**Brainstorming**

Module 2 : **GitHub et la visualisation de données numériques**

Objectif : Devenez autonome sur GitHub et découvrez l'art de visualiser des données numériques et de faire parler les statistiques descriptives.

* Comprendre les principes du versionnage de code avec GitHub.
* Gérer et transformer des données numériques à l'aide de dplyr.
* Visualiser des données numériques avec ggplot2.
* Calculer et interpréter des statistiques descriptives pour des variables numériques.

En détail :

Présentation du versionnage de code et intégration de Github avec Rstudio. Librairie dplyr et gestion de données (filtrage, sélection) Librairie ggplot2, premiers graphiques avec variables numériques : scatter plot, histogramme. Statistiques descriptives pour variable numérique (summarise dans dplyr)

Données : Données biologique sur les pingouins en antarctique avec la librairie « palmerpenguins».

Défi : Tirer des conclusions sur les caractéristiques des penguins (Comparaison des caractéristiques physiques, distribution des espèces par iles, analyse de la taille du bec par espèce et par sexe, tendances entre longueur et prodondeur du bec)

Idée :

Faire comme si ils devaient travailler avec un chercher sur les penguins.

Github : utilisation dans Rstudio

Nombre de bins dans histogramme

Bonne pratique de visualisation : titre dans nos axes, de notre graphique

Est-ce qu’il y a une différente des autres (espece) : avec stat descriptive et visu.

Dans le scatter plot : faire l’interprétation sans les espèces et ensuite avec les espece (paradox de simpson)

Ajouter une donnée aberrantes dans penguins, on leur fait faire l’analyse, on remarque que dans les stat descriptives : on pourrait même l’enlever dans l’analyse. On prend un nouveau jeu de données (dat2 pour pas écraser le premier). Faire le lien avec la reproductibilité (trouver des choses). Faire parler comme un chercheur au fur et à mesure de l’aventure.

Pour le DÉFI : on reprend l’exemple de weather can et on leur demande un graphique de ligne brisé sur la journée de fete sur toutes les années. On va leur demander de prendre des données avec une données manquantes, et leur poser la question de comment ggplot les gèrent (ca trace une ligne entre les deux) : on peut leur faire enlever le segment.

Faire une version corrigée du défi, les valeurs attendus mais ne pas faire de quizz encore.