

# ADV38\_TRANSCRIPT - Faire du 3D en toute légalité

*Johanie Fournier, agr*

*2020-03-07*

Est-ce que ça t'est déjà arrivé d'avoir une intuition? Un vrai sentiment que tu sais quelle est la bonne action à prendre.

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 la semaine.

Tsé, comme quand c'est le printemps et que c'est le temps des semis. La météo annonce les deux plus belles semaines de printemps que tu n'as jamais connu, tous les producteurs voisins sont en train de semer leur maïs, mais tu sais au plus profond que c'est trop top pour semer...

Et ben, à chaque fois que je ne me suis pas écoutée, on n'a pas mangé de maïs...

La dure vie d'avoir un petit jardin de 5000 pi<sup>2</sup> à semer.

C'est exactement ce qui est arrivé quand je me suis fait expliquer que les aires sous la courbe ne sont pas des visualisations optimales...

J'étais convaincue que j'allais un jour trouver une situation qui allait prouver le contraire.

À l'épisode 18 (que tu peux aller écouter en suivant le lien présent dans les notes de l'épisode), j'avais essayé de faire des aires sous la courbe avec les données de la semaine.

À ce moment-là, j'étais déjà déterminé à prouver que c'était possible de faire de présenter des données de manière correcte avec des aires sous la courbe.

Et bien, je n'avais pas réussi à prouver mon point. Le principal défaut des aires sous la courbe, principalement si tu les empiles est que c'est difficile de constater visuellement les valeurs associées à chaque couleur puisqu'il n'y a pas de ligne de base pour comparer l'amplitude. La ligne de base des traitements du haut c'est la variable qui est juste en dessous... et ainsi de suite...

Dans l'épisode 18, j'ai montré quelques alternatives pour éviter de mal présenter ses données et donc utiliser d'autres types de graphiques que les aires sous la courbe.

Par contre, j'ai toujours eu l'intuition qu'il y avait une façon ou bien une structure de données particulière qui se prête bien à l'utilisation des aires sous la courbe.

Après tout, ce type de graphique là n'a pas été inventé pour rien... il faut bien qu'il y ait une situation idéale pour l'utiliser.

Et bien, je l'ai enfin trouvé cette semaine. J'ai trouvé le patron de données 3D parfait pour utiliser des aires sous la courbe. Je dis que c'est en 3D parce que je présente les résultats en fonction de 3 variables. Je t'explique tout ça...

Alors, les données de cette semaine portent sur les résultats des joueurs de hockey et sont disponibles dans mon article de blogue au [johaniefournier.com/tyt2020w10](http://johaniefournier.com/tyt2020w10).

Il y a 3 bases de données de disponibles cette semaine. Mais, comme je viens de le mentionner, j'avais déjà une bonne idée de ce que je voulais faire en voyant la structure des données cette semaine.

Donc, je n'ai pas pris le temps d'investiguer toutes les variables des 3 bases de données et je me suis consacré sur 3 variables en particulier: l'âge, l'année et le nombre de buts comptés. La BD nous donne ces valeurs pour chacun des matchs joués pour chacun des joueurs.

On a donc un terrain de jeu qui comporte 49 384 données pour s'amuser.

J'ai commencé par regarder attentivement les 3 variables, je voulais m'assurer qu'elles étaient sous le bon format pour que je puisse travailler, mais aussi m'assurer qu'elles ne comportaient pas de valeurs aberrantes qui cachent des valeurs nulles.

Si tu ne comprends pas ma paranoïa de cette semaine va écouter l'épisode de la semaine passée l'épisode 37 ça en vaut la peine.

Bon aucun vice caché cette semaine, par contre en regardant de plus près les données, je me suis rendu compte que j'aurais fait un graphique de 3km de long puisqu'il y a des données pour toutes les années entre 1980 et 2020.

J'ai choisi de faire la somme des buts par âge et par tranche de 10 ans, donc par décennie pour avoir un visuel qui se lit bien.

Je dis que j'ai fait un graphique en 3D cette semaine parce que je présente 3 variables. J'ai mis l'âge des joueurs sur l'axe des x, les décennies entre 1980 et 2010 sur l'axe des y et les valeurs que je présente avec des aires sous la courbe sont la somme des buts réalisés par les joueurs. Donc 3 variables.

Les aires sous la courbe sont mises les unes au-dessus des autres et se croisent un peu par souci d'économie d'espace, mais chacune à une ligne de base bien droite. . .

On appelle ce type des graphiques des ridgeslines. . .

On voit clairement sur ce visuel que l'âge à lequel les joueurs marquent le plus de buts se déplacent en fonction des décennies.

Les jours des années 80 comptaient plus de buts début 20n, ceux des années 90, fin 20n et ceux des années 2010, début 30n. Le patron est un peu particulier pour les buts comptés dans les années 2000.

En plus, la plage de temps pour laquelle les joueurs de chaque décennie changent. Plus les années passent, plus les joueurs jouent au hockey plus longtemps. Dans les années 80 la carrière des joueurs se terminait vers 30 ans. Maintenant, les joueurs sont présents jusqu'à l'âge de 45 ans, ce qui représente une belle amélioration pour ce sport.

Les graphiques en 3D, dans leur version traditionnelle comme ajouter une dimension supplémentaire à un histogramme seulement pour "faire beau", c'est-à-dire que l'ajout d'éléments esthétiques sur le graphique n'a aucune autre utilité que pour son aspect visuel, n'ont pas bonne réputation dans le monde de la visualisation de donnée

C'est même la règle de conduite la plus célèbre: Ne pas utiliser de 3D!

Et je suis tout à fait d'accord avec ça, ajouter une dimension inutile si je peux dire ça ainsi, en plus de n'avoir aucune utilité ça complique la lecture des 2 dimensions principales.

Ça complique la vie parce qu'en plus d'avoir des éléments visuels inutiles sur le graphique, la ligne de base qui permet la lecture et la comparaison des variables est souvent inclinée.

Notre cerveau n'est pas fait pour comparer des éléments qui ne sont pas parfaitement alignés sur une ligne horizontale, donc les graphiques en 3D sont très difficiles à lire et malheureusement entraîne souvent une mauvaise interprétation des données.

Dans le graphique que j'ai fait cette semaine, j'ai mis en graphique 3 variables, d'où mes 3 dimensions, mais tout est aligné sur un même plan. Donc, l'ajout de ma dimension supplémentaire ne vient pas compliquer ni la lecture ni la compréhension des données.

Alors, en terminant, je voudrais faire un petit tour des principes de base de la visualisation de données que j'ai appliqués cette semaine.

Alors, premièrement j'ai brisé la règle no 1: Ne pas utiliser de 3D! Je blague un peu parce que je n'ai pas fait du 3D, mais, j'ai présenter 3 variables de mon ensemble de données de manière avec des aires sous la courbe en prenant soins que chacune des séries de données présentées soit facile à lire, c'est-à-dire qu'elle a une ligne de base claire.

Ensuite, j'ai réduit l'encombrement au minimum, pas d'éléments en gras inutile, pas d'icône sur le graphique, bref j'ai conservé seulement l'essentiel: Données, axes, titre.

Pour permettre au lecteur de comprendre le sujet à la base de mon graphique, j'ai utilisé un titre explicatif qui énonce le sujet, mais qui sert aussi de légende. En mettant les mots 'comptent plus de buts' de mon titre de la même couleur que mes séries des données, je viens préciser au lecteur qu'il regarde le nombre de buts en fonction des paramètres âge et décennie qu'il retrouve sur l'axe des x et des y.

Donc, dans cette visualisation attributs préattentifs principal est la couleur. Utilisé judicieusement ici, je viens réduire l'encombrement en évitant d'ajouter une légende tout en mettant l'emphase sur le sujet. C'est presque de la magie!

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 dit à la semaine prochaine!

Quelques liens utiles:

- ADV18: [ici](#)
- ADV37: [ici](#)
- Pour écouter l'épisode: [ici](#)
- Transcription de l'épisode: [ici](#) et en pdf [ici](#)
- L'article de blogue en lien avec cet épisode: [blogue](#)
- Me contacter: [contact](#)

Tu aimerais avoir par écrit ce processus de dataviz? J'ai mis toutes les étapes que je réalise à chaque semaine pour créer mes visuels dans un aide-mémoire: [Le Processus Dataviz](#)