

ADV33_TRANSCRIPT - Spécial R: weathercan

Johanie Fournier, agr

2020-02-02

TEASER: Je ne te partage pas souvent mon enthousiasme envers le logiciel R pour traiter et visualiser des données. Cette semaine, je n'ai pas pu résister, car j'ai découvert l'existence d'un package qui va me faire sauver énormément de temps et me donner accès à des données ultras pertinentes. Alors, alerte à ceux qui sont allergiques à R: c'est un épisode où je ne parler que de ça. Aux passionnés de R comme moi: vous allez voir c'est juste magique!

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 33e! Aujourd'hui, je déroge de ma programmation habituelle, car je n'ai pas participé au TidyTuesday cette semaine. C'est pour une bonne cause, en fait j'ai découvert le package weathercan et je n'ai pas pu m'empêcher de créer un dashboard pour visualiser les données météo de la station météo la plus près de chez moi.

Alors, je ne vous partage pas souvent ma passion pour le logiciel R. Je crois définitivement que faire de la visualisation de données ce n'est pas lié à un outil ou un logiciel en particulier et c'est pour ça que je ne met pas l'emphase sur l'outil que j'utilise moi pour travailler. Je pense sincèrement que la seule chose qui est importante est de bien maîtriser son outil de travail, peu importe lequel. Même Excel peut faire des choses hallucinantes, il faut juste de donner la peine d'apprendre comment le faire et ça, c'est long. Maîtriser un logiciel c'est long et ardu, mais ça en vaut la peine croyez moi, parce que quand on a passé la période d'apprentissage ou l'on doit valider ou mettre chacune des virgules et qu'on commence à travailler un rapidement et avec plus de confiance en nos capacités, ça donne des ailes et ça ouvre tout un monde de possibilité.

Alors, comme je crois fermement que vous avez le choix d'utiliser le logiciel qui vous plait le mieux, j'essaie de ne pas trop vous montrer mon enthousiasme débordant pour R et uniquement vous mettre le code source au cas où il y ait d'autres passionnés comme moi et que ça puisse être utile pour eux.

Mais cette semaine, je n'ai pas pu résister. Alors, pour le temps d'un épisode, je vais fièrement vous casser les oreilles avec R et le merveilleux package que j'ai découvert cette semaine: Weathercan.

Ça fait longtemps que j'essayais de trouver une bonne source de données météo, autant les données historiques que les prévisions. Je ne vous surprendrai sûrement pas en disant qu'un des facteurs qui vient jouer sur le quotidien des producteurs agricoles est la météo. En plus, j'ai aussi personnellement besoin de données comme celle-là parce que je cultive un petit jardin de 5000 pi² et je préfère savoir précisément quand arroser en créant des bilans hydriques au lieu de dérouler 300pi de hosse à jardin juste pour le fun...

Bref, une source de données météo agréable à travailler et complète ça manquait à ma vie. Trouver des données météo sur internet c'est facile, mais y avoir accès sans devoir faire du copier-coller ça c'est une autre affaire.

C'est pour ça que je trouve que R en général et particulièrement le package weathercan aujourd'hui c'est magique. Avec ce fameux package, j'ai accès directement dans R à la base de données historique des stations météo de tout le Canada.

Ok, je redis ça juste une fois pour ceux qui n'ont pas encore réalisé la magie de la chose. Je n'ai plus besoin d'aller sur le site de Environnement Canada de chercher dans leur site où est caché l'historique météo, d'essayer de trouver une station (par essai et erreur) qui est encore active et qui contient des données correctes, télécharger un csv et convertir ce fameux csv en quelque chose de compréhensible à chaque fois

que je veux des données météo. Et juste pour vous mettre en contexte, je veux des données météo tous les jours l'été donc en saison de culture.

Maintenant, la seule chose que j'ai à faire est de cliquer sur « run » et j'obtiens ce superbe dashboard météo à jour - avec l'histoire des données les plus à jour qui existe c'est à dire: hier!

Ok, il faut que j'avoue que j'ai quand même travaillé fort pour créer mon markdown qui m'a permis d'automatiser tout ça, mais ça en a valu le coup vous pouvez me croire.

Il faut aussi que je précise que le visuel que j'ai créé cette semaine n'est pas encore à mon gout. Dans le sens que je veux ajouter des prévisions, parce qu'on a aussi accès à partir de R au serveur metéocode de Environnement Canada qui nous donne accès aux résultats de tous leurs modèles de prédictions et je veux encore travailler sur les éléments esthétiques de mon graphique. Mais en général, je suis assez fière de mon travail et d'avoir réussi à automatisé tout ça.

À mon avis, l'automatisation c'est l'une des grandes forces de R. Je peux créer des markdown qui se génèrent uniquement en cliquant sur un bouton ou créer des programmes R standards et utiliser un taskeduleur pour que l'ordinateur exécute ce programme sans même que j'aie à être là ou à intervenir. Est-ce que vous imaginez le temps que ça sauve! C'est comme avoir un lutin qui travaille pour nous pour faire les tâches répétitives pendant qu'on est concentré à autre chose!

Bon, si je reviens au Weathercan. C'est un package très simple qui nous fournit seulement quelque fonction, mais comme je vous l'ai expliqué ça nous rapporte énormément en économie de temps d'accès aux données.

Comme tous les package dans R, il faut commencer par l'installer et télécharger la librairie. Ensuite, on peut directement dans R, identifier la station météo qui nous intéresse. Si on connaît son ID c'est très simple, sinon, on peut aussi chercher les stations météo les plus près d'un point précis en utilisant les coordonnées géographiques de ce point en spécifiant la distance maximale à respecter.

Une fois qu'on a identifié un groupe de station météo qui est dans un rayon raisonnable de l'endroit qui nous intéresse, on peut télécharger les données de ces stations et visualiser le profil des données manquantes de chacune des stations présentes dans le rayon choisi. Plus besoin de jouer au chat et à la souris avec les données de précipitations (qui ne sont pas prises à toutes les stations...), on a un beau visuel qui nous permet de faire un choix de station éclairé en fonction des éléments qu'on veut travaillé.

Après ça, le fun commence. On a tout ce qu'il faut pour créer tous les visuels qu'on a de besoin. Personnellement, j'avais besoin d'un portrait général qui présente la température, l'humidité, la vitesse du vent et les précipitations. J'ai donc choisi de créer un petit dashboard avec ces 4 graphiques. La position du titre me sert à identifier l'endroit de la station météo. Chacun des 4 graphiques possède son propre titre avec l'unité de mesure qui lui est associé. J'ai évité d'identifier les axes pour ne pas surcharger le visuel. Les dashboard, par définition c'est très chargé. Les axes des y n'ont pas tous la même échelle non plus, ça aurait été l'idéal du point de vue esthétique, mais ça ne fait pas de sens avec les unités de mesure de chacune des valeurs présentées. Des pourcentages et des °C c'est quand même assez différent.

J'ai choisi de présenter chacun des paramètres avec une couleur différente. C'est assez rare que j'utilise 4 couleurs, mais cette fois-ci c'est plus pour le côté esthétique que pour le côté compréhension de la chose et je n'ai pas besoin de légende non plus.

Voilà, comme je l'ai mentionné ce visuel n'est pas parfait parce que je veux encore ajouter les données de prévisions météo et je vais sûrement ajuster les éléments esthétiques en cours de route avec l'évolution de l'utilisation du graphique, mais je pense que c'est un bon point de départ et surtout je peux créer rapidement une mise à jour avec les données récente.

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](#)