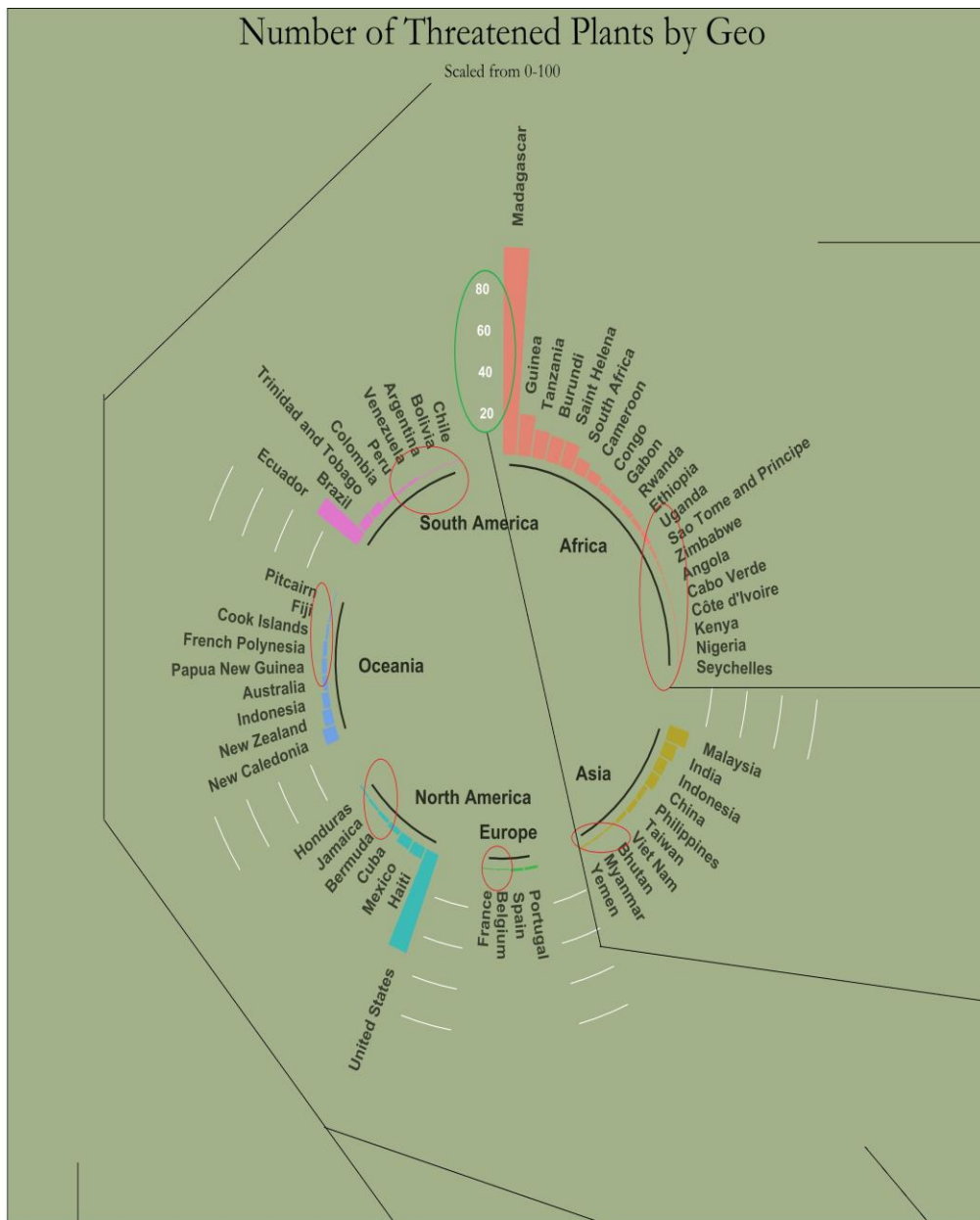


Assignment 1 graph critic and remake with ggplot 2

Lien : <https://twitter.com/joepope44/status/1296544220799537172/photo/1>

Discussion du graphique tweeté par [@joepope44](#) :



1 Pour commencer, on peut souligner le bon usage de la décomposition des pays selon les continents pour différencier entre les continents qui ont un nombre important de plantes menacées comme l'Afrique et les continents qui ont un petit nombre de plantes menacées comme l'Europe.

2 Les pays encerclés apparaissent comme si'il n'y a pas d'observations dans la base de données. C'est à cause du "scale" que l'auteur a fait sur les données.
En plus, on ne peut pas lire les valeurs exactes de nombre de plantes menacées par pays.

3 En regardant les données, j'ai remarqué que l'auteur décompte le nombre de plantes menacées par pays sur des pays qui n'ont pas le même nombre d'observation par exemple, on trouve que l'auteur est en train de calculer le nombre des plantes menacées sur 12 observations en Yémen et sur plus que 1100 observations en Madagascar ce qui rend la comparaison est non significative c'est pourquoi je propose de faire un taux en % (plantes menacées / total des plantes).

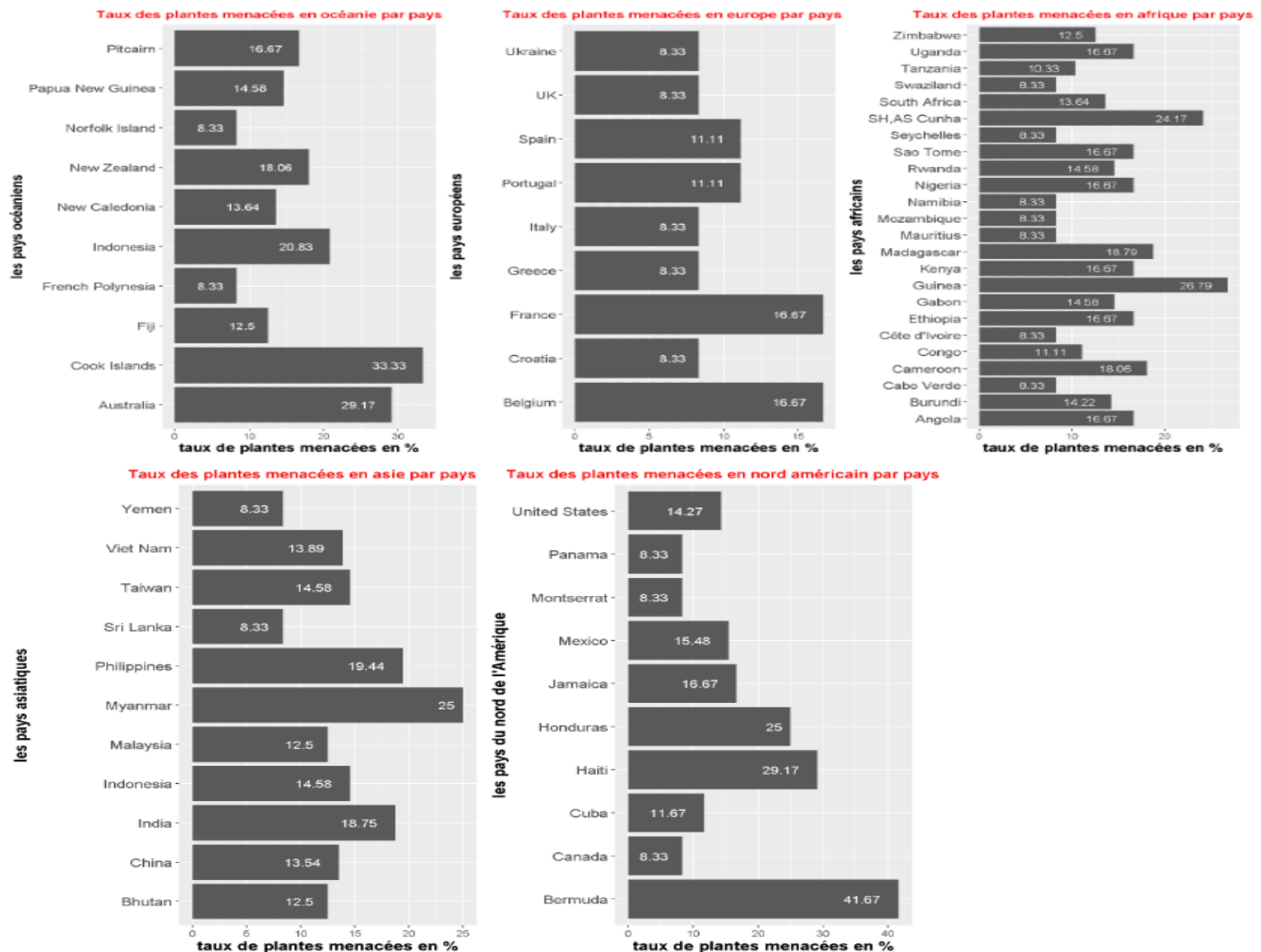
4 On remarque aussi que la figure ne contient pas la source de données, il vaut mieux ajouter d'où l'auteur a pris les données pour construire les plots.

5 L'auteur a décidé de mettre les données à l'échelle "entre 0 et 100", mais, cette décision signifie qu'il va perdre le vrai décompte des plantes menacées.

6 c'est vrai que les pays qui ont les plantes les plus menacées par continent seront toujours évidents mais on perd l'effet visuel comme si on a sous-estimé la différence du nombre de plantes menacées entre les pays (elle est moins accentuée).

7 je voulais finir par une remarque esthétique, je trouve que la couleur de fond est un peu foncée, il est préférable de mettre un fond blanc, mais c'est juste une remarque 100% subjective.

Une solution proposée :



(Source de données : <https://github.com/rfordatascience/tidyuesday/blob/master/data/2020/2020-08-18/threats.csv>)

1. pré-traitement et description :

J'ai fait d'abord un pré-traitement des données en effet :

- J'ai simplifié quelques noms de pays qui rend le graphique plus lisible et moi chargé.
- J'ai ajouté une colonne nommée "total_plants" qui définit le nombre total de plantes par pays car comme je l'ai déjà dit dans les critiques, il est non significatif de comparer le nombre de plantes menacées entre des pays qui n'ont pas le même nombre de plantes total prenant l'exemple des populations on ne peut pas parler d'un phénomène de surpopulation en comparant que les populations des pays mais on compare le nombre d'habitants par km2 qui est plus réaliste, du coup j'ai pensé à comparer le pays par un autre critère qui est le taux des plantes menacées (plantes menacées / total plantes).
- J'ai pensé à changer le diagramme en bâton circulaire en diagramme en bâton inversé et ne pas faire du "Scale" pour résoudre le problème des observations qui ont des petites valeurs qui semblent non claires dans le premier graphique, en plus lors de la visualisation j'ai trouvé que l'auteur a éliminé quelques pays d'Europe par exemple car il n'a pas trouvé de place pour les mettre (par exemple UK, l'Italie, la Grèce etc ...). Pour cela, j'ai décomposé le tableau de données en 5 petits "data frame" (décomposition par continent) pour faire un diagramme en bâtons inversé pour chaque continent (j'ai voulu laisser cette histoire de continent que l'auteur a mis en valeur pour qu'on puisse avoir une vue globale de la différence de taux de plantes menacées entre les continents).
- Dans ce graphe, j'ai essayé de choisir les bonnes couleurs des canaux (simple en point de vue visuel). Pour le thème : j'ai supprimé la légende puisque les continents sont bien identifiés dans le titre, j'ai ajouté les lignes de la grille pour faciliter au lecteur de comparer entre les pays, j'ai utilisé un fond bleu ciel claire au lieu d'un couleur très foncé que l'auteur a utilisé. Tout ça pour alléger le maximum notre graphe. J'ai ajouté aussi la source de données qui est indispensable de chaque graphe par suite le lecteur serait ravi d'explorer le graphique et d'en tirer des conclusions. Enfin, j'ai ajouté les pourcentages des plantes menacées pour chaque pays pour que le lecteur peut avoir une idée sur les chiffres et différencier entre les pays si rapidement et d'avoir une idée de la proportion de chaque pays donc je peux dire que la visualisation est plus claire maintenant.