

ADV35_TRANSCRIPT - Graphique complexe public non averti

Johanie Fournier, agr

2020-02-15

TEASER: Qu'est-ce qu'on fait quand la meilleure façon de présenter les données pour répondre l'objectif est un graphique complexe et qu'on s'adresse à un public pas nécessairement à l'aise avec les graphiques en général?

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 35e épisode! Aujourd'hui je te partage un peu de mon travail du quotidien... Le vrai que je fais entre 8h et 5h là, pas le fun que j'ai à faire des graphiques avant que le soleil se lève!

D'une part, ce que je fais pour des clients ou des collègues de travail est confidentiel et d'autre part parce que ça traite le plus souvent de sujets vraiment précis, le plus souvent en agriculture, et je ne veux pas t'endormir avec ça.

Mais, aujourd'hui je pense que ça pourrait être utile que je te partage le processus de réflexion et le travail que j'ai eu à réaliser semaine.

En fait, comme ça arrive assez souvent, je me suis fait demander de faire un graphique...

Jusque là rien de nouveau, ça me fait toujours extrêmement plaisir surtout parce j'aime voir l'impact que ça a sur la personne qui me donne le mandat. C'est vraiment beau de voir la clareté qu'un peu de ménage dans les données et un ou deux bons graphiques bien fait peu apporter à la personne que m'a donnée le mandat. Et le mieux dans tout ça c'est qu'ils en redemandent des graphiques!

Bref ménage et graphique plus tard, je suis arrivée à la conclusion que la meilleure façon de présenter les données pour répondre à leur l'objectif est un waterfall chart.

Oh, oh... un waterfall chart ou graphique en cascade pour la version française, c'est un superbe graphique, mais qui est un peu plus complexe à comprendre... surtout pour des personnes qui ne sont pas à l'aise avec les graphiques en général.

C'est un beau problème et surtout un beau défi! Je dois d'abord apprendre à faire ce type de graphique, parce que c'est assez particulier comme présentation des données. Ensuite, il faut que je travaille sur la transmission du message et l'éducation de mon public cible.

Alors, cette semaine, j'ai décidé d'utiliser l'opportunité du tidytuesday pour apprendre à faire des waterfall chart. Les données de cette semaine traitent des réservations de chambres d'hôtel. Le lien pour y avoir accès aux données se trouve, comme d'habitude, dans mon article de blogue au johaniefournier.com/tyt2020w7 et j'ai aussi mis le lien dans les notes de cet épisode.

La base de données disponibles cette semaine nous renseigne sur plusieurs paramètres intéressants, ça vaut la peine de prendre le temps de regarder tout ça, mais comme je viens d'expliquer j'étais en mission waterfall chart cette semaine.

Alors, j'ai choisi la variable qui avait le plus de sens pour moi: le prix moyen d'une chambre. Les waterfall chart sont ces graphiques qui permette de visualiser l'évolution dans le temps de la variation d'une variable. Donc, mon objectif cette semaine avec les données des tidytuesday est de visualiser l'évolution du prix moyen des chambres avec un waterfall chart.

Donc, pour bien répondre à mon objectif, j'ai pris le temps de bien explorer les données du prix moyen pour m'assurer que j'allais bien pouvoir travailler avec ça.

Le prix moyen est disponible pour chaque réservation donc on a plusieurs valeurs par jour dans une base de données qui fait 120 000 lignes. On a des données entre 2015 et 2017. Il y aurait eu matière à faire une analyse plus poussée pour aller voir les conditions qui font augmenter ou diminuer le prix de réservation des chambres, mais j'ai gardé ça simple ici et je me suis concentrée sur une variation globale en fonction des semaines dans l'année.

Pour m'assurer de pouvoir faire un waterfall chart, je dois voir un patron de variation dans les données à la fois entre les valeurs de l'axe des x et dans le temps.

Pour valider le tout, j'ai commencé par visualiser la distribution des données. Donc, en visualisant la fréquence relative et la composition avec un box plot, j'ai pu retirer les valeurs qui ne me seraient pas très utiles pour répondre à mon objectif qui sont en fait les valeurs de 0 et celles plus grandes que 225.

Une fois mon petit ménage fait, je devais quand même valider le patron de distribution des moyennes dans le temps pour m'assurer que ça aurait du sens. Donc, j'ai fait un histogramme qui montre le prix moyen des chambres en fonction des semaines dans l'année. Et c'est parfait, on voit une variation entre les semaines et un patron de données qui nous apprend que les chambres coutent plus cher en été.

J'ai tout ce qu'il faut pour faire un waterfall chart! Bon ce n'est pas tout... fallait aussi que je le fasse. Le principe de base de ce type de graphique est qu'on visualise la différence à la fois en fonction d'une valeur de base (semaine 1) et de la valeur de la semaine précédente.

Il faut donc monter une base de données qui va permettre de tracer le tout dans R. Ce qu'il faut savoir c'est que je n'ai pas utilisé de fonction toute prête pour faire mon graphique. J'aime ça me compliquer la vie ça l'air! Mais c'est parfait pour comprendre le principe.

En gros, il nous faut une base de données qui identifie les variations qu'on veut visualiser. Je ne vous explique pas plus ici, mais si vous utilisez R ça vaut la peine de prendre le temps d'aller voir comment je m'y suis prise pour travailler les données pour ensuite les mettre en graphique.

PRÉSENTER LE GRAPHIQUE: Donc, j'ai fait un waterfall chart qui présente le prix des chambres sur l'axe des y, les semaines de l'année sur l'axe des x et une cascade qui nous montre la variation qu'il y a entre les semaines. Lorsque les prix d'une semaine donnée diminuent par rapport à la semaine précédente, le rectangle de cette semaine-là est en bleu et lorsque le prix augmente par rapport à la semaine précédente, le rectangle est en rouge. Donc, globalement, on peut voir que le prix est en augmentation dans la première moitié de l'année pour être en diminution dans la deuxième partie de l'année. Et que le prix diminue plus vite qu'il n'a augmenté. J'aurais pu pousser plus loin l'analyse du graphique, mais ce n'était pas mon objectif personnel pour cette semaine. Donc, j'ai juste ajouté un titre et un sous-titre pour positionner le sujet et c'est tout. Mon graphique est fait, objectif 1 accompli!!

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: J'ai donc réussi à créer un type de graphique qui est assez complexe. On s'entend il y a pire que ça, mais ce n'est pas aussi instinctif à lire qu'un simple histogramme ou une courbe. Même si j'ai fait bien attention de ne pas surcharger mon visuel, de retirer tous les éléments qui pourraient venir distraire mon lecteur, je sais que si j'utilise ce modèle de graphique pour présenter les résultats dans le mandat que j'ai reçu dans le cadre de mon travail, ça risque d'accrocher.

Dans mon cas, ce type de graphique est parfait pour présenter les données. C'est exactement ça l'objectif: avoir une vue rapide des variations chaque semaine et les couleurs rouge et bleu sont parfaite pour faire ce travail. Mais, le moyen pour arriver à présenter ces données est d'utiliser un type de graphique un peu plus complexe.

Le public à lequel s'adresse le suivi que je vais réaliser avec les waterfall chart n'est pas nécessairement à l'aise avec les graphiques en général. On ne peut pas les blâmer ce n'est pas tout le monde qui y trouve un

intérêt et surtout, ce n'est pas tout le monde qui a l'occasion de travailler avec des données et des graphiques tous les jours.

Donc, ma stratégie avec ça est d'y aller graduellement. Déjà, si on réussit à faire un suivi des données avec un graphique au lieu de le faire à la main je crois qu'on aura réussi.

Ensuite, je considère que j'ai un travail d'éducation à faire. Comme ce n'est pas tout le monde qui voit les avantages du waterfall chart du premier coup d'oeil mon travail est de prendre le temps de bien leur expliquer, leur montrer des exemples répondre à leurs questions et d'adapter le graphique en fonction de leur besoin.

Par contre, je ne vais pas imposer ce type de graphique. Je vais prendre le temps de montrer les données avec un histogramme simple et le mettre côte à côte avec le waterfall chart pour voir comment le tout sera perçu.

Voilà, c'est comme ça que je compte présenter des données avec un graphique complexe à un public non averti. Le projet de suivi sera fait cet été, alors cet automne je pourrai faire un petit compte rendu et si j'oublie n'hésite pas à me demander!

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:

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

Es-tu abonné à mon podcast? Sinon, je t'encourage à t'inscrire dès aujourd'hui, il ne faudrait pas manquer un épisode! [Clique ici](#) pour t'inscrire sur iTunes.

Si tu te sens particulièrement gentil aujourd'hui, je te serais très reconnaissante si tu me laissais un commentaire sur iTunes. Les commentaires aident les gens à trouver mon podcast et ils sont aussi très précieux pour moi. Sélectionne "Notes et Avis" et ensuite "Rédiger un avis" et fais-moi savoir quel est ton épisode préféré. Merci!