tp2MapperDonnées

Mamadou Billo Diallo

2023-09-14

# Tâche 1 : Réflexion

-Il est important de visualiser les variables et les données pour comprendre ce que les données reflètent afin de: Faciliter la communication et la compréhension des données Faciliter la prise de décisions importantes Faciliter l’observation des anomalies;

-Mapper des données pour représenter graphiquement l’esthétique consiste a associé des variables importantes de la base de données afin de les visualiser tout en insérant des couleurs, des forme, des position… pour mieux lire le résultat de la visualisation;

-Les données mappées de la vidéo de Roslin sont des données qui décrivent l’avancée des pays pauvre et des pays riches (santé et richesse) au fil des ans de 1810 à 2009 et donc le Pib par habitant en fonction des espérances de la population. Les données ont été mappé esthétiquement avec une couleur qui décrit les continents, et une taille qui décrit la population tout en visualisant en forme de point;

# Tâche 2 : Le Seigneur des Anneaux

## Charger et nettoyer les données

Nous chargeons, restructurons et nettoyons d’abord les données.

# Vous n'aurez besoin que de la bibliothèque tidyverse pour cet exercice  
library(tidyverse)

## Warning: le package 'tidyverse' a été compilé avec la version R 4.2.2

# Charger les jeux de données séparés  
communauté <- read\_csv("data/The\_Fellowship\_Of\_The\_Ring.csv")  
tt <- read\_csv("data/The\_Two\_Towers.csv")  
rotk <- read\_csv("data/The\_Return\_Of\_The\_King.csv")  
  
# bind\_rows() empile les blocs de données les uns sur les autres  
lotr\_wide <- bind\_rows(communauté, tt, rotk) %>%  
 # Faites de la colonne Film une variable catégorielle (facteur) et placez-la dans  
 # ordonner l'apparition des catégories (pour que les films soient dans le bon ordre)  
 mutate(Film = fct\_inorder(Film))  
  
# Rendez ces données larges bien rangées  
lotr <- lotr\_wide %>%  
 # C'est la nouvelle façon de rendre les données longues  
 pivot\_longer(cols = c(Female, Male),  
 names\_to = "Sexe", values\_to = "Mots")

## Nombre total de mots prononcés par race

lotr\_Mots\_total<-lotr %>%  
 group\_by(Race)%>%  
 summarise(total=sum(Mots),  
 number=n())  
lotr\_Mots\_total

## # A tibble: 3 × 3  
## Race total number  
## <chr> <dbl> <int>  
## 1 Elf 3737 6  
## 2 Hobbit 8796 6  
## 3 Man 8712 6

Le nombre total de mots prononcés par race est: Elf:3737 Hobbit 8796 Man:8712 ##Nombre des mots prononcés par les hobbits masculins

lotr\_Mots\_sexe<-lotr %>%  
 group\_by(Race,Sexe)%>%  
 summarise(total=sum(Mots),  
 number=n())

## `summarise()` has grouped output by 'Race'. You can override using the  
## `.groups` argument.

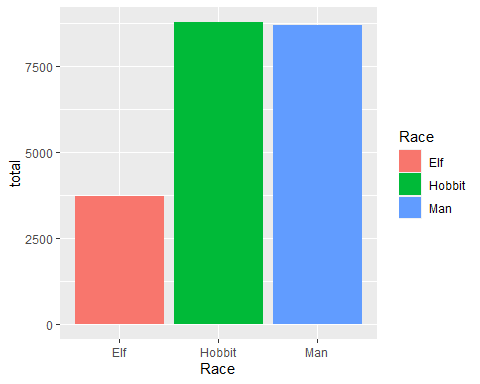
lotr\_Mots\_sexe

## # A tibble: 6 × 4  
## # Groups: Race [3]  
## Race Sexe total number  
## <chr> <chr> <dbl> <int>  
## 1 Elf Female 1743 3  
## 2 Elf Male 1994 3  
## 3 Hobbit Female 16 3  
## 4 Hobbit Male 8780 3  
## 5 Man Female 669 3  
## 6 Man Male 8043 3

Les hobbit masculin ont prononcés 8780 mots.

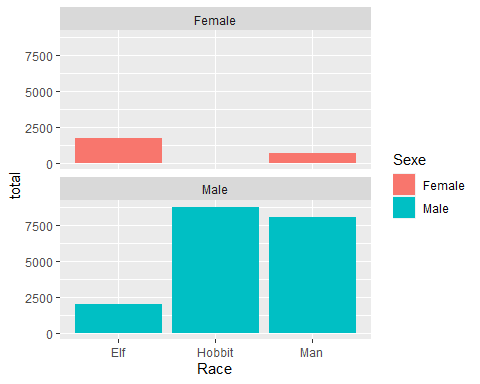
## representation par graphe du Nombre total de mots prononcés par race

ggplot(data=lotr\_Mots\_total,mapping=aes(x=Race,y=total,fill=Race))+  
geom\_col()

 Nous observons que les deux races Hobbit et Man atteignent la barre des 8000 mots contrairement à Elf qui atteigne 4000;

## representation par graphe du Nombre des mots prononcés par les hobbits masculins

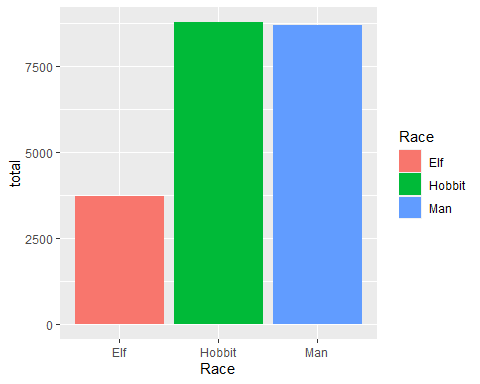
ggplot(data=lotr\_Mots\_sexe,mapping=aes(x=Race,y=total,fill=Sexe))+  
geom\_col(position = position\_dodge())+  
 facet\_wrap(vars(Sexe), ncol = 1)

 Nous remarquons la race Hobbit masculin qui a prononcé plus de 8000 mots.

## Course

Une certaine race domine-t-elle toute la trilogie ? (indice : regrouper par Race)

ggplot(data=lotr\_Mots\_total,mapping=aes(x=Race,y=total,fill=Race))+  
geom\_col()

 La Race Hobbit domine toute la triologie avec juste quelques mots au dessus de la Race Man;

## Genre et cinéma

Est-ce qu’un certain genre domine un film ? (lol bien sûr que c’est le cas, mais quand même, représentez-le graphiquement) (Astuce : groupez à la fois par Gender et Film.) Expérimentez avec le remplissage par Gender ou Film et le facettage par Gender ou Film.

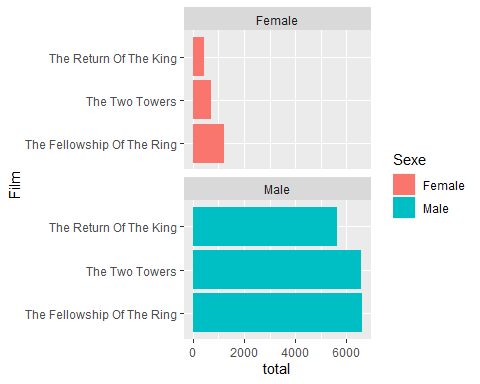
lotr\_Mots\_Gender\_Film<-lotr %>%  
 group\_by(Film,Sexe)%>%  
 summarise(total=sum(Mots),  
 number=n())

## `summarise()` has grouped output by 'Film'. You can override using the  
## `.groups` argument.

lotr\_Mots\_Gender\_Film

## # A tibble: 6 × 4  
## # Groups: Film [3]  
## Film Sexe total number  
## <fct> <chr> <dbl> <int>  
## 1 The Fellowship Of The Ring Female 1243 3  
## 2 The Fellowship Of The Ring Male 6610 3  
## 3 The Two Towers Female 732 3  
## 4 The Two Towers Male 6565 3  
## 5 The Return Of The King Female 453 3  
## 6 The Return Of The King Male 5642 3

ggplot(data=lotr\_Mots\_Gender\_Film,mapping = aes(x=total,y=Film,fill=Sexe))+  
 geom\_col(position = position\_dodge())+  
 facet\_wrap(vars(Sexe),ncol=1)



Nous pouvons observer que le sexe masculin domine

## Course et cinéma

La race dominante diffère-t-elle d’un film à l’autre ? (Astuce : regrouper à la fois par Race et Film.) Expérimentez avec le remplissage par Race ou Film et le facettage par Race ou Film.

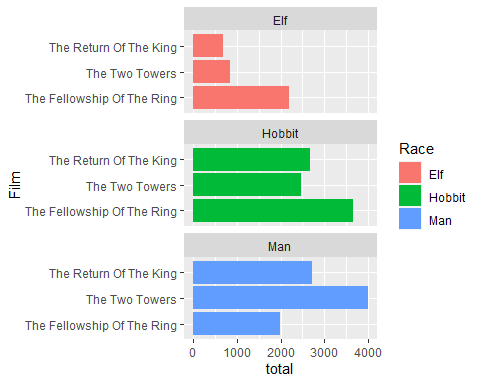
lotr\_Mots\_Race\_Film<-lotr %>%  
 group\_by(Race,Film)%>%  
 summarise(total=sum(Mots),  
 number=n())

## `summarise()` has grouped output by 'Race'. You can override using the  
## `.groups` argument.

lotr\_Mots\_Race\_Film

## # A tibble: 9 × 4  
## # Groups: Race [3]  
## Race Film total number  
## <chr> <fct> <dbl> <int>  
## 1 Elf The Fellowship Of The Ring 2200 2  
## 2 Elf The Two Towers 844 2  
## 3 Elf The Return Of The King 693 2  
## 4 Hobbit The Fellowship Of The Ring 3658 2  
## 5 Hobbit The Two Towers 2463 2  
## 6 Hobbit The Return Of The King 2675 2  
## 7 Man The Fellowship Of The Ring 1995 2  
## 8 Man The Two Towers 3990 2  
## 9 Man The Return Of The King 2727 2

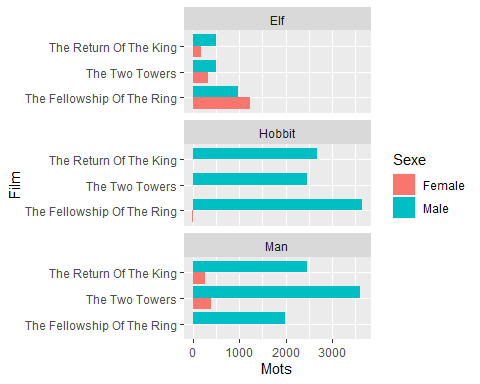
ggplot(data=lotr\_Mots\_Race\_Film,mapping = aes(x=total,y=Film,fill=Race))+  
 geom\_col(position=position\_dodge())+  
 facet\_wrap(vars(Race),ncol=1)

 Effectivement la race dominante diffère d’un film a un autre par exemple nous pouvons observer le Hobbit domine avec le film The Fellowship of The Ring tandisque pour le film The two Towers est dominé par la race Man;

## Race et genre et film

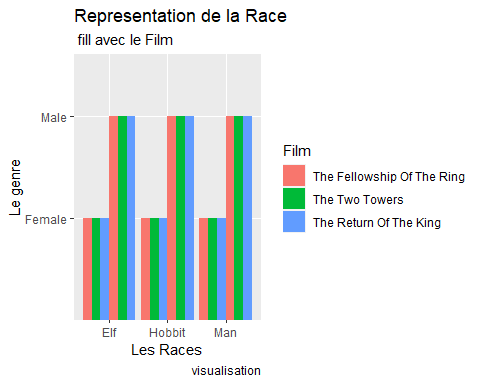
Créez un graphique qui visualise simultanément le nombre de mots prononcés par race, sexe et film. Utilisez le bloc de données “lotr” complet. Vous n’avez pas besoin de créer un nouvel ensemble de données résumées (avec group\_by(Race, Gender, Film)) car les données d’origine ont déjà une ligne pour chacune d’entre elles (vous pourriez créer un ensemble de données résumées, mais il serait identique à la version complète).

ggplot(data=lotr,mapping = aes(x=Mots,y=Film,fill=Sexe))+  
 geom\_col(position=position\_dodge())+  
 facet\_wrap(vars(Race),ncol = 1)

 Nous pouvons observer ce graphique qui visualise simultanément le nombre de mots prononcé par race,sexe et film;

Vous devez afficher Race, Gender et Film en même temps, mais vous n’avez que deux esthétiques possibles (x et fill), vous devrez donc également faire face à la troisième. Jouez avec différentes combinaisons (par exemple, essayez x = Race, puis x = Film) jusqu’à ce que vous en trouviez une qui raconte l’histoire la plus claire. Pour vous amuser, ajoutez un calque labs() pour ajouter un titre, un sous-titre et une légende.

ggplot(data=lotr,mapping = aes(x=Race,y=Sexe,fill=Film))+  
 geom\_col(position = position\_dodge())+  
 labs(title="Representation de la Race ",subtitle = " fill avec le Film",caption = "visualisation",x = "Les Races", y ="Le genre")



#NB: j'ai rechercher jai pas trouvé un moyen de le faire avec x et fill uniquement