

ADV32_TRANSCRIPT - Comment ne pas mettre 2 axes sur un même graphique

Johanie Fournier, agr

2020-01-26

TEASER: Qu'est-ce qu'on fait quand on veut montrer plusieurs variables sur un même graphique, 2 axes, hmmm pas certaine...

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 32e! J'espère que vous allez bien parce que cette semaine on jase du classement de Soptify. Les données proviennent de Spotify via Sporify et on a accès à une base de données de presque 33 000 chansons pour s'amuser. Le lien pour y avoir accès se trouve dans mon article de blogue au johaniefournier.com/tyt2020w4.

Chaque ligne de la base de données nous renseigne sur le nom de la chanson, l'artiste-compositeur, la popularité (qui est un indice qui va de 0 à 100), le nom de l'album, la date de publication, le nom de la liste de lecture à laquelle la chanson est associée, le genre et le sous-genre de la liste, et l'on termine avec 12 paramètres qui viennent décrire les chansons de manière à pouvoir les catégoriser: entre autres on a l'énergie, la dansabilité, le bonheur et le tempo pour ne nommer que les éléments principaux.

Alors, après avoir exploré un peu cette base de données et m'être familiarisé avec les différents paramètres qui décrivent les chansons j'ai pris la peine de lire l'article en lien avec les données de cette semaine, je me suis fixé comme objectif de visualiser les 4 principaux indicateurs qui permettent de décrire les chansons soit: l'énergie, la danseabilité, le bonheur et le tempo pour chaque genre de chanson présente dans la base de donnée.

CHOIX DU TYPE DE VIZ: Pour atteindre cet objectif, j'ai choisi de créer un graphique qui serait un hybride entre un lollipop et un indicateur.

PRÉSENTER LE GRAPHIQUE: Donc, pour chaque genre j'ai tracé une ligne horizontale qui représente la place de données de 0 à 1. Sur cette ligne j'ai placé des points dont la position représente la valeur du paramètre en question. Comme j'ai 4 paramètres à montrer, j'ai utilisé une légende de couleur pour identifier quel point présente quel paramètre. J'ai joué un peu avec la superposition des points pour créer un peu un effet de boutons. Ça, ce n'est absolument pas nécessaire, c'est juste un petit ajout un peu plus artistique juste pour attirer l'oeil. Donc, pour chaque genre c'est possible de voir quel paramètre est le plus présent, le moins présent et quels paramètres se superposent. Ensuite, j'ai ajouté un titre pour attirer l'attention et un sous-titre explicatif et voilà, le visuel de la semaine est fait!

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 expliqué comment j'ai procédé pour créer le visuel, mais ce que je ne vous ai pas dit c'est que mes 4 indicateurs n'étaient pas tous sur une échelle de 0 à 1. En fait, 3 d'entre eux l'étaient. Il y a seulement le tempo qui n'était pas décrit avec le même type d'indice.

Ce n'est pas rare que je doive montrer des valeurs sur un même graphique, mais que les valeurs en question n'ont pas toutes la même unité de base. Et malheureusement, avoir la même unité de base ce n'est pas un prérequis pour la pertinence des éléments qu'on veut montrer. En fait, on dirait même qu'il y a une loi non

écrite qui dit que si tes variables n'ont pas la même unité c'est certain que ce serait pertinent pour une raison ou pour une autre de les montrer sur un même graphique. Alors que fait-on dans ce temps-là??

Est-ce que c'est une bonne idée d'ajouter un second axe? Je ne suis pas convaincue parce que le principal problème des visuels avec 2 axes est qu'une personne extérieure au sujet à beaucoup de difficulté à identifier quelles données sont associées à quel axe. Et s'il y a une chose qu'il faut absolument éviter en visualisation de données est de confondre le lecteur, si ce n'est pas clair au premier coup d'oeil, c'est qu'on a encore du travail à faire pour atteindre la perfection.

Pour éviter de mettre un deuxième axe et de quand même montrer nos variables pertinentes sur un même graphique, il y a quand même quelques solutions qui s'offrent à nous en fonction de ce qu'on présente:

On peut, ne pas mettre le deuxième axe vertical et mettre des étiquettes de données directement sur le graphique à la place. À utiliser seulement si les unités sont clairement définies quelque part, en sous-titre par exemple.

On peut, superposer les graphiques de manière à ce que chacune des séries de données montrées ait son propre axe des y, mais que l'axe de x soit partagé. C'est pratique pour la majorité des situations, mais pas pour tout.

Dernièrement, j'ai eu à créer des graphiques qui avaient plus l'aspect des indicateurs. Tous sur la même ligne comme le graphique de cette semaine ou sur des lignes différentes pour montrer des unités différentes. Dans tous les cas, toutes mes variables devaient être montrées sur un même graphique.

Qu'est-ce qu'on fait dans ce temps-là? En ben, j'ai transformé les données. Cette semaine, il y avait seulement le tempo qui n'était pas sur une valeur de 0 à 1. J'ai donc créé une régression pour convertir les données et les transposer sur une échelle de 0 à 1. Il faut faire attention ici, il faut bien prendre le temps et calculer minutieusement la conversion pour être certaine de ne pas inclure de biais et s'assurer justement que le résultat final de la conversion représente bien les données. Ce n'est pas approprié dans tous les cas, mais dans un contexte où on montre des indicateurs et que la position du point est quand même un peu relative, je pense que c'est une belle solution.

Une autre façon de contourner les graphiques à 2 axes et de faire une transformation moins dommageable sur les résultats est tout simplement d'utiliser les pourcentages pour des variables montrées ou de diviser l'une par l'autre. Le choix de la transformation étant bien sûr guidée par ce qui l'applique le mieux à vos données ou au message à passer. Dans tous les cas, il faut prendre le temps de vérifier si la transformation ne vient pas enlever de l'information importante, un effet d'échelle par exemple.

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 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:

- Transcription de l'épisode en pdf: [ici](#)
- Pour écouter l'épisode : [ici](#)
- Le graphique discuté: [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](#)