

Anthon_Dossier_Individuel_RMD

2025-02-04

ETAT : PAS FINI

Avant-propos

Ce dossier a été réalisé dans le cadre du module OPEN (Outil P E N) dispensé à l'ISARA en 2025 et coordonné par M. PAYET. Ce dossier a pour but de présenter un travail de recherche sous-format R.Mark.Down pour nous familiariser à l'outil et justifier de compétences nécessaires abordées au cours de l'optionnel.

Pour ce faire, nous allons réalisé un support d'information et d'explication sur le concept des *Expressions Régulières* communément appelé *Regex*, qui est utilisé dans plusieurs langages informatiques qui permettre l'identification de structures régulières au sein d'un texte. Puisque le module se fonde essentiellement sur du R, nous aborderons ce concept au travers de ce langage et feront un état des lieux des possibles dans d'autres langages informatique. L'objectif a termes est de fournir un dossier qui permet de permettre l'apprentissage des Regex à toutes personnes en ayant besoin et en fournissant les différents codes adéquats à leur utilisation.

Le REGEX

Le REGEX un concept pour définir l'Homme

Avant de parler d'expression régulières il faut parler d'anatomie. En effet, au cours des années 1940, Warren McCulloch et Walter Pitts souhaitent décrire le système nerveux humain par l'intermédiaire d'automates simples.

INSERER IMAGE REPRESENTATION NEURONE FORMEL

Imports

Les données utilisées sont issues des données disponibles sur R dans le packages *dplyr*. Dans un premier temps nous allons donc installer les packages avec puis les importer.

```
#install.packages("dplyr")  
#install.packages("ggplot2")
```

```
library(ggplot2)  
library(dplyr)
```

```
##  
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##   filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##   intersect, setdiff, setequal, union

data("starwars")
```

Présentation

Tout d'abord, il est nécessaire de réaliser un premier état des lieux des données. Pour ce faire nous allons utiliser la fonction `head()`

```
head(starwars)

## # A tibble: 6 x 14
##   name      height  mass hair_color skin_color eye_color birth_year sex  gender
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>      <dbl> <chr> <chr>
## 1 Luke Sky~    172    77 blond      fair        blue         19  male  mascu~
## 2 C-3PO        167    75 <NA>      gold        yellow        112 none  mascu~
## 3 R2-D2         96    32 <NA>      white, bl~  red          33  none  mascu~
## 4 Darth Va~    202   136 none      white      yellow       41.9 male  mascu~
## 5 Leia Org~    150    49 brown     light      brown         19  fema~  femin~
## 6 Owen Lars    178   120 brown, gr~ light      blue          52  male  mascu~
## # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

On remarque nous avons affaire à une table remplie de nombreuses informations de différents personnages de la saga "Star Wars". Parmi les informations fournies, nous avons : le nom, la taille, le poids, la couleur de cheveux, la couleur de peau, la couleur des yeux, le sexe, le genre, l'origine (homeworld), l'espèce ainsi que les films où ils sont présent, leurs véhicules et leurs vaisseaux.

Nous avons donc ici 14 caractéristiques de 87 personnages.

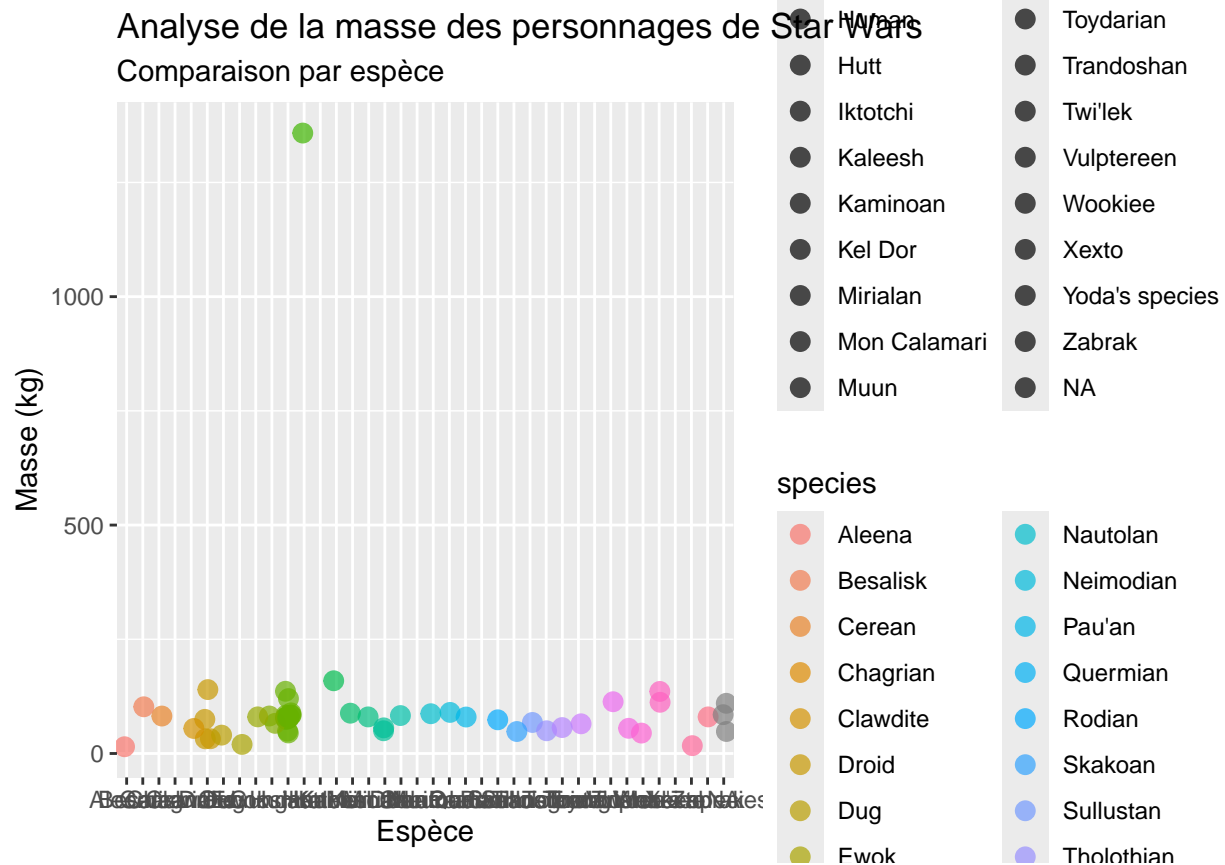
Méthodes et démarches

Le support de notre étude principale

On va une analyse assez simple car l'objectif c'est plus de perdre lecteur que de faire une analyse profonde. Pour ce faire, un simple plot qui met les espèces en abscisses en fonction de leur masse pourrait convenir. On va déjà voir ce que ça donne :

```
ggplot(starwars, aes(x = species, y = mass, fill = species)) +
  geom_jitter(aes(color= species), width = 0.2, size = 3, alpha=0.7)+
  ggtitle("Analyse de la masse des personnages de Star Wars")+
  labs(subtitle= "Comparaison par espèce",
       x = "Espèce", y = "Masse (kg)", fill = "Couleurs des Espèce")
```

```
## Warning: Removed 28 rows containing missing values or values outside the scale range
## ('geom_point()').
```



Dégrader le visuel

Maintenant qu'on a déjà une visualisation confuse des données, on va renforcer la difficulté de lire à partir de différentes commandes. Les explications de chaque ligne sont présentes dans le code. On va grossir les points car on arrive trop à les distinguer. Dans un second temps on change la police à 14 ensuite on va personnaliser les axes et les affichages. Afin de rester la typologie et les couleurs de star wars, on va mettre le fond en noir et mettre le titre et le sous-titre en jaune. Pour l'axe X, on met le texte à 45 degrés pour faire chevaucher les différents textes qui contiennent le nom des espèces. On va mettre un fond rouge et une écriture en cyan pour les textes. L'objectif est de rendre illisible, et difficile à regarder le graphique.

```
ggplot(starwars, aes(x = species, y = mass, fill = species)) +
  geom_jitter(aes(color= species), width = 0.2, size = 5, alpha=0.7)+
  ggtitle("Analyse de la masse des personnages de Star Wars")+
  labs(subtitle= "Comparaison par espèce",
    x = "Espèce", y = "Masse (kg)")+
  theme_minimal(base_size = 14)+
  theme(axis.text.x = element_text(angle=45, hjust =1, face="bold",color="white"),
    axis.text.y = element_text(face="bold", color= "white"),
    plot.background = element_rect(fill = "black"),
    panel.background = element_rect(fill="black"),
    legend.background = element_rect(fill="red"),
```

```

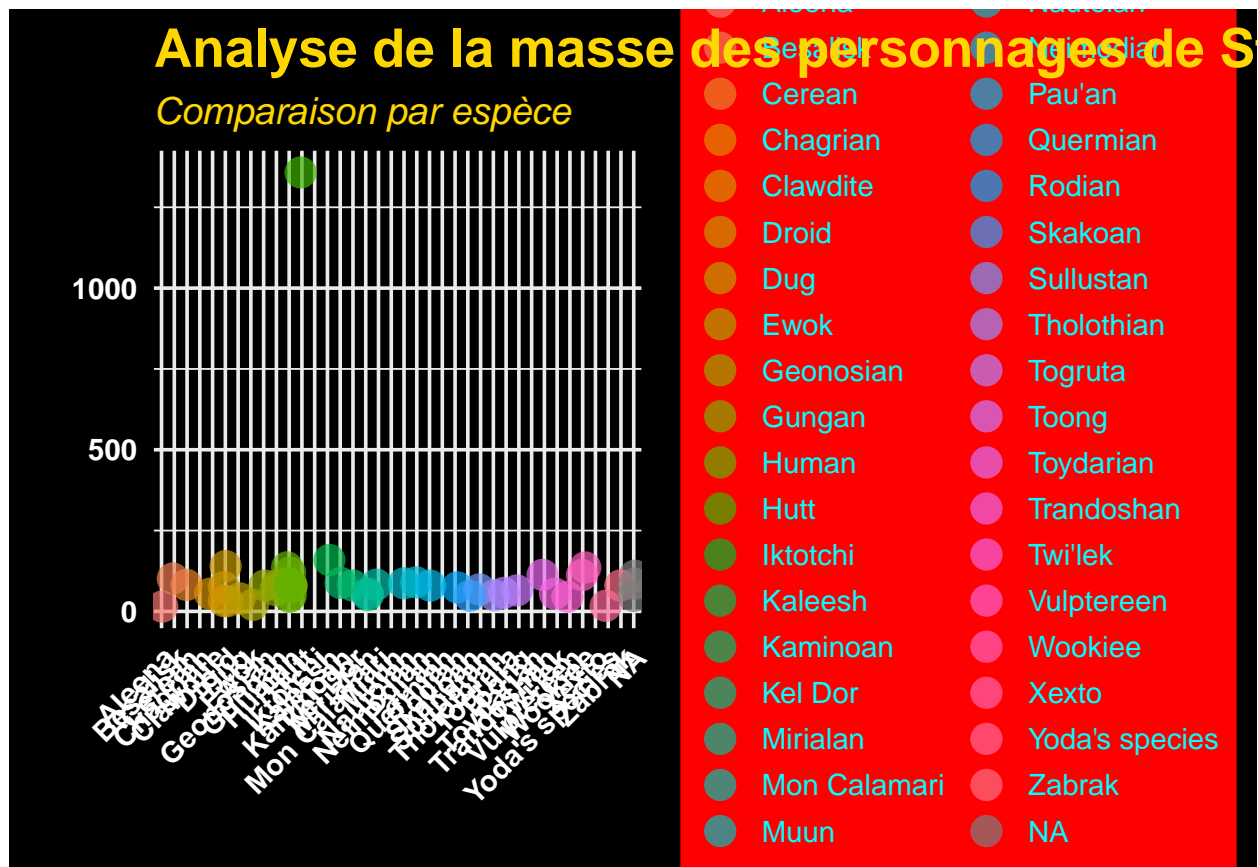
legend.text = element_text(color = "cyan"),
legend.title = element_text(face = "bold", color= "white"),
plot.title = element_text(face = "bold", size = 20, color = "gold"),
plot.subtitle = element_text(face = "italic", size = 14, color = "gold"))+
theme_minimal(base_size = 14)+
theme(axis.text.x = element_text(angle=45, hjust =1, face="bold",color="white"),
axis.text.y = element_text(face="bold", color= "white"),
plot.background = element_rect(fill = "black"),
panel.background = element_rect(fill="black"),
legend.background = element_rect(fill="red"),
legend.text = element_text(color = "cyan"),
legend.title = element_text(face = "bold", color= "white"),
plot.title = element_text(face = "bold", size = 20, color = "gold"),
plot.subtitle = element_text(face = "italic", size = 14, color = "gold"))

```

```

## Warning: Removed 28 rows containing missing values or values outside the scale range
## ('geom_point()').

```



Un petit test de biblio

En vitesse et pour montrer la compréhension voici deux petites références biblios selon le thème de des données [Auteur ?] Star Wars (2025). D'ailleurs, le point vert qui est tout en haut est lié à un personnage très charismatique "Jabba" dont sa page wiki de fan est disponible ici. [Auteur ?] *Jabba Desilijic Tiure* / *Star Wars Wiki / Fandom* ([sans date])



Figure 1: Fig. 1: Jabba sur un échiquier

Bibliographie

Liste des publications :

[Auteur ?] *Jabba Desilijic Tiure* / *Star Wars Wiki / Fandom*, [sans date]. [en ligne]. [Consulté le 4 février 2025]. Disponible à l'adresse : https://starwars.fandom.com/fr/wiki/Jabba_Desilijic_Tiure

[Auteur ?] *Star Wars*, 2025. [en ligne]. janvier 2025. [Consulté le 4 février 2025]. Disponible à l'adresse : https://fr.wikipedia.org/w/index.php?title=Star_Wars&oldid=222503327