans le champ de la toxicovigilance

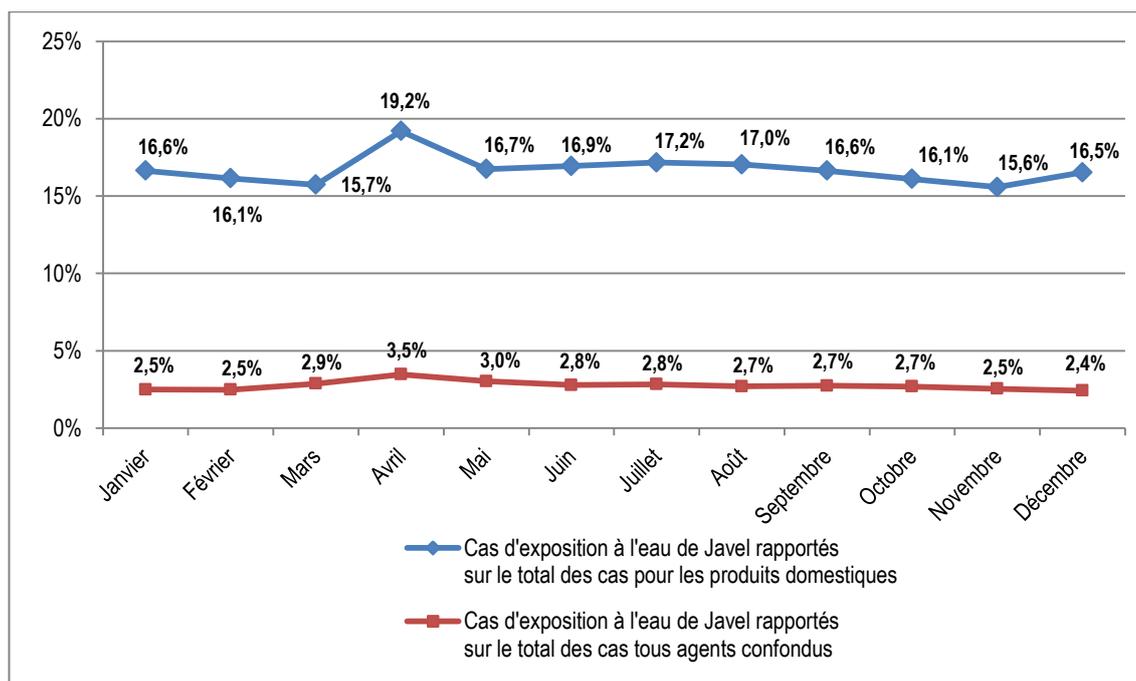
Dans le champ de la toxicovigilance et de la surveillance des intoxications, la variété des agents n'induit pas forcément une très grande diversité dans la représentation graphique des données disponibles : les modalités de la surveillance induisent des méthodes et des indicateurs assez similaires.

Les premiers exemples de graphiques présentés sont extraits d'un rapport en cours de finalisation qui traite des expositions domestiques à l'eau de Javel sur l'année 2011 analysées à partir des données du Système d'information du réseau des CAP TV, les centres antipoison et de toxicovigilance (SICAP). Ce rapport a pour objectif de décrire et analyser ces expositions, professionnelles ou non, symptomatiques ou non, survenues dans le cadre domestique. Les graphiques annexés à ce rapport ont été réalisés dans le cadre de cet objectif de description et d'analyse.

Le premier graphique proposé (figure 75) a pour objectif de représenter la part des expositions à l'eau de Javel, par rapport au total des expositions à des produits domestiques d'une part, et par rapport aux expositions tous agents confondus recensées dans SICAP d'autre part, et ce, mois par mois. Le titre proposé est libellé comme suit :

I Figure 75 I

Distribution mensuelle des cas d'exposition à l'eau de Javel par mois, France 2011 (N=5 452)



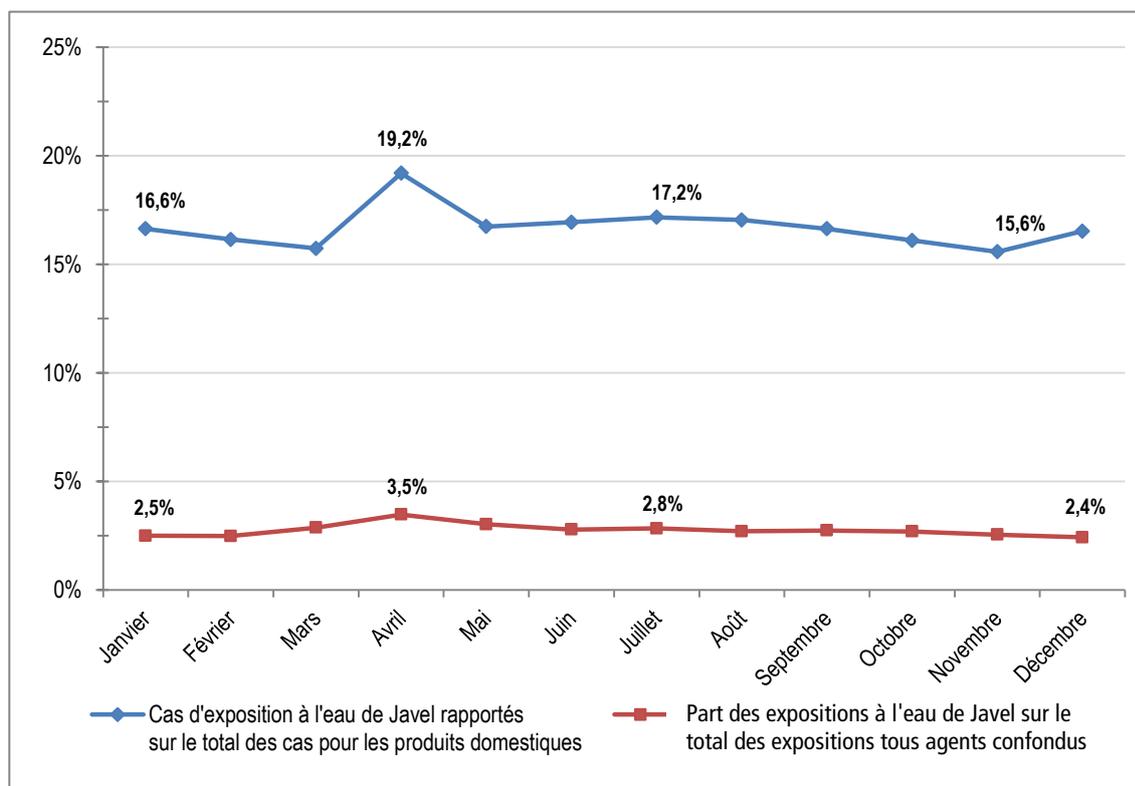
Le choix de représentation graphique semble assez adapté : les points, mois par mois, positionnent la part des expositions à l'eau de Javel par rapport à un dénominateur différent : en bleu, le total des expositions aux produits domestiques, et en rouge, le total des expositions tous agents confondus. La compréhension de l'information n'est *a priori* pas une difficulté. Cependant, elle s'avère ne pas être si aisée et c'est sans doute davantage lié aux textes périphériques (légende, titre), plus qu'au choix de représentation qui a été fait.

Le titre mériterait peut-être d'être modifié en spécifiant plutôt : **Part mensuelle des expositions à l'eau de Javel recensées dans SICAP, France, 2011**. En effet les valeurs représentées sont des pourcentages. Pour les textes de légende il pourrait être pertinent de remplacer : *Cas d'exposition à l'eau de Javel rapportés sur le total des cas pour les produits domestiques* par **Part des expositions à l'eau de Javel sur le total des expositions à des produits domestiques**, et *Cas d'exposition à l'eau de Javel rapportés sur le total des cas tous agents confondus* par **Part des expositions à l'eau de Javel sur le total des expositions, tous agents confondus**.

Par ailleurs, à moins que cela soit expressément demandé par l'un des co-auteurs ou relecteurs, la mention systématique des valeurs au-dessus des points du graphique n'en facilite pas la lecture. Le choix de deux ou trois valeurs maximum à ajouter sur les points peut être fait si ces valeurs particulières sont par ailleurs reprises dans le texte du rapport comme valeurs d'intérêt. Dans la proposition ci-dessous, quelques valeurs ont été isolées et marquées : valeurs minimum et maximum, valeur du premier mois étudié, par exemple. Si l'on veut rendre plus évidentes les variations subtiles des valeurs mois par mois, il y a éventuellement d'autres moyens techniques comme l'ajout d'interlignes au fond du graphique par exemple, ou, comme ci-dessous simplement, les graduations intermédiaires entre les valeurs des graduations principales, et/ou encore l'augmentation de la zone graphique pour mieux faire apparaître ces variations, qui sont ici assez faibles de toute façon, en dehors du cas particulier du mois d'avril.

**I Figure 76 I**

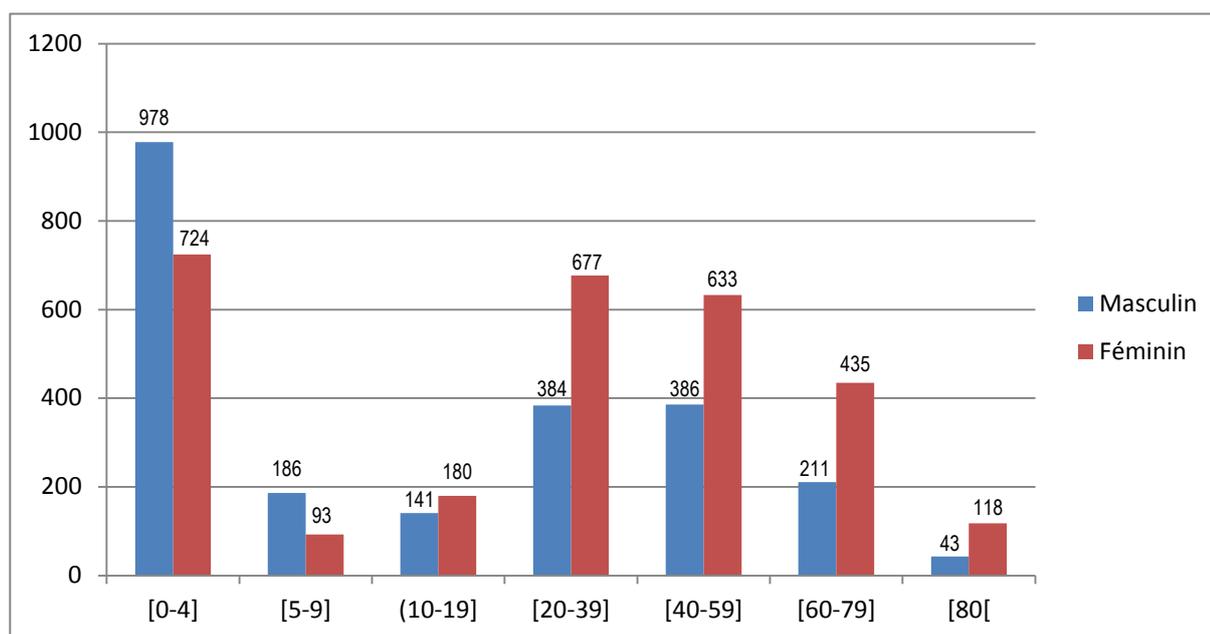
Proposition d'allègement et d'amélioration du graphique de la figure 75



La suite du rapport décrit plus précisément la population concernée par ces expositions à l'eau de Javel : tranche d'âge, sexe, voie d'exposition, etc. Les données sont représentées par une série de graphiques en barres qui figure, par exemple, pour chaque tranche d'âge étudié, la répartition par sexe des cas d'exposition, ou encore la répartition en fonction de la voie d'exposition. Pour chacun de ces graphiques, comme dans les deux exemples ci-dessous (figures 77 et 78), demande a été faite par les CAPTV que soient figurés systématiquement toutes les valeurs exactes sur les graphiques, ce qui en alourdit légèrement la composition et la lecture.

### I Figure 77 I

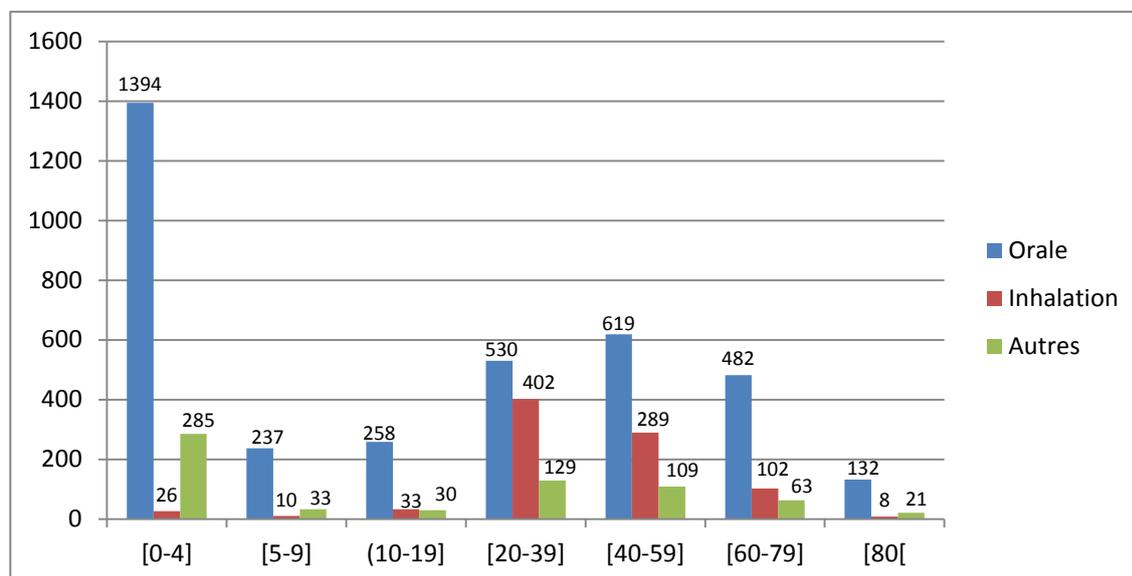
Cas d'exposition à l'eau de Javel selon la tranche d'âge et le sexe, France, 2011  
(N=5 189/5 452)



Un autre exemple présenté est celui de la distribution des voies d'exposition selon la tranche d'âge, France, 2011 (N=5 192/5 452, où 5 192 correspond au nombre de cas pour lesquels la voie d'exposition était renseignée).

I Figure 78 I

Distribution des voies d'exposition selon la tranche d'âge, France, 2011 (N=5 192/5 452, où 5 192 correspond au nombre de cas pour lesquels la voie d'exposition était renseignée)



Remplacer les valeurs brutes des effectifs dans un tableau à part allégerait la lecture des graphiques, mais l'information serait alors doublonnée. Une proposition est faite pour améliorer l'information issue de ces graphiques tout en respectant la demande des CAPTV : il pourrait en effet être pertinent de synthétiser ces données brutes dans un tableau, où les valeurs apparaîtraient telles quelles.

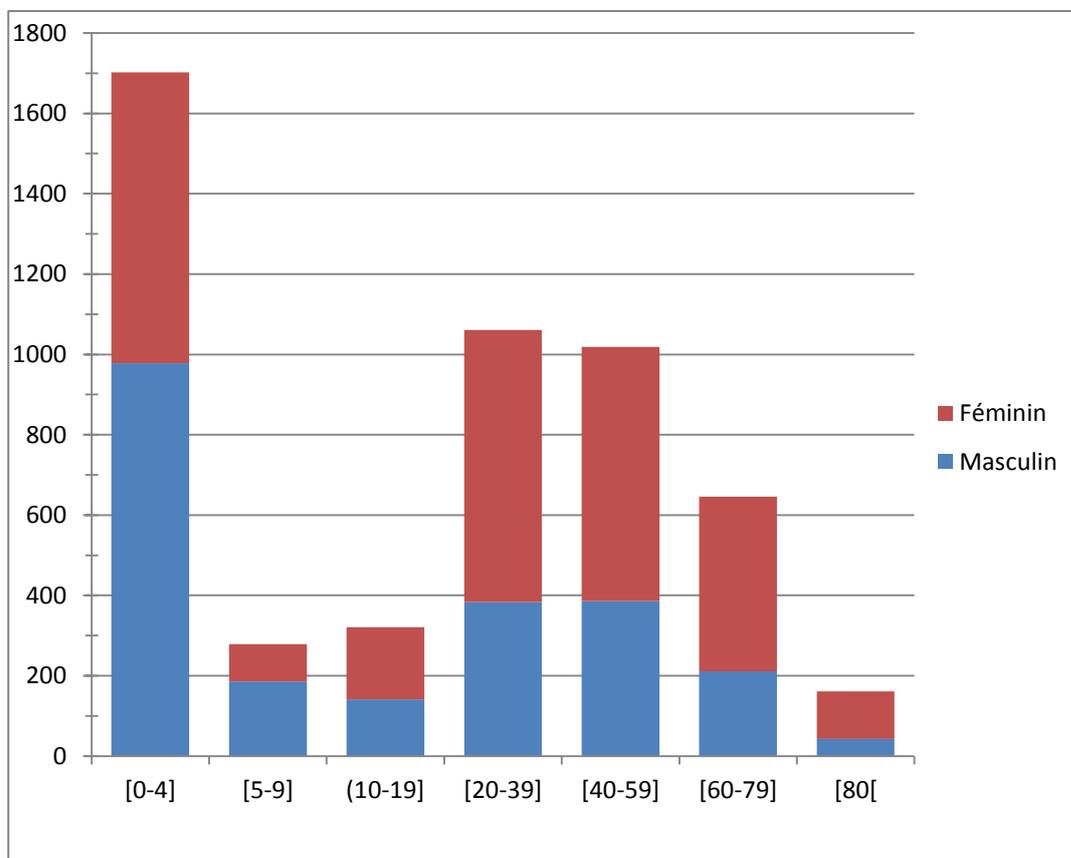
Les graphiques seraient par ailleurs débarrassés des étiquettes numériques encombrantes et pourraient gagner en information si l'on faisait figurer, au lieu des effectifs, les proportions, par classe d'âge, pour le sexe d'une part et la voie d'exposition d'autre part. En procédant de cette manière, on répond aux attentes du partenaire qui souhaite que les effectifs apparaissent, on évite la redondance, et on exploite davantage l'information en en dégagant des proportions qui ne sont pas directement lisibles dans les graphiques tels qu'ils sont réalisés initialement. Des précautions doivent cependant être apportées à la taille de la zone graphique pour que les tout petits effectifs de la catégorie « exposition par inhalation » restent lisibles. Afin de les rendre plus visibles il est possible de faire le choix d'élargir les barres en réduisant l'intervalle qui les sépare.

Voici ci-après une première série de propositions (figures 79 et 80), où l'on conserve les effectifs, ce qui permet de garder directement lisible le fait que toutes les classes d'âge sont loin d'être équivalentes face au risque, et plus loin, une seconde série de propositions, où les proportions sont figurées en pourcentage : les effectifs disparaissent, et les différences d'effectifs entre les classes d'âge également, mais par contre la répartition au sein de chaque classe d'âge est clairement lisible.

Suit une première série de propositions (sans gommer les différences d'effectifs, on reste dans la représentation des effectifs pour chaque classe d'âge).

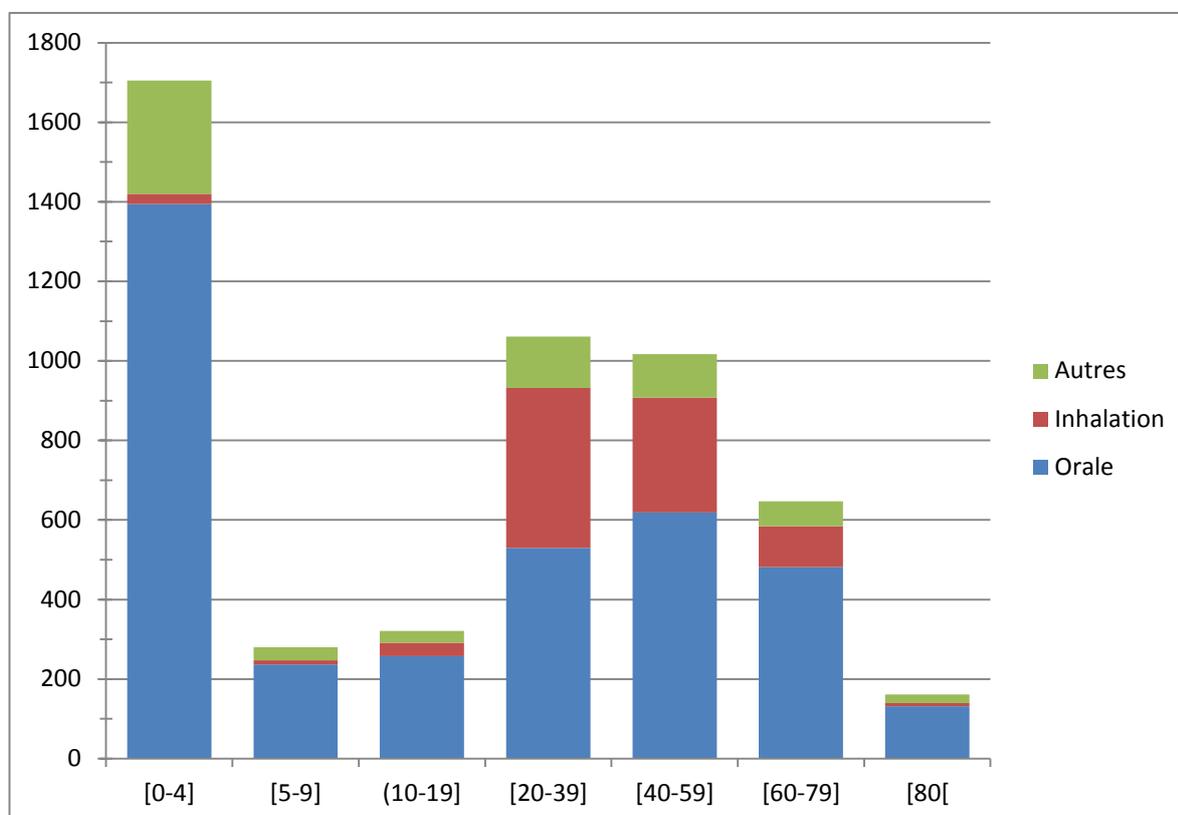
I Figure 79 I

Proposition d'amélioration pour la représentation des cas d'exposition à l'eau de Javel selon la tranche d'âge et le sexe, France, 2011 (N=5 189/5 452)



I Figure 80 I

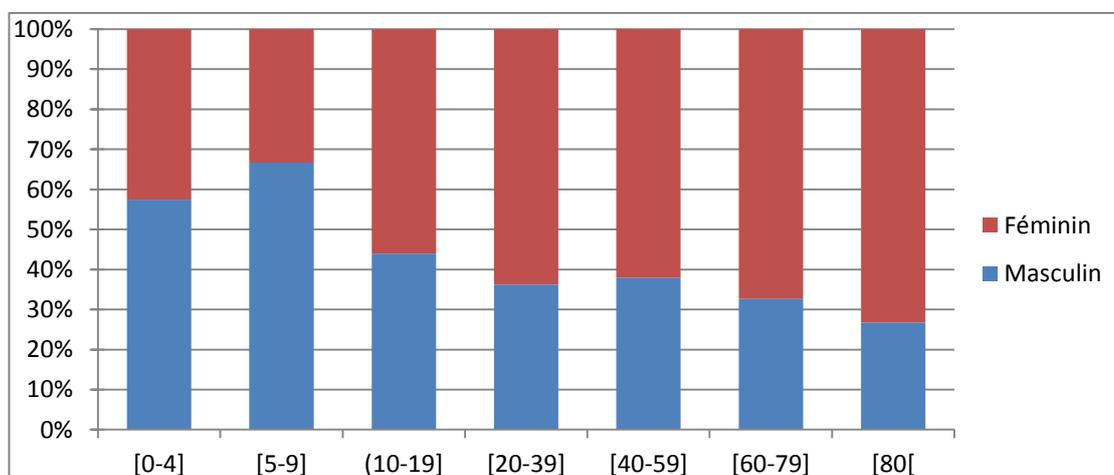
Proposition d'amélioration pour la représentation de la distribution des voies d'exposition selon la tranche d'âge, France, 2011 (N=5 192/5 452, où 5 192 correspond au nombre de cas pour lesquels la voie d'exposition était renseignée)



Dans la seconde série de propositions (figures 80 et 81), on passe à une représentation de la répartition au sein des classes d'âges en base 100, les effectifs bruts étant totalement gommés.

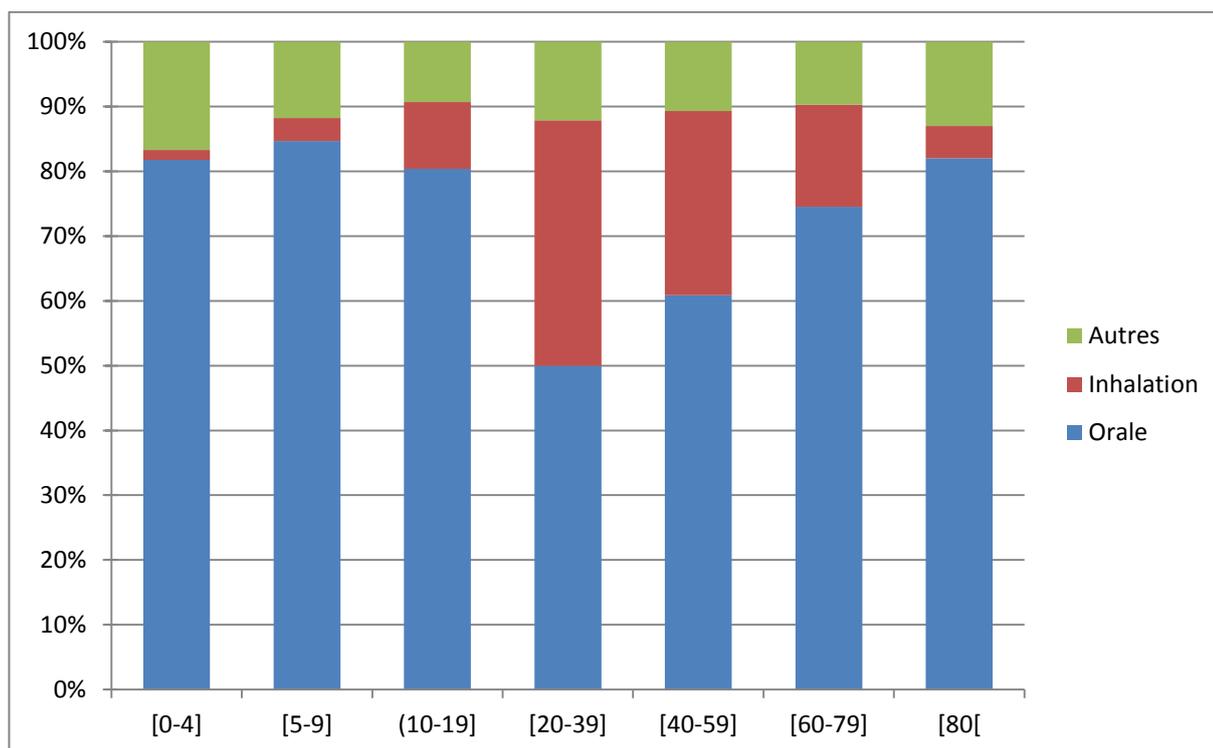
I Figure 81 I

Alternative pour la représentation des cas selon la tranche d'âge et le sexe, en base 100



I Figure 82 I

Alternative pour la représentation des voies d'exposition selon l'âge, en base 100



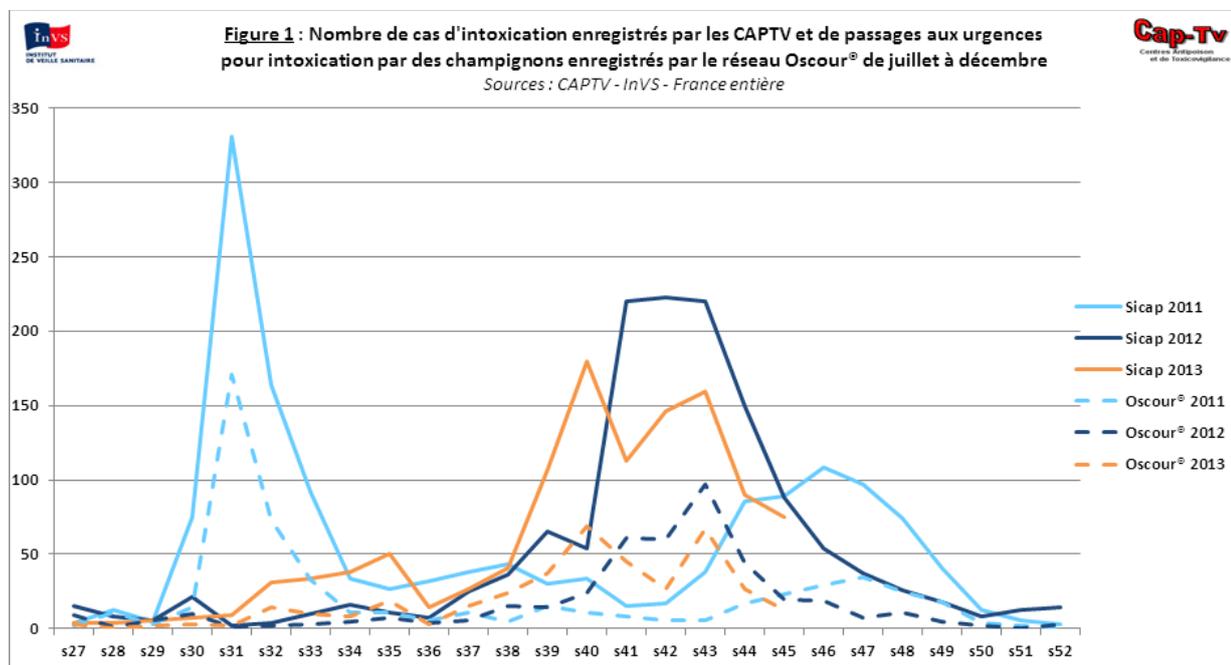
Le choix entre la première et la seconde série de propositions sera à effectuer en fonction des objectifs de la représentation graphique des données. Mais on peut éventuellement considérer que l'insertion de tableaux reprenant les valeurs brutes des effectifs recensés dans SICAP autorise le choix de ne faire figurer dans les graphiques que les proportions en ramenant les hauteurs des barres à 100 % comme dans la seconde série de propositions.

Le champ de la surveillance des intoxications implique également la réalisation assez fréquente de courbes épidémiques. L'exemple suivant est issu de la surveillance hebdomadaire des intoxications aux champignons, conduite à partir des données du SICAP ainsi que des données du réseau OSCOUR® (Organisation de la surveillance coordonnée des urgences) qui couvre environ 67 % des services d'urgences hospitalières de France. Cette surveillance est conduite chaque année de juillet à décembre. Une expertise toxicologique et mycologique est réalisée par le réseau de toxicovigilance, notamment pour les cas graves. Une courbe épidémique est ainsi réalisée semaine par semaine pendant la saison.

Voici une représentation graphique (figure 83) de cette courbe épidémique qui superpose différentes années et les différentes sources de données utilisées. Le résultat est un peu chargé et difficile à lire, et il paraît difficilement envisageable de poursuivre une telle réalisation dans les années à venir : avec trois années x deux sources de données, on atteint déjà six courbes qui se recoupent et se chevauchent, il semble peu probable d'obtenir un résultat qui reste lisible avec une année (soit deux courbes) supplémentaire(s). De plus, la présence de pics oblige d'introduire une échelle de valeurs large, et les variations dans les faibles valeurs sont d'autant moins lisibles. Ce type de graphique vise à synthétiser, en une seule représentation, des objectifs divers : suivre dans le temps un phénomène, mesurer les écarts entre deux sources de données non exhaustives pour vérifier la robustesse d'un système par rapport à l'autre qui fait référence, par exemple, et faire apparaître des variations/des ressemblances de ce phénomène d'une année sur l'autre.

## I Figure 83 I

Représentation de la variation du nombre de cas d'intoxications aux champignons référencées par les différentes sources de données sur 3 ans

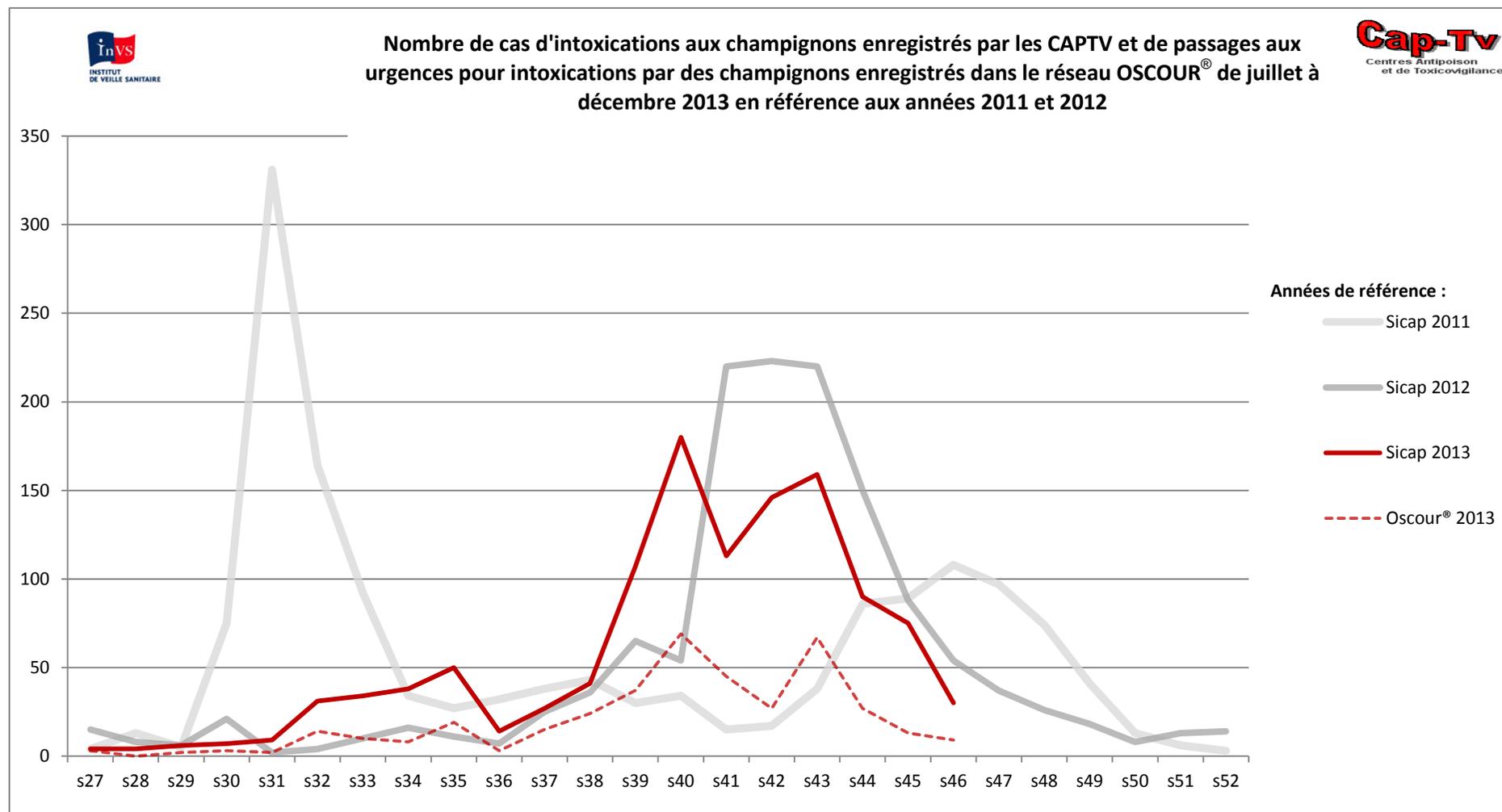


Pour le suivi des intoxications aux champignons, les données d'OSCOUR® ont été introduites dans la surveillance en 2011. Cette année a présenté un pic d'intoxications précoce en juillet assez inhabituel (le pic est plus classiquement positionné en septembre). En 2011, la confrontation des informations remontant des deux systèmes – SICAP et passages aux urgences codés en intoxications aux champignons dans OSCOUR® – a permis de constater la robustesse des informations de ce second système : les tendances sont similaires dans les deux bases. L'année 2012 a consolidé cette observation en mettant en regard, là encore, des informations remontant des deux systèmes.

Pour l'année 2014, l'objectif serait plus précisément d'utiliser les données des années précédentes (mais uniquement SICAP), pour mieux anticiper le début de la saison et l'arrivée d'un pic afin d'avoir une action de prévention plus réactive et efficace. La superposition de l'ensemble des courbes sur les trois années antérieures ne permettra pas une lisibilité suffisante. On pourrait, pour répondre à cet objectif, proposer la solution graphique suivante (figure 84) qui, en hiérarchisant les informations (mise en valeur des courbes de l'année en cours, mise en « fond » des données de référence susceptibles de servir de repères), permettrait de regagner en lisibilité. Les courbes de comparaison entre les deux sources de données pourraient par ailleurs continuer à être réalisées, mais dans des graphiques à part (par exemple le graphique de la figure 85), dans la mesure où elles répondent à un autre objectif (consolidation/validation du système de surveillance reposant sur deux sources de données différentes).

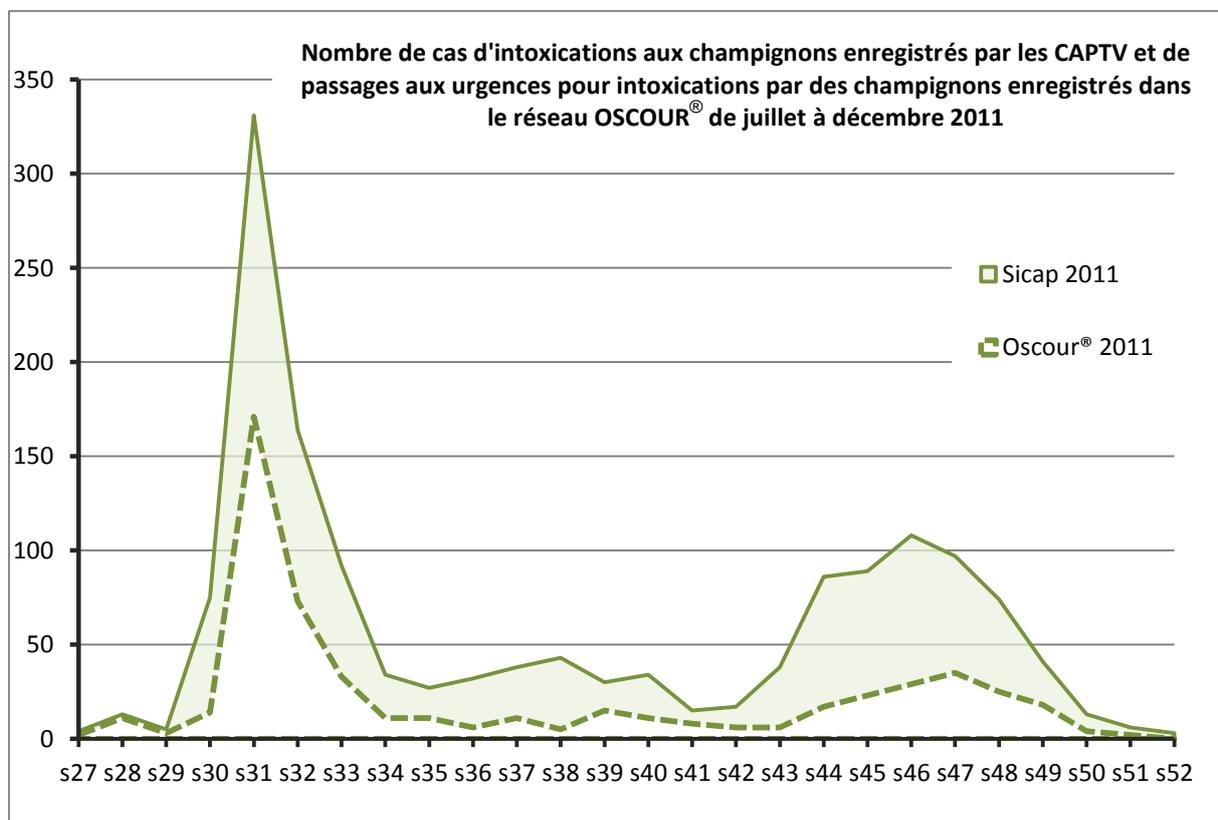
I Figure 84 I

Proposition d'amélioration pour le suivi des intoxications aux champignons sur l'année en cours (ici, exemple de 2013), par sélection et mise en exergue de certaines informations par rapport à d'autres



I Figure 85 I

Proposition pour la comparaison des deux systèmes d'information SICAP et OSCOUR® exemple de l'année 2011



La dissociation de ces différents objectifs pourra probablement permettre de gagner en lisibilité. En réalisant plusieurs graphiques, on obtient un résultat certes moins synthétique, mais plus efficace visuellement.

### 3.1.6 Exemples de représentations graphiques dans le champ de la surveillance du climat

L'épisode de canicule de 2003 et la vague de surmortalité qu'elle a entraînée ont fait l'objet de plusieurs études au DSE. L'un de ces travaux portait plus spécifiquement sur le phénomène des îlots de chaleur urbains. Les publications sur ce sujet présentent l'intérêt particulier d'avoir été déclinées selon différentes modalités afin de diffuser l'information à des publics différents. Les exemples proposés ici sont issus de ces différents supports de valorisation et le public destinataire du graphique constitue un élément décisif dans les choix graphiques qui ont été faits.

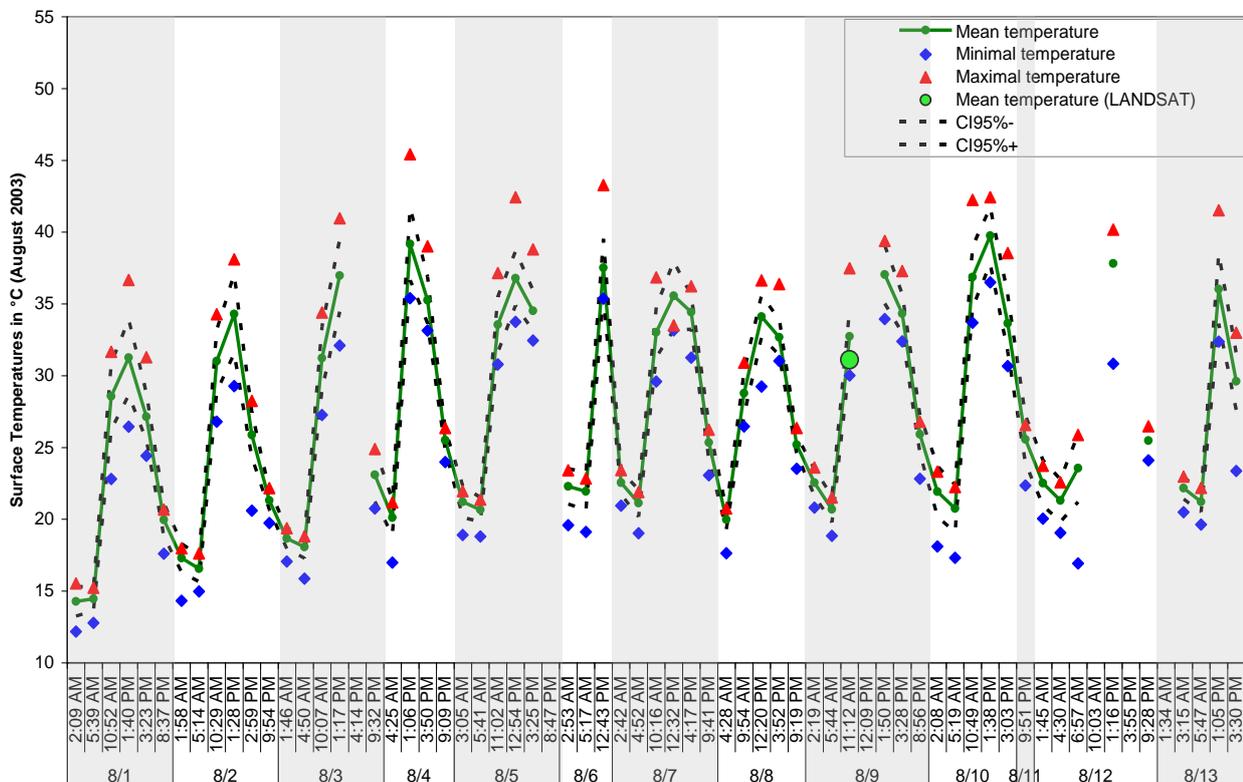
Le premier exemple proposé (figure 86) est un graphique représentant les variations de températures en fonction de l'heure pendant la canicule de 2003. Les données sont issues des images du satellite NOAA qui effectue des passages sur une même zone environ toutes les trois heures. Cependant, certaines données sont manquantes du fait de la non lisibilité de certaines images (en raison de

la présence de nuages, par exemple) ce qui explique que le graphique ne présente pas un nombre de valeurs constant d'un jour à l'autre. Ce graphique a été publié dans un article scientifique qui est paru dans la revue *Environmental Health Perspective (EHP)* en 2011 [11]. Il était destiné à un public averti, *a priori* rompu à la lecture et à la compréhension de ce type de représentations.

Cependant, bien que cette représentation ait été réalisée pour un public scientifique, la lecture n'en est pour autant pas très simple. La multiplicité des informations qui sont figurées brouille la lisibilité, d'autant que ces informations apparaissent toutes avec le même niveau d'importance, sans véritable hiérarchisation. Les pointillés liés à la représentation des bornes des intervalles de confiance 95 % sont assez peu lisibles et surchargent le graphique. On peut par contre saluer le choix d'individualiser les jours par des alternances de bandes grisées et blanches sur l'aire graphique, et ce, d'autant que le nombre de mesures d'un jour à l'autre n'est pas constant du fait des données manquantes liées aux images satellites NOAA à partir desquelles les températures sont calculées.

I Figure 86 I

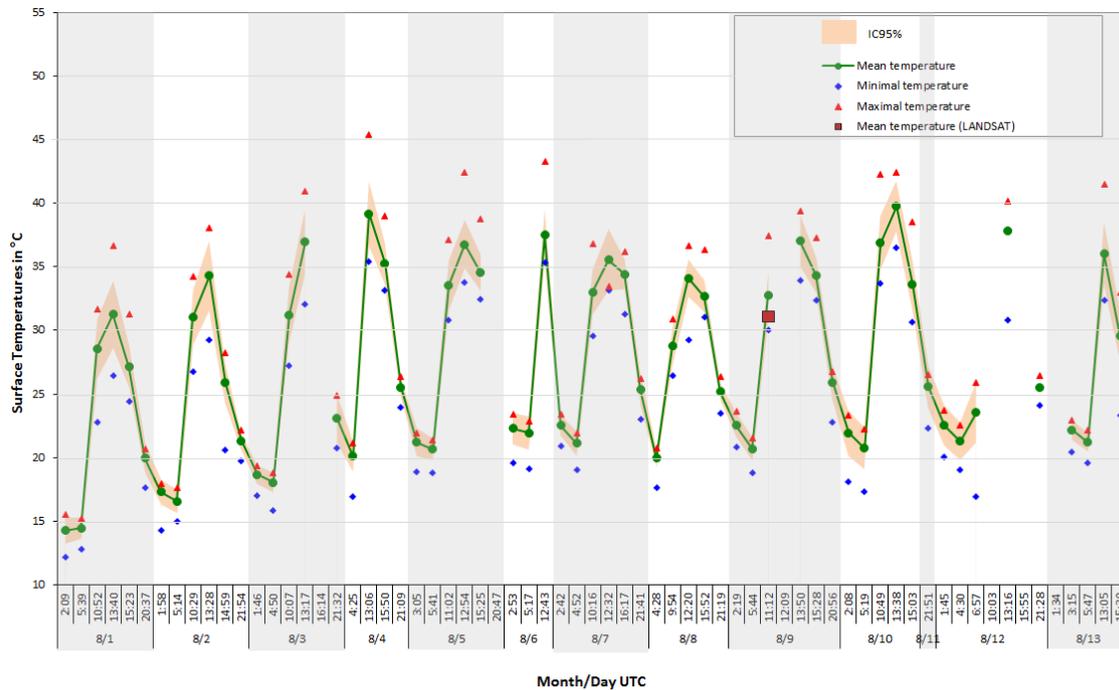
Variations de températures en fonction de l'heure pendant l'épisode de canicule de 2003



Le graphique ci-après (figure 87) est une proposition d'amélioration du précédent. Les bornes de l'intervalle de confiance à 95 % ont été représentées par un « ruban » de couleur claire, qui met l'information légèrement en retrait par rapport à l'objet principal du graphique. Des modifications (corps du texte, grasse, taille des marqueurs, etc.) ont également été réalisées afin d'introduire davantage de hiérarchisation de l'information et de faciliter la lecture et la compréhension de ce document.

## I Figure 87 I

Proposition pour améliorer la lisibilité du graphique précédent (figure 86)



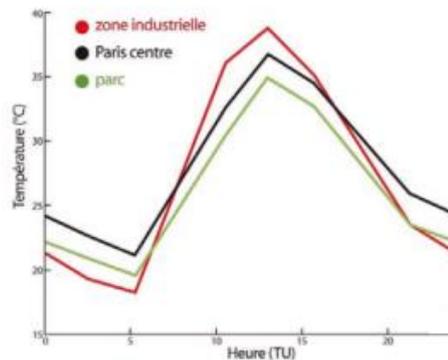
La courbe suivante est issue d'une plaquette [10] traitant des îlots de chaleur urbains (figure 88). Ce document, destiné au grand public, fait suite à l'épisode de canicule de 2003, et fournit une synthèse des études qui ont été réalisées par l'InVS suite à cette vague de chaleur inhabituelle et à la surmortalité qu'elle a entraînée.

Le document étant destiné à un public peu expert, les termes choisis et les représentations adoptées sont simples, et le parti pris ci-dessus est de viser en premier lieu une bonne compréhension par le lecteur du phénomène des îlots de chaleur urbains.

## I Figure 88 I

Illustration grand public des écarts de températures entre la nuit et le jour selon le mode d'occupation du sol

Évolution des températures sur un cycle de 24 heures pour différents modes d'occupation du sol (industriel, urbain, parc).



**Zones industrielles** : très chaudes le jour, plus froides la nuit (les infrarouges s'échappent facilement).

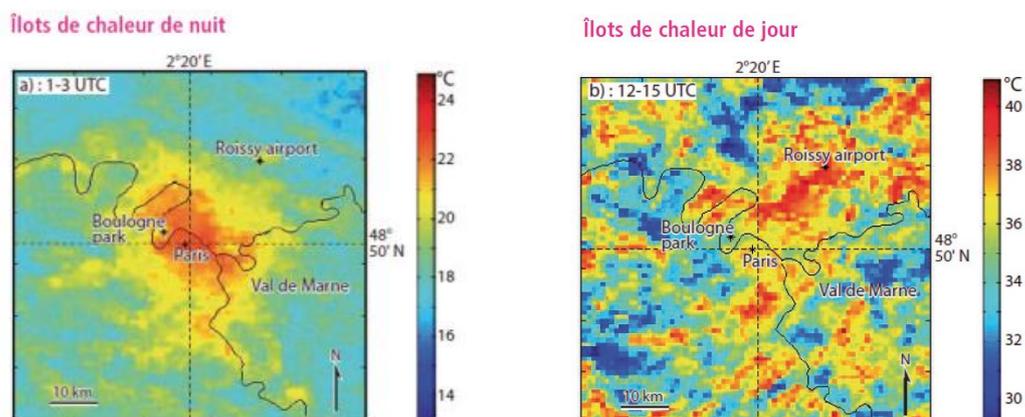
**Centre ville** : très chaud la nuit, ne permet pas à l'organisme de récupérer de la chaleur diurne.

**Parcs** : températures 2 à 3 °C inférieures au centre-ville sur l'ensemble du cycle diurne.

Le graphique (figure 89), qui décrit bien l'évolution journalière des températures en fonction du type d'occupation du sol vient compléter les deux cartes qui y sont associées, qui montrent les déplacements de cet îlot de chaleur en fonction de l'heure : concentré sur le centre très urbain de l'agglomération parisienne la nuit, et plus diffus sur la périphérie pendant la journée.

## I Figure 89 I

Représentation cartographique des îlots de chaleur urbains de jour, et de nuit, sur l'Ile-de-France



Dans des situations de représentation complexe, (information complexe en soi, ou bien combinaison d'informations, nombre d'individus/d'entités à distinguer, période d'étude longue ou divisée en

plusieurs sous-périodes, etc.), une piste à explorer pourrait consister par exemple à tester plusieurs méthodes de représentation puis de les soumettre à diverses personnes pour savoir ce qu'elles comprennent et, éventuellement, choisir une méthode plutôt qu'une autre. Dans la plupart des situations de représentation graphique d'une ou plusieurs informations, il n'y a sans doute pas une seule et unique manière de les représenter et l'exercice pourrait être intéressant de tester la compréhension par un panel de différentes personnes afin d'évaluer et éventuellement remettre en question les choix de représentation que l'on a envisagés. Dans tous les cas, procéder par itération semble une démarche assez adaptée à la recherche du meilleur mode de représentation graphique en fonction d'objectifs préalablement fixés.

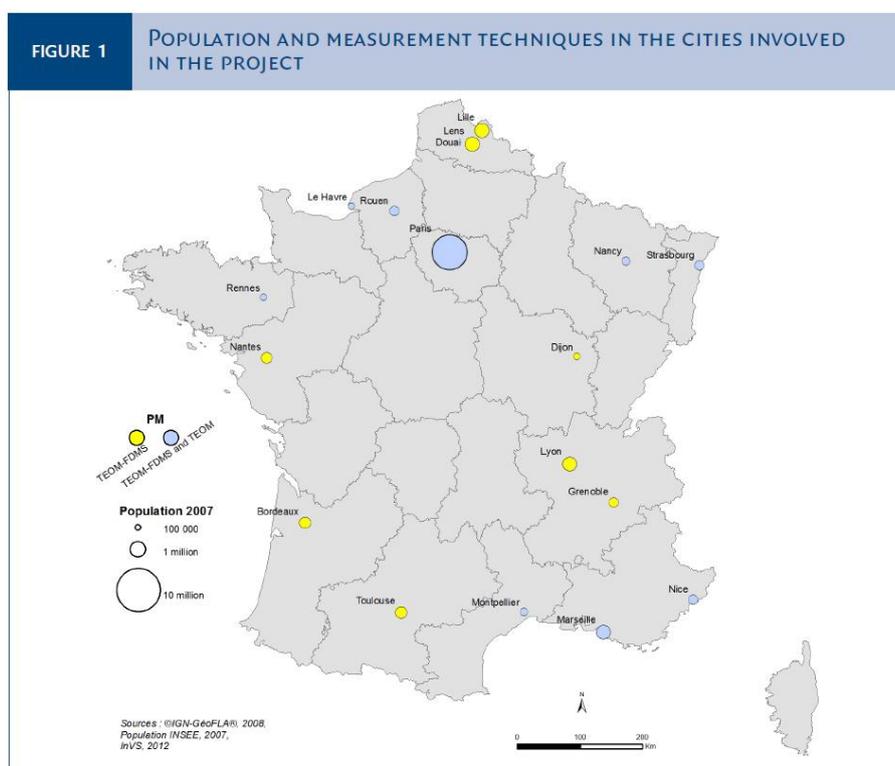
## 3.2 Représentations cartographiques

### 3.2.1 Combinaison d'informations dans une même carte

Représenter sur une même carte et sur les mêmes objets géographiques des informations de natures radicalement différentes

#### I Figure 90 I

Représentation cartographique combinée de deux informations : cartographie extraite du poster "Influence of the PM<sub>10</sub> measurement techniques on the assessment of the short-term mortality impacts of PM<sub>10</sub>", Mathilde Pascal, ISEE, 2013



Dans la carte ci-dessus (figure 90), il s'agissait de faire figurer en une seule illustration cartographique deux informations radicalement différentes : la localisation et l'effectif de population des villes contribuant à l'étude d'une part, et, d'autre part, les deux techniques de mesures utilisées par ces villes pour effectuer les mesures de concentration en PM<sub>10</sub>. La contrainte est forte puisque le format d'un poster impose que l'espace alloué soit restreint et regrouper les informations à cartographier sur une même carte s'avère essentiel. La difficulté ici tient à la nature radicalement différente des informations à représenter conjointement. Le fait d'attribuer une couleur pour catégoriser les techniques de mesures mises en œuvre, mais d'appliquer ces couleurs aux cercles qui eux fournissent l'information sur la taille de population laisse visuellement entrevoir que ces deux informations sont reliées ce qui n'est pas le cas. Une piste alternative aurait pu être explorée en attribuant des couleurs plutôt à l'écriture des noms des villes plutôt qu'aux cercles figurant la population, et ce d'autant qu'il n'y a que deux classes pour ce qui concerne les techniques de mesure ce qui laisse espérer que la lisibilité aurait été préservée.

## 2.2.2 Représentations cartographiques dans le champ des études d'impact sanitaire

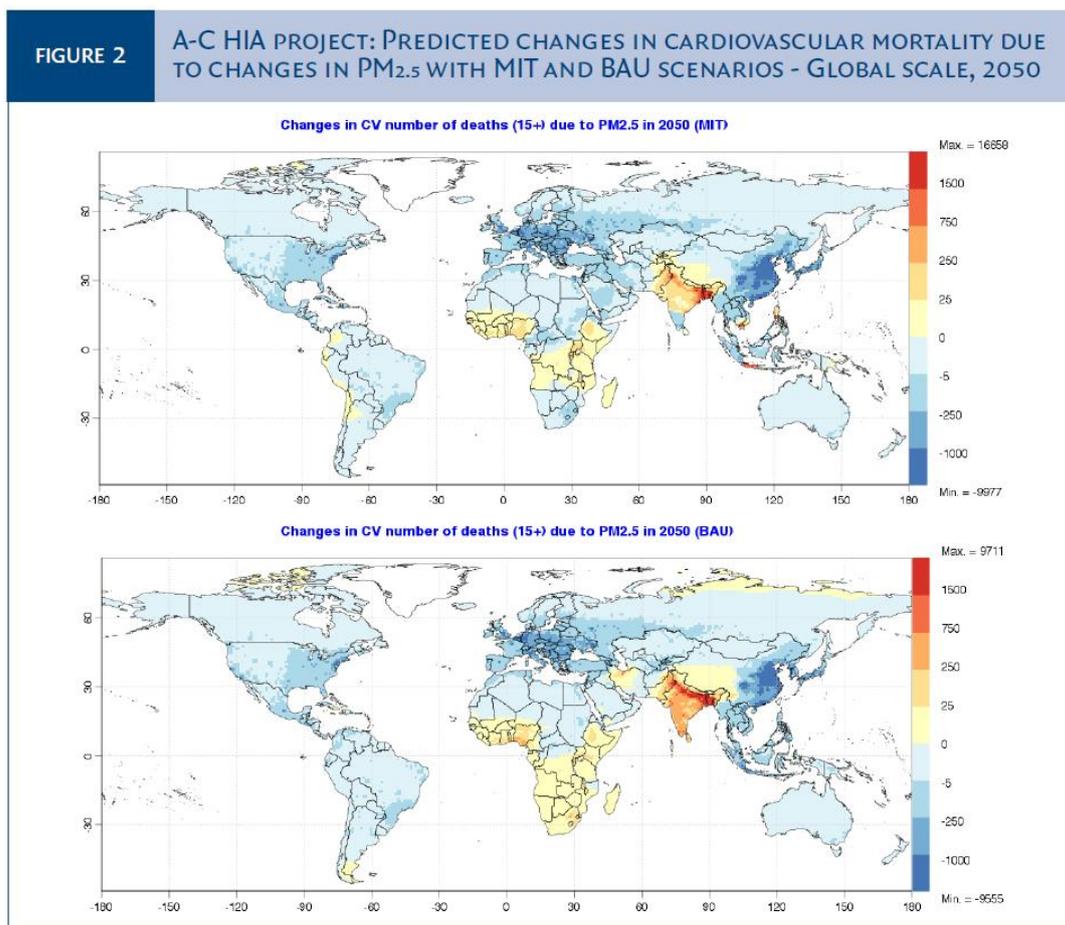
### La cartographie comme support visuel de la comparaison

Une des composantes du projet ACHIA consiste à réaliser une EIS à trois échelles emboîtées (échelle globale, échelle « régionale » menée sur l'Europe, et l'échelle locale, portant sur l'Île-de-France).

L'EIS porte sur l'évolution de la mortalité (toute cause, mortalité cardio-vasculaire, mortalité respiratoire) en fonction de l'évolution des concentrations en  $PM_{2,5}$  à horizon 2050, selon des modélisations qui tiennent compte du changement climatique. Les modélisations ont été effectuées en imaginant deux scénarii possibles : un scénario dit de « mitigation » pour lequel on envisage que des actions sont menées pour faire diminuer les niveaux de concentration en  $PM_{2,5}$  dans l'atmosphère, et un scénario dit « business as usual », c'est-à-dire qu'aucune action particulière n'est entreprise pour faire évoluer ces niveaux.

### I Figure 91 I

Cartographies extraites du poster "Quantifying health impacts of air pollution under a changing climate: from a global to the regional and local scales" – Victoria Likhvar, ISEE, 2013



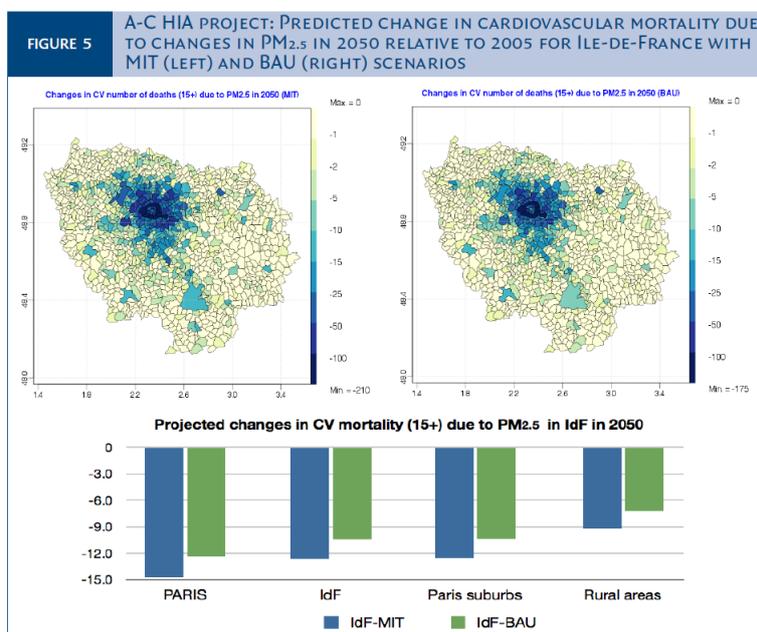
L'objectif de la représentation des résultats de l'EIS à l'échelle globale (figure 91) est de faire apparaître de grandes tendances par région du monde. En effet, à cette échelle et sur un format réduit comme c'est le cas dans un poster, la lecture de l'information « par pays » est inenvisageable.

L'objectif est plutôt de comparer les résultats des deux scénarii en ciblant les zones où ils introduisent de réelles différences. Ici, on voit que les résultats diffèrent assez peu sur l'ensemble du monde, il n'y a guère que sur le sud-est du continent asiatique (péninsule indienne essentiellement) que l'application des deux scénarii (« mitigation » versus « business as usual ») entraîne une disparité plus nette des résultats : la mortalité cardiovasculaire des sujets âgés de 15 ans et plus augmentera d'ici 2050, mais elle augmentera nettement plus, et sur une zone nettement plus étendue si aucune action n'est engagée pour faire baisser les niveaux de concentration en PM<sub>2,5</sub>.

On s'accordera à dire que ce type de représentation ne permet pas une lecture fine par pays et que l'objectif est d'avoir une idée des grandes tendances spatiales du phénomène étudié. On notera ici, comme dans l'exemple suivant (figure 92), la nécessité, pour pouvoir effectuer des comparaisons efficaces entre les deux cartes, d'utiliser strictement la même échelle de couleur et les mêmes seuils entre les différentes classes. C'est une règle incontournable quand on a un objectif de comparaison, qui ne pose pas réellement de problème ici, dans la mesure où l'on obtient des résultats finalement peu différents entre les deux scénarii, avec des ordres de grandeur très proches. Il peut arriver cependant que l'on souhaite comparer les distributions spatiales d'informations qui présentent des différences beaucoup plus nettes : étendues de distribution très disparates, ordres de grandeur très différents, etc. [9].

## I Figure 92 I

Cartographies extraites du poster “Quantifying health impacts of air pollution under a changing climate: from a global to the regional and local scales” – Victoria Likhvar, ISEE, 2013



Les cartographies (figure 92) ci-dessus comparent, de la même façon que les cartographies à l'échelle globale, les résultats de l'EIS en fonction du scénario mis en œuvre, cette fois, en appliquant l'équation à l'échelle « locale » pour le projet ACHIA, soit ici la région Ile-de-France. Les cartographies révèlent des différences dans ces résultats selon que des actions pour faire baisser les niveaux de concentration en  $PM_{2,5}$  seraient engagées ou non. En ceci, les résultats sont conformes à ce qui est observé à l'échelle globale. Cependant, à la lecture de ces deux cartes, les différences entre les deux scénarii ne sont pas très évidentes, un certain nombre de communes changent effectivement de classe et apparaissent en bleu plus foncé (donc avec une baisse de la mortalité cardio-vasculaire plus importante) avec le scénario « mitigation » qu'avec le scénario « business as usual », mais cette différence ne saute pas aux yeux. Le complément à la carte par le graphique qui explore les résultats sur l'ensemble de la zone et en l'éclatant en plusieurs sous-ensembles géographiques est nécessaire pour mieux percevoir cette différence, d'ampleur relativement constante quelle que soit la zone sur laquelle on l'étudie. On retrouve par contre sur ce graphique les codes couleurs (bleu pour le scénario de mitigation, vert pour le scénario « BAU - Business as usual ») utilisés au début du poster ce qui en facilite la lecture.

Par ailleurs, une explicitation plus claire de ce que recouvre géographiquement la distinction qui est faite entre « Paris/Paris suburbs/Rural areas » n'est pas superflue : le lectorat non francophone de ce poster présenté à ISEE 2013 était probablement peu familier de ces subtilités du découpage souvent opéré de la région parisienne.

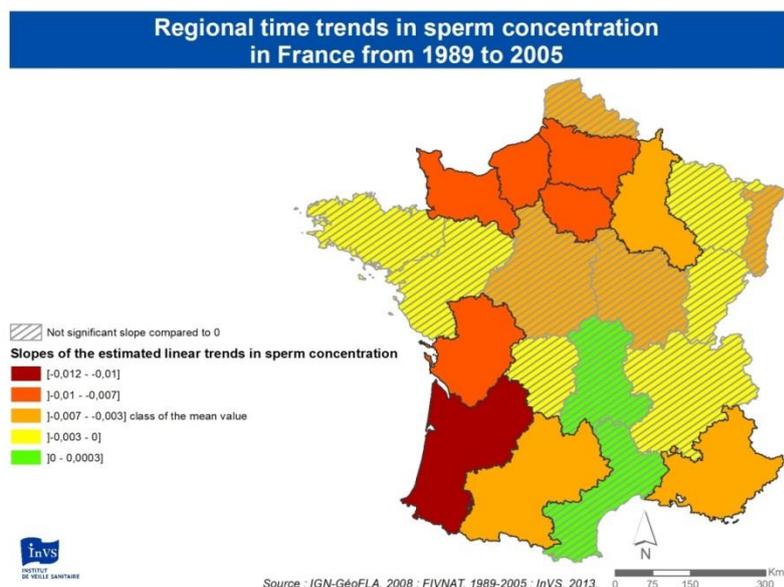
### 3.2.3 Représentation cartographique combinée d'une tendance et de la significativité de l'information

Dans le cadre des travaux menés sur l'évolution de la qualité du sperme en France à partir de la base de données Fivnat<sup>24</sup> [8], des cartographies ont été réalisées, avec une attention particulière portée à la significativité de l'information. En effet, les résultats des analyses laissent apparaître qu'en descendant à une échelle régionale, les tendances qui se dégagent pour les différents indicateurs étudiés ne sont pas toujours significatives ; par contre, leur valeur demeure une information en soi. Sur cette carte (figure 93) il a été choisi de distinguer les régions où les tendances ne sont pas significatives en les voilant d'une trame de gris. Des techniques comparables de « floutage » de l'information non significative ou trop incertaine pourraient être explorées pour améliorer la représentation de l'incertitude.

L'information en couleur, à savoir ici la pente de l'évolution de la concentration du sperme, représentée par des couleurs chaudes pour caractériser une baisse et par du vert pour une hausse, se trouve moins mise en avant que l'information sur la significativité.

I Figure 93 I

Tendances régionales de l'évolution de la concentration du sperme de 1989 à 2005



<sup>24</sup> Fivnat : Base de données nationale française de l'Assistance médicale à la procréation : l'association Fivnat a géré les statistiques de l'Assistance médicale à la procréation (AMP) en France de 1985 à 2005. Sur cette période, elle a enregistré, pour chaque tentative d'AMP réalisée dans des centres volontaires et répartis sur l'ensemble du territoire français métropolitain, une fiche d'information sur les couples concernés.

### 3.3 Représentation d'une démarche au moyen de schémas ou de tableaux synthétiques

La représentation d'une démarche ou d'un process ou encore des relations entre des événements par un schéma sort sans doute quelque peu du champ de la représentation des résultats *stricto sensu*, néanmoins, dans la mesure où elle permet souvent de rendre plus explicite et compréhensible une démarche, éventuellement à destination d'un public moins avisé, mais pas seulement, il semble utile de l'aborder dans l'objectif d'optimiser la communication sur certains travaux ou certaines démarches scientifiques, en particulier les plus complexes. Quelques exemples sont présentés ici, dans le cadre justement de travaux mettant en œuvre des process compliqués à comprendre pour un public non averti, ou bien dans l'objectif de résumer une démarche dans une présentation synthétique.

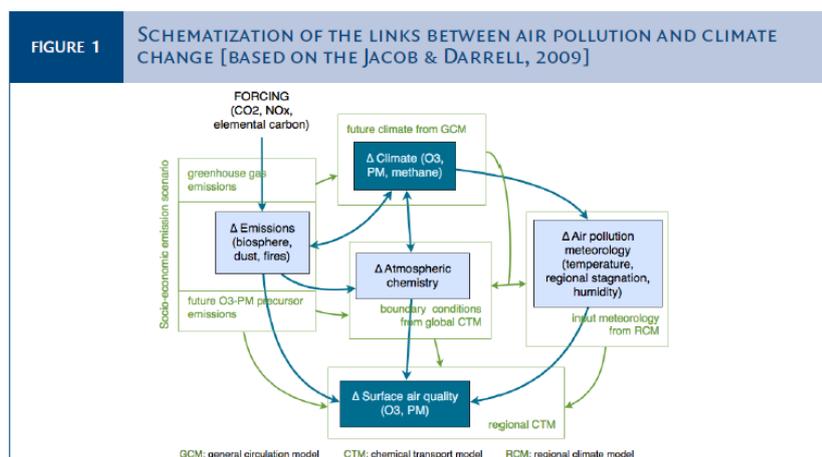
#### 3.3.1 Exemples de schémas utilisés pour résumer une démarche scientifique dans une EIS

La démarche des EIS nécessite souvent, lorsqu'est atteint le stade de la communication des résultats, une explicitation simple et accessible de la démarche en elle-même. Au-delà de la seule communication des résultats, la schématisation de la démarche est une étape en soi, un peu à la marge de la question des graphiques proprement dite, mais qui prend tout son sens à l'ère de la « datavizualisation » qui, bénéficiant d'outils numériques de plus en plus variés et performants, se développe très rapidement à l'heure actuelle.

Résumer dans un schéma une démarche scientifique ou un jeu de relations complexes

I Figure 94 I

Schéma extrait du poster "Quantifying health impacts of air pollution under a changing climate: from a global to the regional and local scales" – Victoria Likhvar, ISEE, 2013

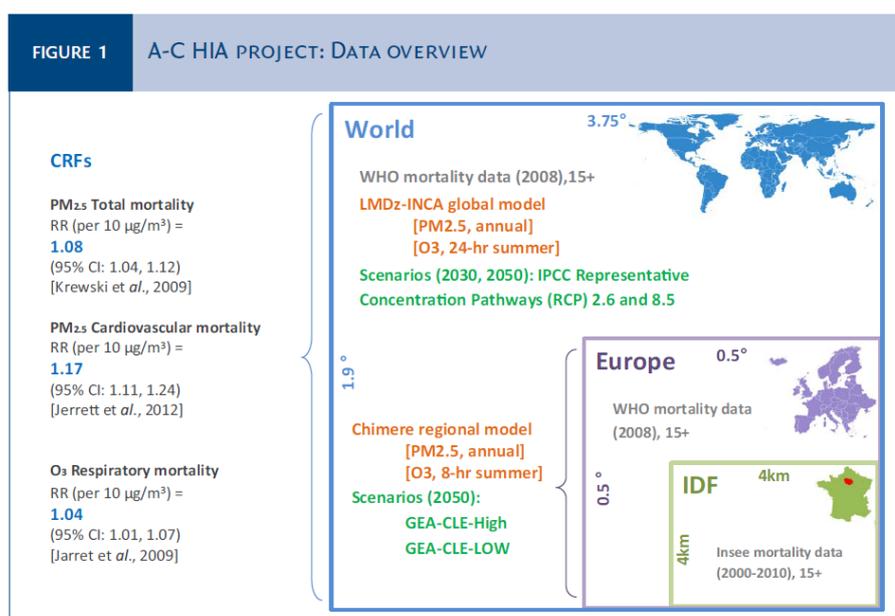


Ce premier schéma vise à illustrer la complexité du jeu de relations et des interconnexions entre la pollution de l'air et le changement climatique. Il est issu d'une source bibliographique externe<sup>25</sup>. Un bon schéma, quand il existe déjà, mérite d'être repris.

### Schématisation d'une démarche scientifique

#### I Figure 95 I

Schéma extrait du poster "Quantifying health impacts of air pollution under a changing climate: from a global to the regional and local scales" – Victoria Likhvar, ISEE, 2013



Ce schéma (figure 95) est intégré à un poster réalisé pour donner les premiers résultats du projet ACHIA, qui s'appuie sur des EIS intégrant des modélisations de l'évolution de la pollution de l'air en fonction du changement climatique, et qui travaille à trois échelles imbriquées : l'échelle globale, l'échelle « régionale » (c'est-à-dire l'échelle de l'Europe quand on la rapporte à l'échelle mondiale), et l'échelle « locale » (soit, la région Ile-de-France). Le schéma ci-dessous illustre l'emboîtement des trois échelles et les données mises en œuvre dans le projet. Les couleurs sont utilisées pour distinguer les divers éléments entrant en jeu dans le calcul des EIS : en gris, sont précisées les sources de données de mortalité ; en orange, les sources de données de pollution de l'air ; en vert, sont précisés les scénarii envisagés d'évolution de la pollution atmosphérique. La taille de la maille d'étude est indiquée pour chacune des trois échelles dans la couleur qui lui est liée (bleu pour l'échelle globale, violet pour l'échelle « régionale », et vert pomme pour l'échelle « locale ». En marge du schéma sont indiquées les trois constantes concentration/exposition/risque utilisées pour toutes les échelles de travail. Une réflexion importante a été menée pour améliorer la lisibilité de ce schéma, notamment sur les couleurs utilisées. Il en résulte un visuel qui résume sur un espace limité des informations nombreuses et relativement complexes qui auraient nécessité une quantité de texte élevée pour être précisées. Le nom de l'Ile-de-France aurait pu sans doute être indiqué en toutes lettres pour une meilleure compréhension d'un public non familier du découpage administratif français.

<sup>25</sup> <http://dash.harvard.edu/handle/1/3553961>

## Synthétiser par des tableaux un travail de revue de la littérature

Présenter un travail de revue de la littérature dans l'espace contraignant d'un poster comporte des difficultés liées notamment à l'espace restreint laissé au texte, et au fait que pour que le poster soit lu il n'est pas envisageable qu'il se résume à du texte seul dont une grande partie serait constituée de références bibliographiques. La dimension esthétique, pour une communication efficace, a réellement son importance et passe, dans ce type de support de communication, par l'utilisation de moyens visuels.

Le document présenté ici est en fait un tableau élaboré sur lequel des informations complémentaires sont apportées par l'utilisation d'un code couleur qui reprend une codification par lettre a-b-c de façon plus immédiate visuellement : les couleurs comme les lettres correspondent à un niveau de classification explicité dans le texte lui-même.

### I Tableau 5 I

Extrait du poster "A review of quantitative health impact assessments of ozone and particulate matter under a changing climate" – Victoria Likhvar, ISEE, 2013

TABLE 1		SUMMARY OF THE STUDIES QUANTIFYING THE HEALTH EFFECTS OF PM AND OZONE										
Type	Reference	Ref. N	Area	1960	1990	2000	2020	2030	2050	2060	2080	2100
<b>Ozone</b>												
b	West <i>et al.</i> 2006	[19]	World				Short-term: CVR					
b	West <i>et al.</i> 2007	[20]					Short-term: Total (non-acc.), CV, Respiratory					
b	Anenberg <i>et al.</i> 2012	[2]					Long-term: Respiratory					
c	Selin <i>et al.</i> 2009	[13]					Short-term: Total, RHA (65+), Respiratory SD, Minor RAD, Asthma "attacks", BU, LRS in children.					
b	Thompson & Selin 2012	[17]	North America				Short-term: BU, Minor RAD					
a	Bell <i>et al.</i> 2007	[3]					Short-term: Total (non-acc.), CV, Respiratory, COPD (65+), RHA (65+), Asthma (<65)					
a	Chang <i>et al.</i> 2010	[4]					Short-term: Total (non-acc.)					
a	Post <i>et al.</i> 2012	[11]					Short-term: Total (non-acc.), RHA, Minor RAD, SDL, Asthma ERV					
a	Tagaris <i>et al.</i> 2009	[15]					Short-term: Total (all ages), RHA (adults), Asthma ERV (all ages), Days of acute respiratory symptoms (18-64), SDL (5-17)					
a	Cheng <i>et al.</i> 2008	[5]					Short-term: Total					
c	Knowlton <i>et al.</i> 2004	[8]					Short-term: All internal causes					
c	Sheffield <i>et al.</i> 2011	[14]					Short-term: Asthma ERV					
c	Tagaris <i>et al.</i> 2010	[16]					Short-term: Total					
a	Anderson <i>et al.</i> 2001	[1]		Europe (EU)				Short-term: Total, RHA				
c	Kovats <i>et al.</i> 2008	[9]					Short-term: Total, RHA					
c	Heal <i>et al.</i> 2012	[7]					Short-term: All-cause, RHA					
c	Ornu <i>et al.</i> 2013	[10]					Short-term: RHA					
<b>PM</b>												
b	Anenberg <i>et al.</i> 2012	[2]	World				Long-term: All-cause, CP, Lung Cancer					
b	Saikawa <i>et al.</i> 2009	[12]					Long-term: Total					
a	Tagaris <i>et al.</i> 2009	[15]	North America				Short- & long-term: Total (30 and 2-12 months old), Onset of new cases of chronic bronchitis (27), CV HA (18-64 and 65+), Days of aggravation of existing asthma (6-18), Cases of acute bronchitis (8-12), Days with URS (9-11), Days with LRS (7-14)					
c	Tagaris <i>et al.</i> 2010	[16]					Long-term: Total					
a	Dias <i>et al.</i> 2012	[6]	EU				Short-term: All internal causes					
b	Wang & Mauzerall 2006	[18]	Asia				Long-term: Total					

CV, cardiovascular; CVR, cardiovascular and respiratory; RHA, respiratory hospital admission; SD, Symptom day; RAD, restricted activity days; URS, upper respiratory symptoms; LRS, lower respiratory symptoms; SDL, school days loss; ERV, emergency room visits; BU, bronchodilator usage; CP, cardiopulmonary.

- a. Studies examining the potential impact of climate change on air quality and health. The impact of PM or ozone emission mitigation measures are not taken into account (7 papers: [1;3-6;8;11;14-15]).
- b. Studies examining the potential impact of PM or ozone precursor emission mitigation measures on air quality and health. The impact of climate change is not taken into account (6 papers: [2;12;17-20]).
- c. Studies examining both, the potential impact of climate change and PM or ozone emission mitigation measures, on air quality and health (7 papers: [7-10;13-14;16]).

Ce tableau (tableau 5) vise également à catégoriser les références en fonction des éléments récurrents de l'EIS qu'elles abordent : pathologie cardio-vasculaire ou/et respiratoire, effets court-terme ou long-terme, polluant incriminé, zone géographique concernée, période d'étude, etc. L'intrication de ces classifications complique nécessairement le travail de synthèse en un seul document. Le résultat est relativement compréhensible si ce n'est que le tableau gagnerait peut-être en lisibilité si les catégories (a-b-c correspondant également aux codes couleurs) étaient explicitées dans le tableau lui-même ou bien juste en-dessous. Cette explicitation se trouve dans le corps du texte du poster lui-même et le lien avec le schéma est plus délicat à établir. De plus chaque information est représentée par un seul code (le polluant par un regroupement des lignes sous un bandeau violet, les dates pour les périodes d'études par une échelle chronologique sur les colonnes de la moitié droite du tableau, etc.), sauf la classification explicitée par ces trois lettres qui est reprise par les couleurs. Le choix de deux codes symbolisant la même information, même fondamentale, peut éventuellement induire des erreurs d'interprétation : si la légende n'a pas été clairement comprise, on n'associe *a priori* pas forcément visuellement ces deux codes à la même information.

## I Tableau 6 I

Tableau extrait du poster "A review of quantitative health impact assessments of ozone and particulate matter under a changing climate" – Victoria Likhvar, ISEE, 2013

TABLE 2		SUMMARY OF THE MODELS AND CLIMATE CHANGE AND EMISSION SCENARIOS USED IN THE STUDIES		
AQ Model	Scale	Resolution	Scenario	Reference
STOCHEM	Global	5° x 5°	IPCC SRES (2000): A2; IS92a	[1]
NASA GISS GCM: GISS-PUCCINI <sup>a</sup> (Shindell <i>et al.</i> , 2006)	Global	[4°lat. x 5°lon.] <sup>1</sup> [2°lat. x 2.5°lon.] <sup>2</sup>	IPCC SRES (2000): A2, A1B	[2 <sup>a</sup> ;3;8 <sup>1</sup> ;14; 15 <sup>1</sup> ;16 <sup>1</sup> ]
ECHAM-HAMMOZ <sup>1</sup> , ECHAM4 <sup>2</sup>	Global	2.8° x 2.8°	IPCC SRES (2000): A2, A1B	[2 <sup>1</sup> ;10 <sup>2</sup> ]
NCAR MATCH + RCA3 for downscaling	Global	50km	IPCC SRES (2000): A2, A1B	[10]
GEOS-CTM	Global	4°lat. x 5° lon.	IPCC SRES (2000): A1B	[13]
MOZART-2 (Horowitz <i>et al.</i> , 2003)	Global	1.9° x 1.9°	IPCC SRES (2000) A2; current legislation (CLE); maximum feasible reduction (MFR)	[12;19]
LMDz-INCA	Global	3.8°lon. x 2.6°lat.	IPCC SRES (2000) A2	[20]
HadCM3: HadAM3P <sup>a</sup> , HadRM3P <sup>b</sup>	Global <sup>1</sup> / Regional <sup>2</sup>	[3.8°lon. x 2.6°lat.] <sup>1</sup> [50km] <sup>2</sup>	IPCC SRES (2000): A2, A1B; CLE	[6 <sup>a</sup> ;9 <sup>b</sup> ;10]
EMEP4UK	Regional	50km → 5km	IPCC SRES (2000) A2; CLE + IPCC SRES B2; MFR+ IPCC SRES B2	[7]
CAMx (www.camx.com)	Regional	36km → {12km, 4km, 2km}	IPCC SRES (2000): A2	[17]
CHIMERE	Regional / Urban	50km, 10km	IPCC SRES (2000): A2	[6]
CMAQ + MM5 for downscaling	-	108km <sup>1</sup> → 36km <sup>2</sup> → 12km <sup>3</sup>	IPCC SRES (2000): A2, A1B; BAU, BACT, ACGT ( <i>based on energy technology improvements</i> )	[3 <sup>2</sup> ;8 <sup>1</sup> ;14 <sup>2</sup> ;15 <sup>2</sup> ; 16 <sup>1,2</sup> ;18 <sup>2,3</sup> ]

Deux autres tableaux sont ajoutés à ce poster, ils introduisent chacun des classifications des références selon un type d'information bien précis. L'un des deux tableaux vise à catégoriser les références en fonction des modèles et des scénarii d'évolution sur les émissions retenus (tableau 6). L'autre tableau a pour objectif de classer les références en fonction de la fonction concentration/réponse utilisée dans l'étude. Ces deux tableaux s'avèrent plus « rebutants » à lire du fait du nombre élevé d'informations chiffrées, de l'usage abondant d'acronymes dans la dénomination des scénarii et des modèles : ce type de tableau s'adresse à un lecteur très averti.

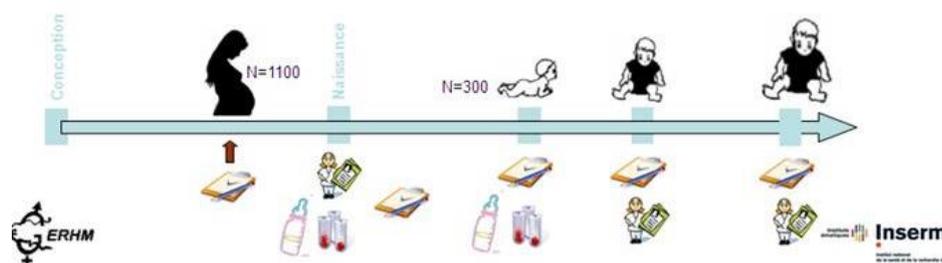
Par ailleurs, on notera l'usage impropre des exposants chiffrés pour le renvoi à une note de bas de page conjointement à l'abréviation de l'indication de mesure « km » pour kilomètre, car l'écriture « km² » correspond dans la nomenclature internationale des systèmes de mesure à l'écriture d'une surface. Il faudrait dans ce cas de figure utiliser un autre moyen pour renvoyer le lecteur à une note de bas de page.

### 3.3.2 Utilisation de schéma pour résumer une démarche scientifique en biosurveillance

Le schéma ci-dessous (figure 96) présenté dans le cadre d'une communication orale [4] détaille le déroulé d'une étude s'appuyant sur la cohorte mère-enfant TIMOUN : examens, questionnaires et prélèvements biologiques. Le choix des icônes symbolisant les différentes composantes de l'étude (questionnaires, prélèvements, examens, etc.) aurait pu être axé sur des symboles plus simples et immédiatement interprétables, mais pour le reste la lisibilité ne pose pas de problème. Les étapes se déroulant après la naissance auraient pu être datées plus précisément sur l'axe qu'elles ne le sont : on n'a ici comme repère que le nourrisson sur le ventre et deux enfants assis, un petit, et un légèrement plus grand, mais les deux seuls repères précis sur l'axe chronologique sont la conception et la naissance.

I Figure 96 I

Exemple d'un schéma utilisé pour résumer le déroulement d'une enquête



### 3.3.3 Utilisation de schémas en EQRS pour s'adresser au grand public

Exemple de la présentation faite en réunion publique sur les jardins familiaux d'Aubervilliers (93)

Le travail d'EQRS réalisé dans le cadre de cette demande d'appui de la Cire au DSE a fait l'objet d'un rapport et d'un avis détaillé transmis au décideur d'une part, et d'une présentation en réunion publique d'autre part. Un travail d'adaptation des contenus au public destinataire s'est avéré nécessaire, notamment pour la réunion publique, pour laquelle les quelques informations essentielles ont été résumées dans des schémas ou des visuels dont toute information superflue ou complexe avait été édulcorée.

En effet, les notions de VTR, d'analyse de sensibilité, l'existence de plusieurs scénarii possibles pour la réalisation d'une EQRS, etc. sont des éléments trop complexes qui brouilleraient excessivement la compréhension du message. Dans la présentation qui est faite au public, les informations transmises se limitent à :

- montrer dans quelle mesure il y a des dépassements (tableau 7) des normes dans les légumes du potager qui posent problème et pour quelles substances ;

## I Tableau 7 I

Mise en exergue de l'information d'importance dans un tableau à destination du grand public

	Nb	fréquence dépassement norme PLOMB	fréquence dépassement norme CADMIUM
Légumes fruits	Tomates	6	0/6
	courgettes	4	0/4
	aubergine	1	0/1
	citrouille	1	0/1
légumes racines	radis	1	0/1
	betterave	2	2/2
tubercules	Pomme de terre	3	2/3
légumes feuilles	blettes	3	2/3
	salade	1	1/1

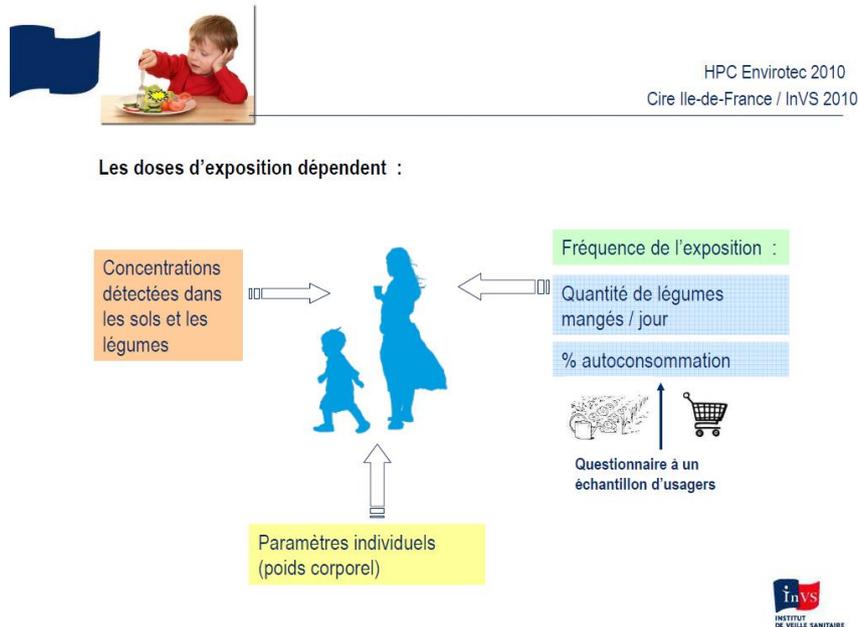
	Nb	fréquence dépassement norme PLOMB	fréquence dépassement norme CADMIUM
Légumes fruits	Tomates	6	0/6
	courgettes	4	0/4
	aubergine	1	0/1
	citrouille	1	0/1
légumes racines	radis	1	0/1
	betterave	2	2/2
tubercules	Pomme de terre	3	2/3
légumes feuilles	blettes	3	2/3
	salade	1	1/1

- résumer l'analyse de sensibilité dans un schéma (figure 97) pour montrer que la dose d'exposition est liée, notamment, au poids des individus et que les enfants sont de ce fait davantage susceptibles d'être concernés par des dépassements de dose ;

## I Figure 97 I

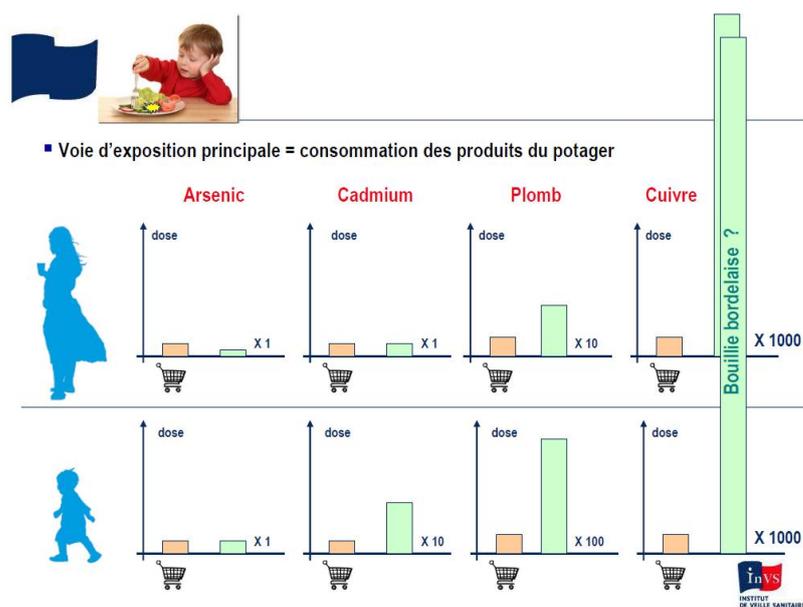
Schéma résumant l'analyse de sensibilité qui établit l'importance du poids corporel des individus dans le risque de dépassement de dose



- montrer en quoi la consommation des produits du jardin constitue la principale voie d'exposition (figure 98) en comparant les doses estimées par autoconsommation (figurées en vert) aux doses reçues par consommation de produits issus du commerce (figurées en orangé) ;

## I Figure 98 I

Les produits du jardin, principale voie d'exposition



- synthétiser le message explicitant quelle est la population et le contexte pour lequel le risque est le plus élevé (figure 99).

## I Figure 99 I

### Résumé de l'information d'importance pour le grand public



Les outils de représentation et de communication sont clairement différents selon le public qui est destinataire du message. Dans le cas de la présentation de résultats d'EQRS, cette attention au destinataire de l'information est sans doute primordiale dans la mesure où les concepts et les équations utilisés ne sont pas très maniables et nécessitent une certaine familiarité pour être manipulés aisément.

#### Résumer les étapes d'une étude à destination du grand public

Le schéma ci-dessous (figure 100) a été réalisé dans le cadre de l'enquête Elfe (Étude longitudinale française depuis l'enfance) afin de résumer, à l'intention du grand public, les étapes du suivi des enfants de la cohorte de la naissance à trois ans. Le schéma synthétise différentes informations : les personnages de couleur bleu ou mauve figurent les personnes qui sont concernées par l'étape qui est détaillée, les symboles de couleur verte illustrent les modalités d'enquête (entretien de vive voix à la maternité, ou au téléphone, ou au domicile, questionnaire courrier, ou courriel, etc.). Le détail des étapes est résumé dans un texte très court. Enfin, pour une bonne compréhension du public concerné, les dates des étapes sont signifiées à la fois par l'année de réalisation, et par l'âge de l'enfant au moment de la mise en œuvre de chacune de ces étapes. Les couleurs utilisées sont celles de la charte graphique d'Elfe. Ce schéma, réalisé à des fins de compréhension du public et de synthèse des étapes du suivi, ne permet pas de connaître précisément le contenu des informations recueillies lors de ces différentes étapes (certaines sont précisées : « questionnaire sur l'alimentation », « remise d'un capteur à poussières », mais d'autres non).

I Figure 100 I

Schéma résumant le déroulement du suivi de la cohorte Elfe de la naissance à trois ans<sup>26</sup>



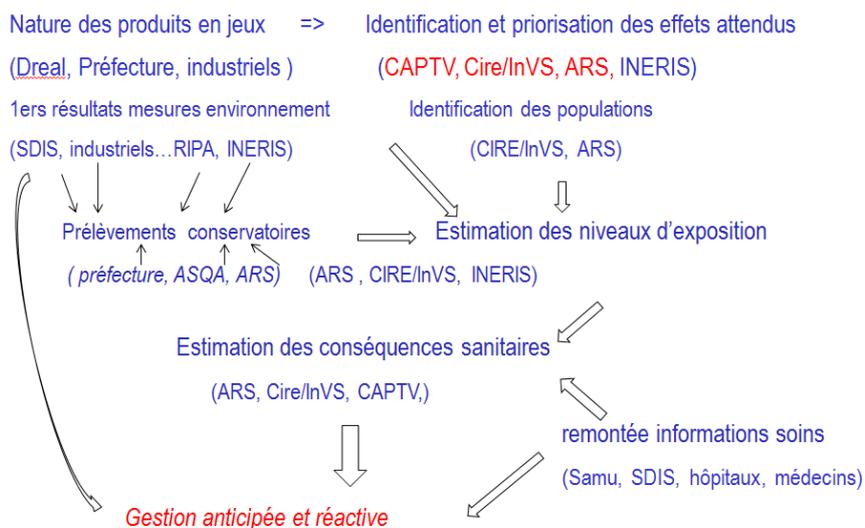
Représenter un circuit complexe de circulation de l'information

L'exemple ci-après (figure 101) illustre bien la difficulté qui peut se présenter lorsqu'il s'agit de faire comprendre rapidement et efficacement à un public plus ou moins averti la complexité d'un circuit de l'information entre des acteurs nombreux et plus ou moins aisément identifiables. Ici, il s'agit du partage de l'information dans un contexte d'urgence.

I Figure 101 I

Schématisation du partage de l'information en contexte d'urgence

Un partage d'informations en urgence indispensables à l'expertise sanitaire



<sup>26</sup> <http://www.elfe-france.fr/index.php/fr/comment-ca-marche/grandes-etapes>



## 4. Les nouvelles possibilités de visualisation des résultats

Le champ de la représentation graphique des résultats est en perpétuelle évolution. Il suffit pour s'en convaincre de suivre, même de loin, les nouvelles propositions qui émergent, puis circulent aujourd'hui à grande vitesse sur Internet. De nombreux blogs et sites sont consacrés aujourd'hui à ce nouveau champ que les Anglo-Saxons qualifient de data-visualisation et pour laquelle le Français n'a pas encore d'équivalent. La généralisation, dans de nombreux domaines, de l'open-data, la conviction, de plus en plus répandue, que le partage de données est plus productif que la rétention de l'information, et, enfin, l'explosion du big-data depuis quelques années amènent nécessairement à une accélération des développements dans ce domaine.

Toutes les représentations qui émergent ne sont pas adaptées à la présentation de résultats dans le domaine de la surveillance et de l'alerte en santé-environnement, loin s'en faut. Par ailleurs, nombre de ces nouvelles représentations s'appuient sur des technologies de communication qui ne sont pas celles qui sont couramment pratiquées dans les champs qui sont les nôtres. Enfin, nombre de ces représentations passent par un travail conséquent d'infographie à laquelle il n'est recouru, pour les travaux de l'InVS, que pour quelques cas de communications bien ciblées à destination du grand public.

Cependant, se tenir informé de ces émergences, de la multiplicité des formes qu'elles peuvent prendre, permet de rester à même d'améliorer les pratiques, et, peut-être, de parvenir à identifier, une méthode qui serait susceptible, à un moment donné, de mieux répondre à certains objectifs spécifiques de communication qui n'auraient pas été atteints jusqu'alors avec les outils habituels. Ici sont présentés quelques exemples de modes de représentation innovants, adaptés à des contextes particuliers, dont il faudrait tenir compte pour une éventuelle reprise sur des sujets différents.

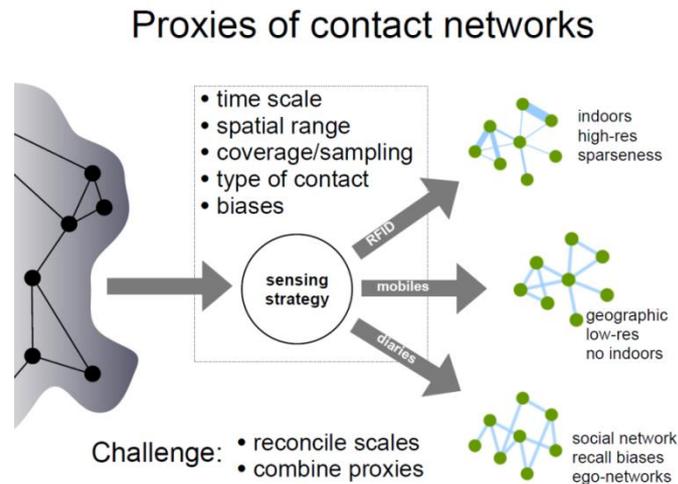
### 4.1 De nouvelles possibilités techniques et de nouveaux outils pour des données qui évoluent

La diversité des données collectées est en perpétuelle croissance, du fait, notamment, des moyens en constante évolution des moyens existant pour recueillir ces informations. Aujourd'hui, la généralisation des smartphones et autres outils connectés fait exploser le volume de données recueillies en permanence dans le monde. Ce chapitre ne présente que quelques exemples pris parmi une foison de possibilités, dans l'idée d'illustrer un tant soit peu cette diversité.

Les exemples ci-après illustrent les travaux spécifiques de l'équipe d'Alain Barrat [5;6] sur les réseaux de contacts humains. Ces travaux permettent notamment d'améliorer les modélisations de la transmission de certaines maladies infectieuses. Dans cette présentation, la complexité des informations recueillies sur les contacts humains donne lieu à des représentations spécifiques. Celle-ci peut porter sur la diversité des contacts mis en évidence selon le mode de recueil de l'information (figure 103).

## I Figure 103 I

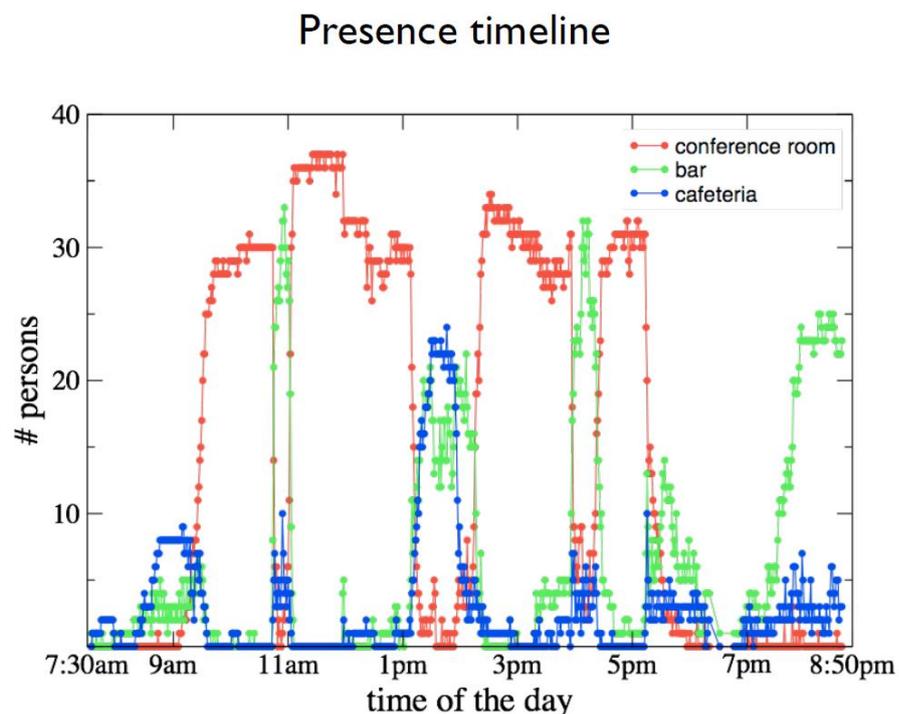
Représentation de réseaux de contacts mis en évidence selon les modalités du recueil de l'information mises en œuvre



Ce type d'étude sur les données de contacts entre individus peut également s'intéresser aux variations dans le temps (ici, au fil de la journée) de ces contacts et des lieux où ils s'opèrent (figure 104).

## I Figure 104 I

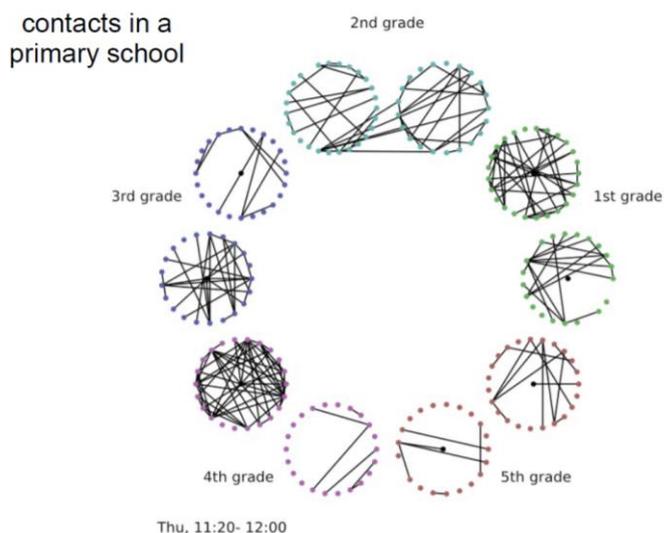
Présentation d'une étude de contacts au cours d'un congrès scientifique : analyse de la durée et du nombre de personnes impliquées dans les contacts en fonction du lieu et de l'heure de la journée



Une autre étude ayant porté sur les individus au sein d'une école permet de mettre en évidence les liens privilégiés qui subsistent entre les enfants d'une même classe même à l'heure de la récréation (figures 105 et 106). Le graphique 106 illustre le poids plus grand de la distinction entre les sexes chez les plus jeunes, distinction qui s'estompe par la suite.

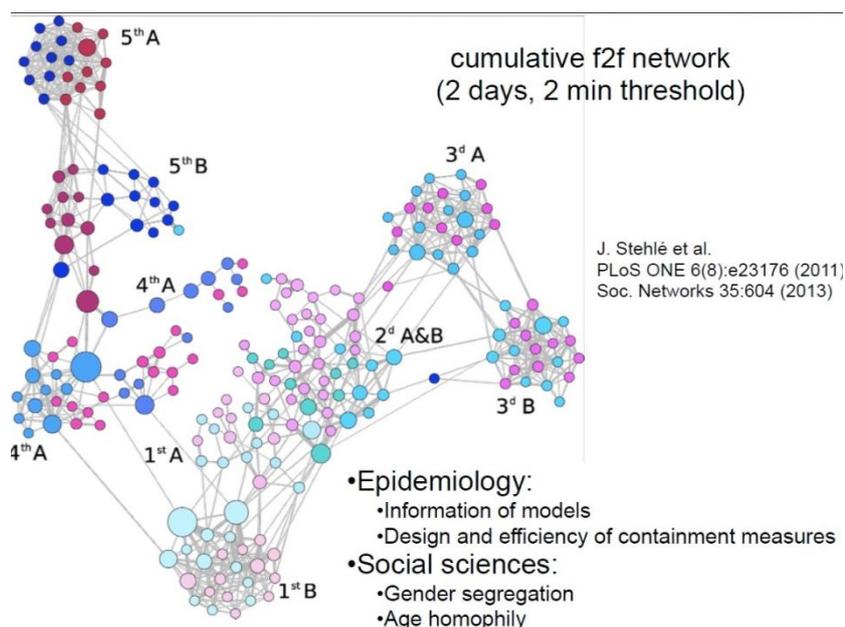
### I Figure 105 I

Analyse des contacts dans une école élémentaire : des contacts privilégiés entre enfants des mêmes classes



### I Figure 106 I

Autre représentation des contacts privilégiés entre individus des mêmes classes ou du moins des mêmes âges

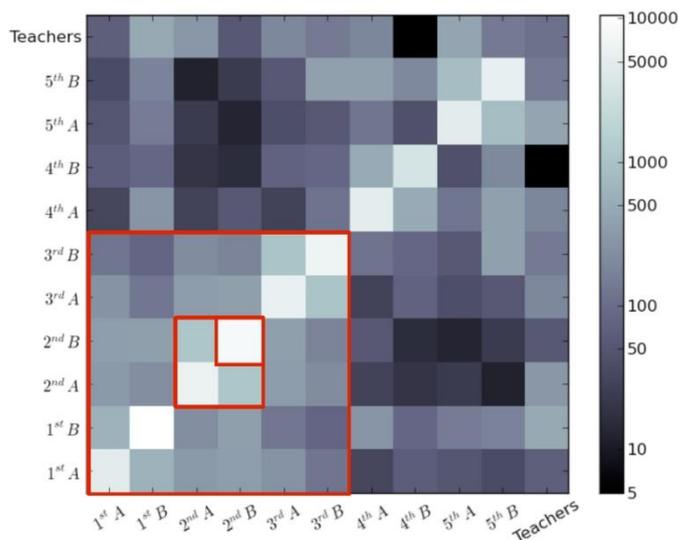


Ces contacts peuvent également être illustrés par des matrices (figure 107). Dans celle-ci, plus la case est blanche, plus le nombre de contacts renseignés est important : on voit là aussi qu'en dehors des enseignants, les contacts s'établissent préférentiellement entre les individus des mêmes classes.

I Figure 107 I

Représentation des contacts par une matrice (ici encore l'exemple d'une école)

### class contact matrix from contact behavior

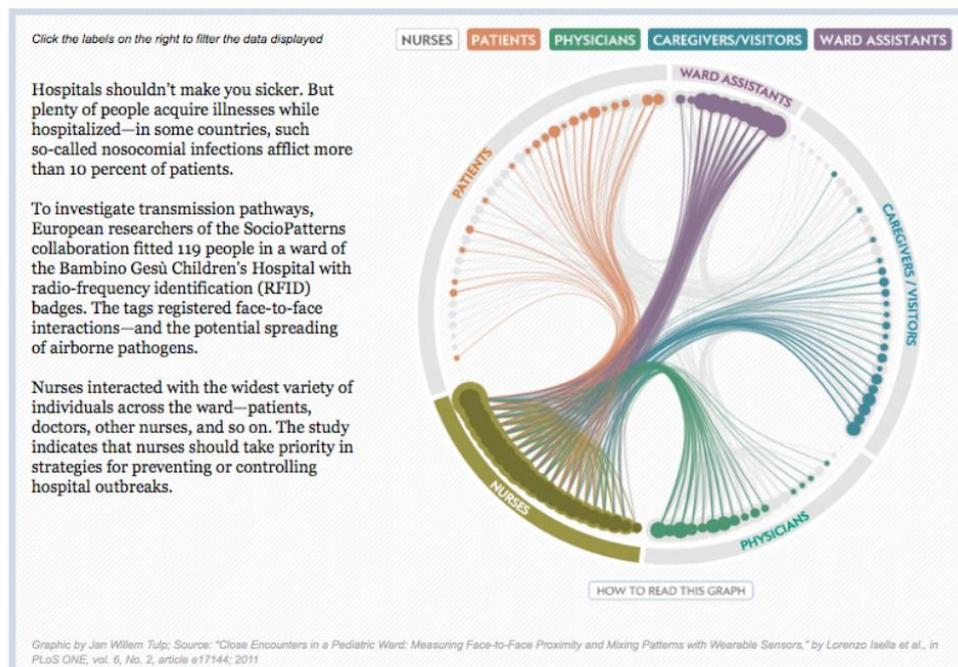


J. Stehle, et al.  
*High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School*  
 PLoS ONE 6(8), e23176 (2011)

Il existe d'autres modalités encore pour représenter la diversité et le nombre des contacts entre individus. L'illustration suivante, qui relève quasiment de l'infographie, représente les résultats d'une étude de contacts conduite entre individus travaillant ou usagers d'un service hospitalier. La densité des liens illustre le nombre de contacts entre les différentes catégories d'individus classées comme suit : les patients et les visiteurs, usagers de l'hôpital, les médecins, les infirmières et le personnel du service. La représentation graphique qui suit permet de voir que les infirmières sont en contact avec toutes les autres catégories d'individus, et ce de façon assez intensive, tandis que les médecins, par exemple, n'ont que peu de contact avec les usagers de l'hôpital, leurs contacts ayant surtout lieu avec les infirmières. C'est ainsi que l'étude pouvait avancer la conclusion que le personnel infirmier devait être particulièrement attentif aux mesures de prévention contre les maladies nosocomiales (figure 108).

## I Figure 108 I

Représentation proche de l'infographie illustrant les contacts entre catégories d'individus usagers/travaillant dans un service hospitalier



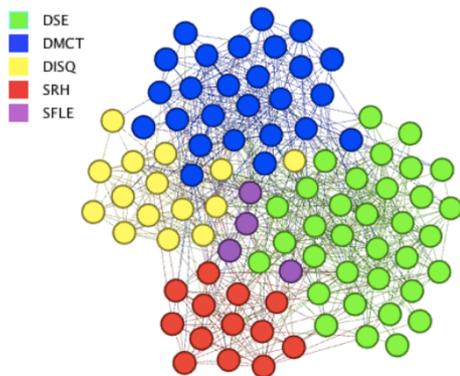
En juin 2013 a été menée une étude de contacts au sein du bâtiment Villermé de l'InVS. Les participants étaient munis de puces RFID<sup>27</sup> susceptibles d'enregistrer les contacts inter-individus dans le but de contribuer aux connaissances qui sont mises en œuvre dans la modélisation de la propagation des maladies infectieuses à transmission par voie respiratoire. Comme le montraient déjà de nombreuses études sur ce sujet, les contacts s'opèrent là encore essentiellement entre individus des mêmes catégories, les personnels se mélangeant finalement assez peu au cours de la journée, en dehors des agents du Service financier, logistique et économique (SFLE) qui, de par leur fonction dans le bâtiment Villermé (regroupement des agents de la « logistique ») sont amenés à être en contact avec davantage de monde. Les représentations ci-après illustrent les résultats des données collectées pour cette étude pendant deux semaines (figure 109).

La représentation par des points permet assez bien de visualiser le fait que les individus ont davantage de contacts au sein d'un même groupe (ici, les départements ou services d'origine des agents) qu'entre groupes et qu'il existe même des agents, en périphérie de ces groupes (sauf pour les agents de la logistique), qui développent très peu de contacts avec les autres groupes. Cependant, sur ce graphique par points, il est impossible de voir quelle est « l'intensité » de ces contacts (nombre, fréquence, durée). Cette information est plus lisible dans la représentation « en matrice » de cette même figure, alors que celle-ci ne permet par contre pas autant de faire apparaître le fait que certains sujets ont, par exemple, très peu de contacts, y compris au sein de leur groupe ; la spécificité des contacts des agents de la logistique est également moins évidente dans la représentation matricielle.

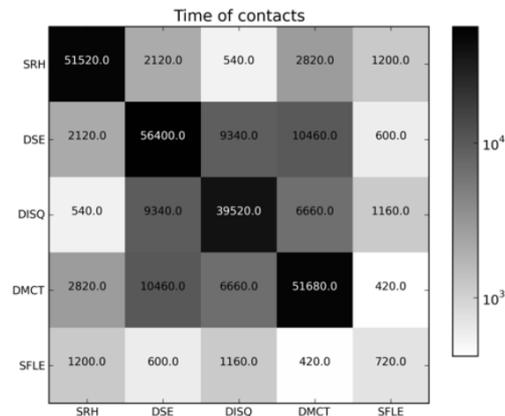
<sup>27</sup> RFID : Radio-Frequency Identification, technologie qui permet d'identifier un objet, d'en connaître les caractéristiques et le cheminement à distance grâce à une étiquette émettant des ondes radio fixée ou incorporée à l'objet.

Présentation des résultats de l'étude de contacts menée à l'InVS (Villermé, juin 2013)

### Two weeks aggregated data



Effectifs:  
 DISQ: 15/19  
 DMCT: 30/46  
 DSE: 38/60  
 SRH: 13/15  
 SFLE: 4/5



Non-homogeneous mixing  
 Strong homophily by department,  
 except for SFLE

## 4.2 Des modes de représentation à explorer dans le champ de la santé-environnement

Un travail est actuellement conduit au DSE sur les automatisations possibles des traitements statistiques et des représentations des résultats des analyses effectuées sur les données collectées dans le cadre de l'étude du volet périnatal de la cohorte Elfe. Une réflexion particulière est conduite dans ce cadre sur les possibilités d'automatisation et d'interactivité des résultats.

En effet, l'émergence de nouvelles techniques s'accompagne aussi de nouveaux outils et de nouveaux supports. La lecture de documents *via* le Web, facilitée aussi par la généralisation de l'Internet, sur ordinateur, tablettes ou smartphone, remet en cause le monopole de la publication papier ou PDF (qui s'apparente au papier numérisé). De nouveaux supports plus interactifs, remettant en cause le caractère très linéaire de la publication traditionnelle, permettent aussi des visuels plus attractifs et dynamiques que l'on ne peut envisager d'intégrer à une publication papier (ou PDF) traditionnelle.

Le logiciel R permet, moyennant quelques connaissances pour mobiliser des librairies javascript, de générer des graphiques de ce type. Le package ggplot2, qui offre de belles possibilités pour réaliser des graphiques traditionnels, s'interface ainsi avec javascript pour concevoir simplement des graphiques dynamiques intégrables dans une page web au format HTML<sup>28</sup>. Le package rCharts permet également d'utiliser de nombreuses librairies javascript<sup>29</sup>. La librairie javascript d3 est actuellement une des plus performantes pour ajouter de l'interactivité et dynamiser ses graphiques. C'est d'ailleurs la librairie la plus couramment utilisée pour la constitution des infographies interactives de référence que publie le New York Times dans son édition Web<sup>30</sup>. Les graphiques sont, *a minima*, exportés sous forme de fichiers images, au format « gif » le plus souvent car c'est un format léger qui peut intégrer la dimension dynamique pour l'insertion dans un document, sachant que dans le cas des représentations dynamiques, l'interactivité n'est possible qu'en restant en HTML.

Les exemples de ces graphiques et représentations d'un nouveau genre sont difficilement intégrables dans un rapport papier aussi il est recommandé, pour visualiser davantage d'exemples, de consulter les très nombreuses ressources qui existent sur le sujet<sup>31</sup>.

---

<sup>28</sup> <http://www.r-bloggers.com/make-your-ggplots-shareable-collaborative-and-with-d3/>

<sup>29</sup> Exemples : <http://ramnathv.github.io/rCharts/>

<sup>30</sup> <http://www.smallmeans.com/new-york-times-infographics/>  
<http://blog.visual.ly/10-things-you-can-learn-from-the-new-york-times-data-visualizations/>

<sup>31</sup> On trouve des exemples et recommandations pour l'optimisation des représentations : <http://eagereyes.org/about>

Il existe également de très beaux exemples de visualisations très régulièrement mis à jour : <http://www.visualisingdata.com/>

Enfin, on trouve des exemples de vidéos, mais qui nécessitent un haut niveau de programmation pour les mettre en œuvre : <http://www.thefunctionalart.com/>

Et pour finir, R permet également par la mobilisation de librairies javascript de réaliser des cartographies dynamiques : <http://rmaps.github.io/>

## 4.3 Des modes de représentation spécifiques liés à l'émergence du « big data »

Le big data, qui se traduit littéralement par « grosses données », et qu'il semble plus recommandé de traduire par mégadonnées ou données massives désigne des ensembles de données dont le volume atteint de telles proportions qu'il en devient difficile de les exploiter avec des outils classiques de gestion et de traitement de bases de données. L'émergence de ces big data s'accompagne de celle de nouveaux ordres de grandeur qui affectent aussi bien la capture, le stockage, la recherche, le partage, l'analyse et le traitement de ces données, que la représentation et la communication de résultats. Avec le big data, ces volumes atteignent plusieurs téraoctets par jour (en 2013 on estimait par exemple que les réseaux sociaux généraient environ une dizaine de téraoctets d'information par jour pour Twitter, un peu moins pour Facebook). Mais les volumes les plus importants restent générés par les installations scientifiques (50 téraoctets de données analysées par jour générées par le radiotélescope « Square Kilometer Array », par exemple). Outre le volume, c'est aussi la diversité des données et des relations qui existent entre elles qui caractérisent ce big data. Enfin, le big data, de par les volumes qu'il induit, nécessite des vitesses de traitement toujours plus importantes qui mettent au défi les systèmes de stockage (les « data centers »), de traitement et de circulation de ces informations.

L'émergence de ce big data concerne de très nombreux domaines, mais la santé et l'épidémiologie sont particulièrement concernées. Les liens ci-après permettent de visualiser des présentations d'exploitations du big data, notamment dans le champ de l'épidémiologie<sup>32</sup>.

L'émergence du big data, en amenant de nouvelles possibilités mais aussi de nouveaux besoins, est probablement un moteur non-négligeable dans le développement de nombreuses alternatives dans le champ de la data-visualisation et de la représentation graphique.

## 4.4 Les limites des nouvelles possibilités

Il existe néanmoins des limites, assez nombreuses, à ces nouvelles possibilités. Elles sont de nature diverses.

Pour commencer, la qualité de ces représentations est souvent très « logiciel-dépendante ». En effet, on ne pourra pas produire des résultats similaires avec R, Excel, SAS, Stata, etc. Il convient de rester très vigilant quant aux capacités d'un logiciel à produire effectivement la sortie graphique souhaitée dans le respect des règles qui y préside. Enfin, à moins de maîtriser un large panel d'outils, l'auteur de graphique en recherche d'originalité se heurtera assez souvent aux limites de son outil de prédilection, surtout s'il n'est pas à l'aise avec les logiciels qui nécessitent d'être un peu à l'aise avec le développement et le code.

Par ailleurs, même en maîtrisant bien un outil, il est parfois malaisé d'obtenir un résultat visuellement esthétique et attrayant. C'est là aussi qu'est le pas à franchir entre le graphique « brut » sorti du logiciel, et l'infographie qui, en retraitant le visuel, tente d'y apporter une touche d'esthétique visant à le rendre plus attrayant. Plus attrayant, mais aussi parfois, plus faux. Car il n'est pas rare en la matière de rencontrer des représentations qu'un traitement par l'infographie a déformées et de ce fait, faussées. L'infographie, bien que souvent très esthétique, n'introduit donc pas toujours que du « joli » dans l'univers éventuellement un peu austère du graphique.

---

<sup>32</sup> [http://www.ted.com/talks/larry\\_brilliant\\_wants\\_to\\_stop\\_pandemics.html](http://www.ted.com/talks/larry_brilliant_wants_to_stop_pandemics.html)  
[http://www.ted.com/playlists/56/making\\_sense\\_of\\_too\\_much\\_data.html](http://www.ted.com/playlists/56/making_sense_of_too_much_data.html)

Enfin, sans même introduire d'erreurs comme on le voit parfois, certaines solutions alternatives, en tentant de rendre la représentation plus attrayante, peuvent éventuellement arriver à un résultat à l'opposé de son objectif, et à brouiller le message, ou encore à mettre l'accent sur un aspect qui est en fait loin de constituer l'essentiel du message. De ce fait, le destinataire de la représentation est susceptible de ne pas recevoir le message initialement souhaité par l'auteur.

Il y a donc certainement un dosage subtil à atteindre, entre attrait visuel du graphique et sobriété, celle-ci sera bien souvent la clef majeure d'une bonne communication et d'une bonne compréhension du message sous-jacent à un graphique. Cette sobriété n'est pas impossible à faire coexister avec une certaine esthétique, mais le pas qui privilégie l'attrait visuel sur la simplicité du message est vite franchi. Elle demeure sans aucun doute un gage de bonne compréhension d'un graphique, quel qu'il soit.

## 5. Conclusion

L'époque actuelle voit exploser la diversité des possibles en matière de représentation des résultats. Notre œil est ainsi quotidiennement confronté à cette multiplicité de représentations. Il apprend à les décoder, il se trompe aussi sûrement très souvent. C'est pourquoi prendre le temps d'une réflexion sur ce sujet, et d'un rappel des bonnes pratiques en la matière s'avérerait particulièrement pertinent. La tentation est grande de suivre les médias contemporains en allant vers toujours plus de visuel, d'esthétique. Elle ne doit pas, si l'on y succombe, faire oublier le message qui préside à la réalisation d'un graphique : que souhaite-t-on montrer des résultats ou des données que l'on représente, à quelles fins, et à qui l'on s'adresse.

Cette réflexion, en se confrontant à la multiplicité des champs couverts en santé environnement, à la diversité des thématiques et des méthodes scientifiques mises en œuvre, prend tout son sens. La question du public destinataire a été ici rapidement abordée, car elle doit impérativement être posée au moment du choix d'une représentation graphique au détriment d'une autre, puisqu'en effet, certains graphiques font appel à des notions, statistiques par exemple, qui ne sont pas forcément à la portée du plus large public, et à l'inverse, un lectorat constitué d'experts en pleine possession des concepts afférents à une thématique ou à une méthodologie n'aura pas besoin d'un rappel sur ces prérequis de leur champ de compétences. La question, abordée ici assez rapidement, mérite néanmoins d'être approfondie et un travail spécifique sur ce sujet est prévu dans les suites à donner à ce guide.

## Références bibliographiques

- [1] Tufte E. *The Visual Display of Quantitative Information*. 2<sup>nd</sup> ed. Cheshire : Graphic Press ; 2001. 51p.
- [2] Bertin J. *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes*. 3<sup>e</sup> éd. Paris : EHESS ; 1999. 431 p.
- [3] Lebel B. *Convaincre avec des graphiques efficaces*. Paris : Eyrolles ; 2012. 257 p.
- [4] Guldner L. Apports alimentaires en contaminants chimiques et nutriments, impact sur l'issue de la grossesse [Thèse d'épidémiologie dirigée par Cordier S]. Paris : Université Paris XI ; 2008. 246 p.
- [5] Stehlé J, Charbonnier F, Picard T, Cattuto C, Barrat A. Gender homophily from spatial behavior in a primary school: a sociometric study. *Social Networks* 2013;35(4): 604-13.
- [6] Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, *et al.* Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* [en ligne] 2013 [consulté le 29/08/2014];8(9):73970. Disponible à partir de l'url : <http://www.cpt.univ-mrs.fr/~barrat/journal.pone.0073970.pdf>
- [7] InVS, Anses. Groupe scolaire des Bourdenières de la commune de Chenôve (21 300) : Évaluation des risques sanitaires liés aux composés de traitement du bois. Rapport d'expertise. Comité scientifique et technique [en ligne]. Maisons-Alfort : Anses ; 2009. 234 p. [consulté le 27/08/2014] Disponible à partir de l'url : <https://www.anses.fr/sites/default/files/documents/AIR2007et9900Ra.pdf>
- [8] Le Moal J, Rolland M, Gorla S, Wagner V, de Crouy-Chanel P, Rigou A, *et al.* Semen quality trends in French regions are consistent with a global change in environmental exposure. *Reproduction* 2014;147 (4) 567-74.
- [9] Stempfelet M, de Crouy-Chanel P. La cartographie : quelques règles de représentation des données géographiques [en ligne]. Saint-Maurice : InVS ; 2010 ; 26 p. [consulté le 02/09/2014] [Document à usage interne]. Disponible à partir de l'url : [http://intranet\\_vosespaces/dse/SIG/Documents%20partages/Documents%20méthodologiques/Guide%20des%20bonnes%20pratiques%20en%20cartographie.pdf](http://intranet_vosespaces/dse/SIG/Documents%20partages/Documents%20méthodologiques/Guide%20des%20bonnes%20pratiques%20en%20cartographie.pdf) (InVS-Saint-Maurice)  
<http://vosespaces.invs.sante.fr/dse/SIG/Documents%20partages/Documents%20méthodologiques/Guide%20des%20bonnes%20pratiques%20en%20cartographie.pdf> (Cire)
- [10] Laaidi K. Rôle des îlots de chaleur urbains dans la surmortalité observée pendant les vagues de chaleur - Synthèse des études réalisées par l'Institut de veille sanitaire sur la vague de chaleur d'août 2003 [en ligne]. Saint-Maurice : Institut de veille sanitaire ; 2012. 4 p. [consulté le 28/08/2014] Disponible à partir de l'url : [http://opac.invs.sante.fr/doc\\_num.php?explnum\\_id=8058](http://opac.invs.sante.fr/doc_num.php?explnum_id=8058)
- [11] Laaidi K, Zeghnoun K, Doussot B, Bretin P, Vandentorren S, Giraudet E, *et al.* The Impact of Heat Islands on Mortality in Paris during the August 2003 Heat Wave. *Environ Health Perspect* [en ligne] 2011 [consulté le 28/08/2014];120:254-9. Disponible à partir de l'url : <http://ehp.niehs.nih.gov/1103532/>
- [12] Anses. Formamide : Usages dans les produits de consommation et évaluation des risques sanitaires liés aux jouets en mousse « tapis puzzle ». Avis et rapport d'expertise collective. Maisons-Alfort : Anses ; 2011. 241 p. Disponible à partir de l'url : <https://www.anses.fr/sites/default/files/documents/CHIM2010sa0302Ra-2.pdf>
- [13] Dhée F. Cartographie pour les déficients visuels de la couleur. Propositions d'amélioration des cartes pour les daltoniens [Thèse dirigée par Palsky G. et Ruas A]. Paris : Paris Sorbonne, 2013.

[14] Dhée F. Amélioration des cartes topographiques pour les daltoniens. 25<sup>th</sup> International Cartographic Conference (ICC'11); 3–8 juillet 2011, Paris (France). [consulté le 28/08/2014] Disponible à partir de l'url : [http://icaci.org/files/documents/ICC\\_proceedings/ICC2011/Oral%20Presentations%20PDF/E3-Mapping%20for%20color-blind%20or%20blind%20users/CO-469.pdf](http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/E3-Mapping%20for%20color-blind%20or%20blind%20users/CO-469.pdf)

[15] Béguin M, Pumain D. La Représentation des données géographiques. Paris : Armand Colin ; 1996. 191 p.

[16] Bentayeb M, Stempfelet M, Wagner V, Zins M, Bonenfant S, Songeur C, *et al.* Retrospective modeling outdoor air pollution at a fine spatial scale in France, 1989-2008. *Atmospheric Environment* 92 (2014) 1-13.

## Ressources complémentaires accessibles sur Internet

Elles ne peuvent évidemment être énumérées de façon exhaustive. On trouvera ici un certain nombre de ressources susceptibles de compléter la lecture de ce document et d'assurer une veille sur les outils et les méthodes de représentation et de visualisation qui sont en perpétuelle évolution.

### Documentation en anglais :

<http://www.visual-literacy.org/>

<http://www.r-bloggers.com>

<http://rgraphgallery.blogspot.fr/> - Exemples de code pour réaliser des représentations graphiques dans R.

<http://excelgraphs.blogspot.fr/> - Ressources pour réaliser des graphiques en mobilisant des ressources R dans Excel, via le téléchargement ici des ressources implémentées dans Excel : <http://rcom.univie.ac.at/download.html#RExcel>

<http://labs.juiceanalytics.com/chartchooser/index.html> - Ressources pour réaliser des graphiques simples avec Excel ou Powerpoint, en choisissant le mode de représentation en fonction de ses objectifs.

<http://docs.ggplot2.org/current/>

<http://eagereyes.org/about> - Ressources et recommandations pour améliorer ses représentations.

Quelques recommandations pour l'optimisation des représentations :

<http://jknowles.github.io/Jan2013SDPtalk/#39>

Sur l'utilisation de la couleur :

[http://icaci.org/files/documents/ICC\\_proceedings/ICC2011/Oral%20Presentations%20PDF/E3-Mapping%20for%20color-blind%20or%20blind%20users/CO-469.pdf](http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/E3-Mapping%20for%20color-blind%20or%20blind%20users/CO-469.pdf)

<http://www.personal.psu.edu/cab38/ColorSch/Schemes.html>

<http://www.infovis.net/printMag.php?num=183&lang=2>

<http://sparklines-excel.blogspot.fr/>

<http://www.visual-analytics.eu/> et notamment : <http://www.visual-analytics.eu/2010/12/200countries200years4minutes/>

<http://www.scoop.it/t/r?page=2>

<http://dd.dynamicdiagrams.com/> et notamment : <http://dd.dynamicdiagrams.com/2008/01/nightingales-rose/>

**Documentation en français :**

<http://www.datavisualization.fr/blog/>

<http://www.spatial-modelling.info/Darcy-2-module-de-comparaison> (anamorphoses)

Réflexions sur l'utilisation de la couleur en cartographie :

<http://couleurs.hypotheses.org/>

Publications de l'équipe de recherche AVIZ de l'Institut national de recherche en informatique et automatique (Inria) :

[http://www.inria.fr/equipes/aviz/\(section\)/publications](http://www.inria.fr/equipes/aviz/(section)/publications)

<http://www.inria.fr/centre/saclay/recherche/vismaster-un-projet-remarque>

# Annexes

## Annexe 1 : Ressources R pour la génération d'un Mosaic Plot

Les données sont formalisées de la manière suivante dans une table :

	FILIERE2stat2	Sexe_Etude2
1	SUR	Homme
2	SUR	Homme
3	SOC	Femme
4	SUR	Homme
5	SUR	Homme
6	SUR	Homme
7	SUR	Homme
8	SUR	Homme
9	SUR	Homme
10	ADM	Femme
11	SUR	Homme
12	SUR	Femme
13	SOC	Homme
14	SUR	Homme
15	SUR	Homme

Displayed 1000 rows of 41,942 (40,942 omitted)

Puis on fait tourner le code suivant :

```
library(vcd)

base<-read.table("mosaic08b.CSV",header=T,sep=",")

#Mise en forme du code des variables

sex<-factor(base$Sexe_Etude2,level=c("Femme","Homme"))

filstat<-
factor(base$FILIERE2stat2,level=c("ADM","DIR","SOC","SUR","TEC","Retr
aité","Sorti"))

sb3b<-structable(filstat,sex)

mosaicplot(sb3b,xlab = 'Filière ou statut', ylab = 'Sexe',

main = '',cex.axis = 0.575 ,color=c('Green','Orange'),las=3)
```

## Annexe 2 : Ressources R pour la génération d'un graphique en Tree Map

```
library(treemap)

treemap(calcul_pa_epicar2, index=c("geh_tt"),
vSize=c("pa"),title="Répartition des personnes-années par
GHE",fontsize.labels=c(30,20,12))
```

## Recommandations pour la représentation des résultats au DSE

Le Département santé environnement (DSE) de l'Institut de veille sanitaire (InVS) produit de nombreux travaux scientifiques dont les résultats sont diffusés auprès d'un public très varié. Ce guide propose un rappel des bonnes pratiques, une aide au choix du mode de représentations visuelles le plus adéquat, ainsi qu'une présentation des nouvelles possibilités de représentation et de visualisation. Il vise à harmoniser et enrichir les représentations, graphiques ou non, des données et résultats produits au DSE, afin de les rendre plus informatifs et plus accessibles, de les adapter au mieux aux publics auxquels ils s'adressent, et de faciliter ainsi leur utilisation par les principaux partenaires de l'InVS.

**Mots clés :** graphique, méthodologie, santé environnement, guide

## Representing data and results in the Department of Environmental Health: advices and good practices

*The Department of Environmental Health (DSE) of the French Institute for Public Health Surveillance (InVS) produces numerous studies whose results are disseminated to a large and various public. This work intends to be a reminder of good practices, a decision support tool to choose the most appropriate way of representing data and results. It also intends to present more recent and alternative possibilities which are continuously emerging in representing data and visualization. It aims to harmonize and enrich the practices in graphic or not graphic representations of the results produced by the DSE. It eventually pursues to make these data and results more informative and understandable, to adjust them to the public and to facilitate their use by the different partners of the InVS.*

Citation suggérée :

de Crouy-Chanel P. Recommandations pour la représentation des résultats au DSE. Saint-Maurice : Institut de veille sanitaire ; 2015. 121 p. Disponible à partir de l'URL : <http://www.invs.sante.fr>

### INSTITUT DE VEILLE SANITAIRE

12 rue du Val d'Osne  
94415 Saint-Maurice Cedex France  
Tél. : 33 (0)1 41 79 67 00  
Fax : 33 (0)1 41 79 67 67  
[www.invs.sante.fr](http://www.invs.sante.fr)

ISSN : 1958-9719  
ISBN-NET : 979-10-289-0142-4  
Réalisé par Service communication – InVS  
Dépôt légal : juillet 2015