

# ADV30\_TRANSCRIPT - Éviter l'encombrement visuel

*Johanie Fournier, agr*

*2019-12-07*

TEASER: Tous les éléments qu'on ajoute à une page blanche augmentent la charge cognitive du lecteur. Sachant ça, c'est plus qu'important de faire du ménage dans les éléments qu'on laisse dans notre visuel et surtout de s'assurer que chaque élément qui reste soit utile pour faire passer le message.

INTRODUCTION: Ici Johanie Fournier et bienvenue à un nouvel épisode d'Agriculture, Données et Visualisation. Le podcast où je vous apporte avec moi dans le processus de traitement et de visualisation de données pour apprendre à présenter vos propres données de la manière la plus efficace possible. Sans plus tarder, voici l'épisode de cette semaine.

Bonjour et bienvenue dans ce 30e épisode. J'espère que vous allez bien! Avant de débiter, j'aimerais vous dire que c'est mon dernier épisode avant Noël. Je prends une petite pose pour la période des fêtes et je vous reviens le 20 janvier avec un nouvel épisode. Alors aujourd'hui, on regarde les contraventions de stationnement qui ont été données à Philadelphie en 2017.

Les données de cette semaine proviennent du Open Data Philly. Le lien pour y avoir accès se trouve dans mon article de blogue au [johaniefournier.com/tyt2019w49](http://johaniefournier.com/tyt2019w49)

Donc, cette semaine on dispose d'une base de données de plus de 1 million de lignes d'informations. Chaque ligne nous renseigne sur la raison de la contravention, le moment (date et heure) auquel la contravention a été donnée, l'autorité qui l'a émise, le montant et des informations de géolocalisation (coordonnées gps)

Alors, après avoir regardé ces données je me suis fixé comme objectif de présenter l'évolution du prix moyen des contraventions par autorité. J'étais curieuse de savoir si le prix variant par autorité, mais aussi de savoir qui émet les tickets les plus chers.

CHOIX DU TYPE DE VIZ: J'ai choisi de présenter ces données avec des lignes, tout simple de même. En voulant présenter l'évolution dans le temps, c'est certain que j'ai choisi de travailler avec des données chronologiques cette semaine. Donc, les lignes ou les courbes sont un des choix à notre disposition dans ces cas là. C'est possible de présenter des données chronologiques avec d'autres types de graphiques, comme par exemple des histogrammes. Par contre, il y a 10 autorités qui ont donné des contraventions à Philadelphie en 2017, donc ça complique un peu les choses dans un histogramme.

PRÉSENTER LE GRAPHIQUE: Donc, j'ai utilisé des lignes pour faire mon visuel de cette semaine et je ne veux pas trop expliquer maintenant le processus de réflexion qui a mené à la création de mon graphique parce que, en deuxième partie d'épisode, je vais justement vous expliquer comment j'ai désencombré le graphique de base pour obtenir mon visuel.

ADD: Hey! tu travailles avec R et ça t'intéresse de voir le code que j'ai utilisé pour nettoyer et visualiser mes données? Va voir dans les notes de cet épisode, j'ai mis un lien vers l'article de blogue dans lequel tu pourras trouver tous les détails dont tu as besoin.

REVOIR LES RÈGLES D'OR DE LA DATAVIZ: Alors, j'ai mis les 3 versions des graphiques que j'explique ici dans les notes de l'épisode que tu pourras trouver au [johaniefournier.com/adv30](http://johaniefournier.com/adv30).

Donc, pour montrer l'évolution dans le temps du prix moyen des contraventions par agence. J'ai commencé par créer une base de données qui fait la moyenne, par mois, du prix moyen des tickets par agence. Avec cette base de données synthétisée, j'ai créé un graphique montrant les mois de 2017 sur l'axe des x et le prix moyen sur l'axe des y. Dans mon graphique, chacune des 10 lignes représente une autorité différente.

Le premier graphique que j'ai mis dans les notes de l'épisode est le modèle de graphique qu'on obtient par défaut dans R. Il y a beaucoup d'éléments dans ce graphique qui doivent être revus pour améliorer l'expérience de notre lecteur. Le visuel que j'ai créé cette semaine est le 3e présenté dans les notes. C'est ces deux versions que je vais comparer ici pour vous expliquer ma démarche.

D'abord, je ne suis pas fan d'utilisé des fonds de graphique de couleur. Je ne parle pas ici de l'arrière-plan du visuel, je fais souvent des visuels qui ne sont pas sur fond blanc et cette semaine mon visuel final est sur fond gris foncé, je parle ici de mettre la zone de graphique d'une couleur différente de l'arrière-plan du visuel. Ça ajoute de l'encre inutilement sur le visuel. C'est certain que les lignes blanches sur le fond gris permettent de créer une grille qui guide la lecture des valeurs, mais encore une fois j'ai réussi à le faire sans ajouter un encadrement supplémentaire sur le visuel.

C'est important de considérer l'utilisation des grilles quand on crée un graphique. On les utilise pour éviter au lecteur de devoir consommer de l'énergie pour tracer avec leurs yeux ou leurs doigts des lignes horizontales pour repérer les valeurs des points. Mais, il faut les utiliser intelligemment si on souhaite vraiment leur faciliter la vie. Ce que je veux dire par là c'est que si on choisit d'utiliser des grilles, il ne faut pas toutes les mettre! Les grilles doivent servir à guider la lecture si on inclut toutes les grilles (axe des x et axe des y) personne ne saura par où commencer ou ce qui est important à lire. Donc, ici comme mon objectif était de mettre l'accent sur les agences qui ont des tickets les plus chers, j'ai choisi d'utiliser seulement les grilles sur l'axe des y. Je voulais que la lecture se fasse en constatant le prix des tickets. L'axe des x, les dates, c'est plus intuitif comme lecture. On sait que ça va de janvier à décembre, pas besoin de grille pour comprendre ça.

Comme je l'ai déjà mentionné, tous les éléments visuels présents sur le graphique ajoutent une charge cognitive. C'est le cas ici pour les marqueurs et les axes. Dans mon exemple, on a seulement les marqueurs et pas de ligne définie pour identifier les axes. Dans certaines situations, ces marqueurs et axes sont utiles pour soutenir notre message, mais on doit prendre le temps d'y réfléchir au lieu de les laisser là seulement parce que ce sont des éléments ajoutés par défaut par tout bon logiciel qui conçoit des graphiques. Dans mon exemple j'ai retiré les marqueurs et j'ai choisi de ne pas mettre d'axes supplémentaires. Je pense que mettre seulement les grilles principales de l'axe des y fait le travail escompté c'est-à-dire de guider la lecture des prix moyens des tickets.

Ensuite, je pense qu'il y a du travail qui doit être fait sur l'organisation des valeurs et étiquettes des axes. Premièrement, sur l'axe des y. La première règle d'or est de toujours commencer à 0 si on présente des courbes ou des histogrammes. Donc, il faut revoir la graduation. J'ai choisi de présenter les données sur une plage allant de 0 à 120. Je trouve que cette plage de donnée permet de bien centrer les données sur l'espace prévu par le graphique. Avec une graduation au 20\$ je pense que ça apporte suffisamment de précision sans pour autant surcharger le visuel.

Ensuite, il faut revoir l'axe des x. Deux choses ici: il faut éviter les répétitions inutiles et les abréviations. On perd de l'espace utile à répéter 2017 à chacune des graduations de l'axe. On a tout avantage à retirer ça des étiquettes des axes et d'apporter la précision en sous-titre. Aussi, il faut éviter le plus possible les abréviations. On ne doit pas partir du fait que les abréviations qu'on met dans un visuel seront comprises de tous. Les mois de l'année sont assez standards me direz vous. Oui, c'est vrai. Mais c'est quand même une bonne habitude à prendre parce que quand on tombe dans les abréviations de pays par exemple, ça peut devenir plus mélangeant. Je ne vois dit pas combien de fois j'ai vu l'abréviation CA, qui pour moi veut dire Canada, c'est un acquis, un lien qui se fait automatiquement dans ma tête. Alors qu'en fait, dans le visuel CA est mis pour Californie...

Ensuite, il y a du travail à faire du côté de la légende. 10 autorités dans un même graphique, si chacun à sa couleur, ça devient rapidement un arc-en-ciel de couleur... ce n'est pas que je n'aime pas les arcs-en-ciel, mais j'ai une petite tendance à les préférer dans le ciel et non pas dans mes visuels. La couleur c'est l'élément esthétique le plus important, celui qui a le plus d'impact... Il ne faut pas le gaspiller comme ça en utilisant trop de couleur sur ne même graphique. On perd carrément l'effet de focus. Alors, quand je me retrouve dans une situation comme celle-ci, c'est-à-dire où il y a surcharge d'information. Je prends le temps de m'arrêter quelques minutes et de m'interroger sur le message que je veux faire passer. Au début, je vous ai mentionné que j'étais curieuse de voir quelle était l'autorité qui avait les tickets les plus chers en moyenne. Je pense qu'utiliser les couleurs pour mettre cette conclusion en valeur est la meilleure chose à faire ici pour capter l'attention du lecteur et la diriger directement où on la veut...

Donc, j'ai mis la ligne qui représente la série de données des agents de stationnement résidentiels en rouge et toutes les lignes de données qui représentent les autres autorités sont en blanc. J'aime aussi beaucoup

utiliser le titre pour accrocher l'attention, le sous-titre pour expliquer un peu plus le sujet et les couleurs dans ses deux éléments comme légende pour expliquer le graphique au lieu de rajouter d'autres éléments visuels pour créer une légende séparée.

Je me retrouve donc avec un visuel qui ne contient pas d'éléments inutiles, qui identifie clairement le sujet traité et qui fait passer son message en un seul coup d'oeil.

Par contre, il faut faire attention de ne pas désencombrer à l'excès... Trop, c'est comme pas assez. Il faut être attentif à laisser sur le visuel uniquement des éléments qui sont utiles au lecteur pour lui permettre de comprendre notre conclusion, mais il faut justement aussi être prudent pour laisser assez d'éléments dans le visuel pour que le lecteur comprenne justement. Et ce n'est pas toujours facile de trouver le juste équilibre... Quand ça fait plusieurs heures qu'on travaille avec les mêmes données, il y a des éléments qui sont devenus des acquis pour nous, mais il ne faut pas oublier que l'audience à qui s'adresse notre visuel n'ont jamais vu ces données. Il faut donc leur apporter suffisamment de détails pour qu'ils puissent de faire une idée...

Juste pour montrer de quoi ça peut avoir l'air, le deuxième graphique dans les notes de l'épisode montre justement une version qui est trop désencombrée. En générale, une bonne règle du pouce est qu'il faut minimalement bien identifier les axes. C'est essentiel pour que le lecteur puisse comprendre de quoi on parle.

Donc en conclusion, une fois qu'on a créé notre graphique, c'est important de prendre quelques minutes pour personnaliser les paramètres de base des logiciels qu'on utilise pour créer des graphiques. Ces modèles de base sont fait justement pour donner une base. Mais, ils ne sont pas suffisamment intelligents pour savoir comment mettre l'accent sur le message qu'on veut passer à notre audience. Donc, les quelques minutes qu'on investit à peaufiner le graphique final pour améliorer la transmission de notre message ne sont pas perdues ou mal investies. À mon avis, c'est le temps le mieux investi de notre travail. Par ce qu'en fin de compte, la meilleure analyse, de données au monde ne sera pas entendu si personne ne peut comprendre les résultats qu'on veut leur présenter...

Si tu as aimé la démarche de désencombrement que j'ai présenté dans cet épisode, tu vas aimer le contenu de mon nouveau ebook: Principes de visualisation de données: comment améliorer ses graphiques. Dans ce ebook, je reprends un exemple de désencombrement comme j'ai présenté ici, mais je le traite plus en profondeur avec plus d'éléments esthétiques. Une fois que tu maîtrises bien la notion de désencombrement des graphiques, tu pourras te pencher sur mes deux autres sections présentes dans ce ebook soit le manque d'ordre et les stratégies pour attirer l'attention du lecteur question de perfectionner tes compétences et d'améliorer tes visuels. Si ça t'intéresse d'en savoir plus, va voir dans les notes de l'épisode, j'ai mis un lien vers mon ebook.

**CONCLUSION:** Voilà, ça fait le tour de ce que je voulais présenter aujourd'hui. Si jamais tu as des commentaires ou des questions, n'hésite pas à me contacter. Tu peux aller au [johaniefournier.com/contact](http://johaniefournier.com/contact) pour m'écrire directement ou aller dans la section commentaire de l'épisode pour poser tes questions, ça va me faire plaisir de te répondre. Alors, j'espère que cet épisode a été utile et que tu as appris quelque chose, merci de m'avoir écouté et on se reparle le 20 janvier! d'ici là, je te souhaite de belles fêtes, de passer du bon temps en famille et de créer de superbes visualisations de données!

Quelques liens utiles:

- Transcription de l'épisode en pdf: [ici](#)
- Pour écouter l'épisode : [ici](#)
- Les graphiques discutés: [ici](#), [ici](#) et [ici](#)
- L'article de blogue en lien avec cet épisode: [blogue](#)
- Me contacter: [contact](#)
- Ebook: Principes de visualisation de données: [Comment améliorer ses graphiques](#)