

Introduction

Données

Les données viennent du site kaggle, nous les avons choisis car on voulait travailler sur un jeu-vidéo que nous aimons. Nous les avons également choisi car on a trouvé plusieurs dataset complémentaires et riche. Elles s'intègrent dans le contexte de l'esport et plus particulièrement dans le jeu-vidéo Overwatch

Nous avons 4 dataset différents donc nous allons faire 4 parties afin que ce soit plus clair. ## Overwatch 2 statistics

Les données ont été ajoutées par Mykhailo Kachan, qui les a pris du site Overbuff
Ce dataset présente des statistiques sur les héros Overwatch (pick rate, pourcentage de victoires, dégats, etc...) en fonction du niveau du joueur (de bronze à diamond) et des saisons.

Le dataset se compose de 131 variables et 288 observations ; les colonnes les plus précises ne seront pas étudiées car leur contenu trop pointu ne nous intéresse pas. Nous nous intéresserons à une dizaine de colonnes :

- Hero** : Nom du héros (type texte, ex. Ana, Mercy)
- Skill tier** : Définit le niveau de classement du joueur (type texte, 8 valeurs possibles : All, Bronze, Silver, Gold, Platinum, Diamond, Master ou Grandmaster)
- KDA Ratio** : La somme de kills et des assistances divisée par le nombre de fois où le joueur est mort (type numérique)
- Pick rate** : Le taux de sélection de chaque héros en jeu (type numérique)
- Win Rate** : Le taux de victoire de chaque héros (type numérique)
- Eliminations / 10min** : Nombre d'éliminations en moyenne sur 10 minutes de jeu (type numérique)
- Damage / 10 min** : Nombre moyen de dégâts du héros sur 10 minutes de jeu (type numérique)
- Healing / 10 min** : Nombre moyen de points de vie récupéré/redonné sur 10 min de jeu (type numérique)
- Role** : Rôle du héros en jeu (type texte, trois valeurs possibles : Damage, Tank ou Support)

Ce dataset est divisé en 5 parties : Une première reprend les infos ci-dessus en général depuis que le jeu est sorti. Les quatres suivantes correspondent aux mêmes données mais pour chaque saison.

eSports Earnings

Les données du dataset proviennent du site eSportsEarning.com , ce dataset comporte 3 parties distinctes :

- 1. country-and-continent-codes-list.csv **Nombre d'observations** (lignes) : 262
Nombre de variables (colonnes) : 6
Détails des variables :

Nom de variable	Définition de variable	Type de variable
Continent_Name	Nom du continent	type texte, 7 valeurs uniques, ex. Asia, Europe
Continent_Code	Code du continent	type texte, 6 valeurs uniques, ex. AS, EU
Country_Name	Nom complet du pays	type texte, 254 valeurs uniques
Two_Letter_Country_Code	Code pays à deux lettres	type texte, 253 valeurs uniques
Three_Letter_Country_Code	Code pays à trois lettres	type texte, 250 valeurs uniques
Country_Number	Numéro du pays	type numérique, 250 valeurs uniques

Les données peuvent être regroupées par continent (Continent_Name et Continent_Code), offrant une catégorisation géographique des pays (Country_Name, Two_Letter_Country_Code, Three_Letter_Country_Code). Ce fichier semble principalement fournir une correspondance entre les pays et leur appartenance continentale, sans sous-groupes internes évidents au-delà de cette classification géographique.

- 2. highest_earning_players.csv

Nombre d'observations: 96
Nombre de variables : 8
Détails des variables :

Nom de variable	Définition de variable	Type de variable
PlayerId	Identifiant du joueur	type numérique, 998 valeurs uniques
NameFirst	Prénom du joueur	type texte, 756 valeurs uniques
NameLast	Nom de famille du joueur	type texte, 637 valeurs uniques
CurrentHandle	Pseudo actuel du joueur	type texte, 990 valeurs uniques
CountryCode	Code pays du joueur	type texte, 56 valeurs uniques
TotalUSDPrize	Total des gains en USD	type numérique

Nom de variable	Définition de variable	Type de variable
Game	Jeu	type texte, 10 genres de jeu différents
Genre	Genre du jeu	type texte, 5 genres différents

Les données sont organisées autour des joueurs et leurs gains dans différents jeux vidéo. Catégories Principales : Les joueurs peuvent être catégorisés par le jeu (Game) auquel ils participent, et par genre de jeu (Genre), comme les jeux de tir à la première personne ou les arènes de bataille en ligne multijoueur. Sous-groupes Possibles : Au sein de chaque jeu ou genre, des analyses peuvent être effectuées sur des sous-groupes, tels que les joueurs par pays (CountryCode), ou les tranches de gains (TotalUSDPrize).

3. highest_earning_teams.csv

Nombre d’observations : 87

Nombre de variables : 6

Détails des variables :

Nom de variable	Définition de variable	Type de variable
TeamId	Identifiant de l’équipe	type numérique, 505 valeurs uniques
TeamName	Nom de l’équipe	type texte, 505 valeurs uniques
TotalUSDPrize	Total des gains en USD	type numérique
TotalTournaments	Total des tournois disputés	type numérique
Game	Jeu	type texte, 10 genres de jeu différents
Genre	Genre du jeu	type texte, 5 genres différents

Semblable aux données des joueurs, celles-ci sont organisées autour des équipes et de leurs performances financières dans le domaine des esports. Catégories Principales : Les équipes peuvent être catégorisées par jeu (Game) et par genre de jeu (Genre), offrant un aperçu des domaines où elles ont le plus de succès. Sous-groupes Possibles : Il pourrait y avoir des analyses sur le nombre de tournois disputés (TotalTournaments) par rapport aux gains totaux, ou des comparaisons entre les genres de jeux pour voir où les équipes réussissent le plus.

Overwatch Heroes

Overwatch_edopic.csv est un dataset décrivant tous les héros présents dans le jeu ainsi que leurs statistiques. C’est un dataset que nous avons trouvé sur le site kaggle. Les personnes ayant publié ce dernier n’ont pas précisé la provenance des données.

Ce fichier est composé de 8 variables et de 58 observations. A savoir qu’il n’y a pas 58 héros dans le jeu, mais quelques-uns possèdent plusieurs formes qui ont des statistiques différentes. Le jeu de données ayant été publié il y a 7ans, les héros qui sont arrivés depuis n’y figurent pas, il faudra donc prendre des pincettes quand aux résultats obtenus puisqu’un total de 12 personnages ne sont pas présents dans ce dataset, et auraient pu changer les résultats.

Description des variables : - Heroes : Nom des héros présents dans le jeu (type texte) - Damage par seconds : Nombre de dégâts infligés par seconde (type numérique) - Headshot DPS : Nombre de dégâts infligés par seconde lorsque tous les tirs touchent à la tête (type numérique, certaines observations sont vides car plusieurs héros ne peuvent pas mettre de « Headshot ») - Single shot : Nombre de dégâts causés par coup/balle (type numérique) - Nanoboost DPS : Nombre de dégâts infligés lorsque le personnage voit ses dégâts augmentés par le Nanoboost de Ana (sa capacité ultime) (type numérique) - Health per Sec : Nombre de soins prodigués par secondes (type numérique, 12 valeurs uniques) - Life : Nombre de points de vie (type numérique) - Reload : Nombre de balles par chargeur (type numérique)

Professional Overwatch Player Mouse/Sens. Settings

Ce dataset vient de 3 autres datasets qu’une personne à réuni pour son plaisir. Il présente des statistiques à propos des joueurs professionnels d’Overwatch et de leur sensibilité. Principalement la DPI, le modèle de souris ou encore la sensibilité dans le jeu.

Dans ce premier dataset nous avons 3 fichiers différents, le premier contient 99 observations 6 variables :

- Model : modèle de la souris (type texte, 99 valeurs uniques, ex. Logitech G Pro, Raze Viper) - Size : taille de la souris (type caractère, 5 valeurs uniques, ex. XS, M, L) - Length : longueur de la souris en cm (type numérique) - GripWidth : la taille de la zone où les utilisateurs vont tenir la souris, en centimètres (type numérique) - Weight : poids la souris en gramme (type numérique) - WeightToSizeRatio : ratio entre le poids de souris et la longueur fois la taille de la zone ou on tient la souris (type numérique)

Le deuxième fichier contient 173 observations et 7 variables :

- ID : pseudo du joueur (type texte, 173 valeurs uniques) - Team : nom de l’équipe professionnel du joueur (type texte) - Status : statut actif ou non (type texte, 1 valeur unique) - Roles : rôle du joueur (type texte, 3 valeurs uniques, ex. Support, Tank) - HaveSettings : indique si les paramètres du joueurs sont présents dans le fichier suivant (type booléen, 2 valeurs uniques) - Hero : le héro principal joué par le joueur (type texte, 23 uniques valeurs) - TwitchStream : lien vers le twitch du joueur (type texte)

Le troisième fichier contient 148 observations et 12 colonnes :

- Player : pseudo du joueur (type texte, 148 valeurs uniques) - IsProfessional : vrai si leur ID est présent dans le fichier précédent (type booléen) - Role : rôle du joueur (type texte, 3 valeurs uniques, ex. Tank, Support, Damage) - Hero : le héro principal joué par le joueur (type texte) - MouseModel : modèle de la souris (type texte) - MouseWeight : poids de la souris (type numérique) - MouseWeightToSizeRatio : ratio entre le poids de souris et la longueur fois la taille de la zone ou on tient la souris (type numérique) - DPI : dpi des joueurs (type numérique) - Sensitivity : sensibilité en jeu (type numérique) - eDPI : la dpi multiplié par la sensibilité (type numérique) - cmPer360 : nombre de centimètre d’un mouvement horizontal de la souris pour faire un tour complet dans le jeu (type numérique) - TwitchStream : lien vers le twitch du joueur (type texte)

Plan d'analyse

Overwatch 2 statistics

Voici quelques exemple d'interrogations sur ce dataset :

- Quels héros sont les plus/moins joués ?

Cette question seule sera difficile à analyser ; on aura des informations sur les héros les plus joués sans pouvoir définir pourquoi. Mais elle deviendra utile liée à d'autres questions ci-dessous, qui pourront (ou non) nous aider à comprendre les raisons poussant les joueurs à jouer certains héros plus que d'autres.

- Quels héros provoquent le plus/moins de dégâts ?

Pour cette question, on s'attend à ce que les héros les plus offensifs (ayant le rôle "damage") soient ceux qui provoquent le plus de dégâts en moyenne. Au contraire, nous nous attendons à voir que les héros défensifs (rôle "support") provoquent bien moins de dégâts, sachant que ce n'est pas leur but premier.

- Quels héros ont le plus/moins de taux de victoire ?

Il est difficile de prévoir le résultat de cette visualisation. Pour autant, elle peut être intéressante à lier à d'autres questions, ou en la spécifiant plus (par exemple, en se demandant quels sont les rôles des héros ayant le plus de victoires en moyenne, etc.)

- Comment le ratio des héros a évolué en fonction des saisons ? (augmentation/diminution)

Étudier l'évolution du KDA ratio des héros en fonction des saisons dans Overwatch pourrait permettre de comprendre les changements de popularité des héros au fil du temps (première question). Cela offre un aperçu des héros qui ont gagné en impact aidant à comprendre les tendances du jeu en fonction des saisons.

Dans l'ensemble, toutes ces questions peuvent être découpées : - En fonction du niveau des joueurs (skill tier) - En fonction des différentes saisons

Permettant d'étudier sous différents angles toutes nos données.

eSports Earnings

Voici quelques exemples d'interrogations sur ce dataset :

- Moyenne des gains des joueurs en fonction de leur nationalité ?

Cette analyse vise à déterminer si certaines nationalités sont plus susceptibles de gagner des sommes d'argent plus élevées dans le monde d'Overwatch 2. Cela pourrait révéler des tendances intéressantes, comme la possibilité que certaines régions offrent un meilleur soutien, formation, ou plus d'opportunités pour leurs joueurs. Les résultats pourraient indiquer des écarts de performance ou de popularité du jeu selon les pays.

- Moyenne des gains des joueurs par continent ?

Semblable à l'analyse par nationalité, mais à une échelle plus large, cette question cherche à comparer les gains moyens sur différents continents. Cela peut aider à comprendre comment les infrastructures régionales, la popularité du jeu, et le niveau de compétition influencent les récompenses financières. Cela pourrait également révéler des continents où le jeu est plus compétitif ou plus prisé.

- Moyenne de gains des équipes par tournoi ?

Cette analyse vise à évaluer le succès financier des équipes à travers différents tournois. Cela peut révéler quelles équipes dominent la scène d'Overwatch 2 et comment la distribution des prix varie entre les tournois. Cela peut également aider à identifier des tournois particulièrement lucratifs ou compétitifs.

- Nationalités des joueurs les mieux payés ?

En identifiant les nationalités des joueurs les mieux rémunérés, on peut tenter de déduire non seulement où se trouvent les talents les plus prometteurs ou les plus accomplis, mais aussi où le jeu est le plus valorisé. Cette information peut indiquer des tendances de recrutement pour les futures équipes ou des zones cibles pour le développement des ligues et des tournois.

Overwatch Heroes

Voici quelques exemples d'interrogations sur ce dataset :

- Quel héros fait le plus de dégâts ?

Nous pouvons penser que ce sera un DPS puisque leur rôle dans une partie est de faire un maximum de kills et de dégâts. Sachant que Bastion possède une capacité qui lui permet de se transformer en tourelle et tirer avec une énorme cadence, nous pouvons nous attendre que ce soit lui qui sorte en tête du classement. Pour répondre à cette question nous n'aurons qu'à utiliser la variable « Damage per sec » . Afin de donner plus de sens à cette question, nous pourrions élire le personnage qui fait le plus de dégâts par rôle (Tank, DPS, Healer). Pour cela, il faudrait avoir le rôle de chaque personnage pour pouvoir les trier selon leur rôle, mais cette variable n'est pas présente dans le dataset, nous devons donc la rajouter nous-même.

- Quel héros fait le plus de dégâts ?

Nous pouvons penser que ce serait un Tank car ces derniers sont ceux qui sont sensés pouvoir encaisser le plus de dégâts, ceux qui sont en ligne

de front. Nous pensons donc à Reinhardt. Pour répondre à cette question, nous aurons à utiliser la variable « life ». Afin de donner plus de sens à cette question, nous pourrions élire le personnage qui a le plus de vie par rôle (Tank, DPS, Healer). Pour cela, il faudrait avoir le rôle de chaque personnage pour pouvoir les trier selon leur rôle, mais cette variable n'est pas présente dans le dataset, nous devons donc la rajouter nous-même.

- Quel est le héros avec le plus de soins par seconde ?

Cette question est importante car elle nous permettrait de savoir quel personnage a le plus de potentiel de soin pour son équipe. Nous nous attendons donc logiquement à ce que ce soit un healer. Nous pouvons penser que le personnage que nous cherchons soit Ange, car celle-ci délivre un flux de soin continu. Pour répondre à cette question, nous utiliserons la variable « Health per Sec ». Le seul problème que nous pourrions rencontrer serait que dans ce dataset se trouvent tous les personnages, ainsi que les formes alternatives qu'ils peuvent prendre selon leurs pouvoir ou leur capacité ultime. Il faudra donc départager les personnages, des personnages sous capacité pour avoir un résultat juste.

Professional Overwatch Player Mouse/Sens. Settings

Voici quelques exemples d'interrogations sur ce dataset :

- Est-ce que les joueurs qui jouent le même rôle ont des sensibilités similaires ?

Nous pensons que oui car chaque rôle a tendance à jouer de la même façon donc les sensibilité des joueurs devraient se ressembler. Pour cette interrogation nous utiliserons les variables Rôle, DPI et Sensitivity. Le problème que nous pourrions rencontrer c'est que des joueurs peuvent avoir une DPI élevée et une sensibilité en jeu basse et certains peuvent avoir l'inverse, mais ces deux paramètres peuvent "renvoyer" une sensibilité similaire dans le jeu.

- Est ce que les joueurs du rôle damage ont une plus grosse sensibilité afin de pouvoir mettre des flickshot ?

Nous pensons que oui car la plupart des héros de type damage sont au cœur du combat et ont besoin de réagir très vite. Pour cette question nous utiliserons les variables Rôle, DPI et Sensitivity. Le problème que nous pourrions rencontrer est le même que celui de l'interrogation précédente.

- Est ce que les joueurs damage privilégient les souris plus légères contrairement aux autres rôles ?

Dans un premier temps nous aurions eu tendance à penser que oui mais finalement nous ne sommes pas sûrs que le rôle du joueur impacte à ce point le poids de la souris. Nous avons donc envie de savoir ce qu'il en est. Pour cette interrogation nous utiliserons Rôle et MouseWeight. Il n'y a pas de problème particulier à part qu'il n'y ait aucune relation entre le rôle et le poids de la souris et donc que nous puissions rien en tirer.

Analyse du dataset

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(readr)
```

```
#Importation des données  
player_data <- read_csv("./data/player_settings.csv")
```

```
## Rows: 148 Columns: 12  
## —— Column specification ——  
—  
## Delimiter: ","  
## chr (5): Player, Role, Hero, MouseModel, TwitchStream  
## dbl (6): MouseWeight, MouseWeightToSizeRatio, DPI, Sensitivity, eDPI, cmPer360  
## lgl (1): IsProfessional  
##  
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data <- read_csv("./data/ow2_quickplay_heroes_stats__2023-05-06.csv")
```

```
## Rows: 288 Columns: 131
## —— Column specification ——
—
## Delimiter: ","
## chr (3): Hero, Skill Tier, Role
## dbl (128): KDA Ratio, Pick Rate, %, Win Rate, %, Eliminations / 10min, Objec...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
players <- read.csv("./data/highest_earning_players.csv")
country_codes <- read.csv("./data/country-and-continent-codes-list.csv")
heroes <- read_csv("./data/Overwatch_edopic.csv")
```

```
## Rows: 58 Columns: 8
## —— Column specification ——
—
## Delimiter: ","
## chr (3): Heroes, Life, Reload
## dbl (5): Damage per second, Headshot DPS, Single shot, Nanoblast DPS, Health...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1) Quel sont les 10 héros qui ont le plus haut taux de dégâts par seconde ?

Pour répondre à cette question, nous nous appuyerons sur le nombre de dégâts par seconde infligés par chaque héro. On peut s'attendre à trouver une grande partie de personnages Damage dans le lot puisque c'est leur utilité première. Nous ne prendrons dans ce graphique que des héros sans capacité (capacité ultime) puisque ces dernières peuvent fausser les stats.

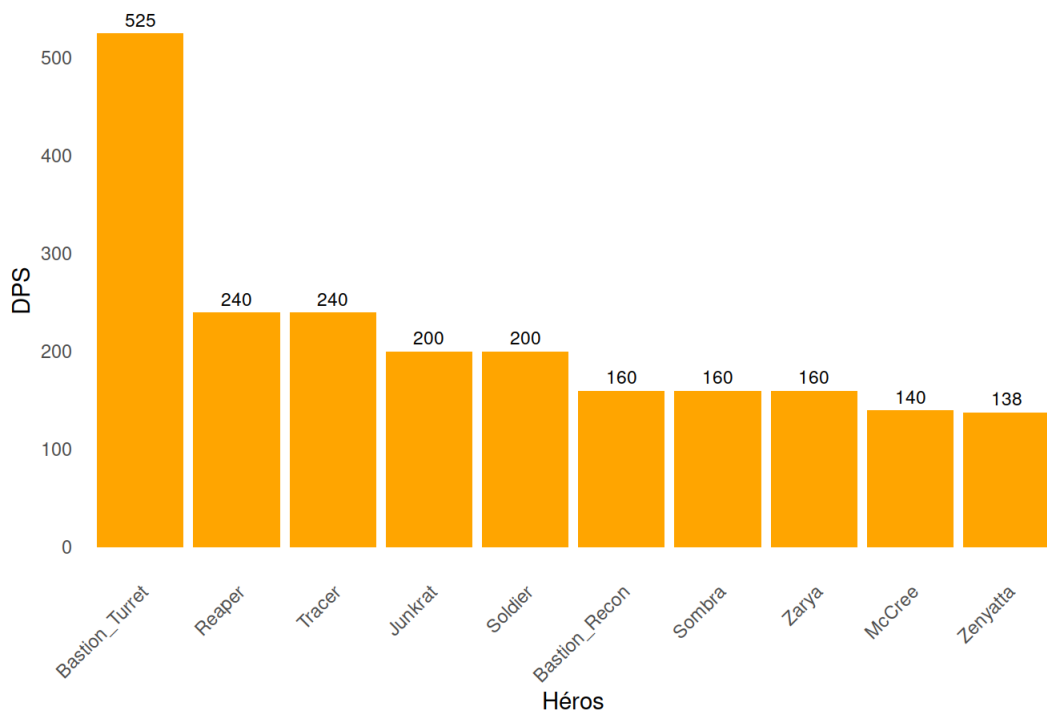
```
heroes <- rename(heroes, Damage_per_sec = `Damage per second`)
heroes <- mutate(heroes, DPS = Damage_per_sec)

#On garde Bastion qui est une exception : il possède deux formes de base dans le jeu
heroes <- mutate(heroes, Heroes = case_when(
  Heroes == "Bastion Recon" ~ "Bastion_Recon",
  Heroes == "Bastion Turret" ~ "Bastion_Turret",
  TRUE ~ Heroes
))
#supprimer tous les héros utilisant une capacité, ex : lucio soundwave
heroes <- filter(heroes, !grepl(" ", Heroes))
heroes <- arrange(heroes, desc(DPS))

top_10_heroes <- head(heroes, 10)

ggplot(top_10_heroes, aes(x = reorder(Heroes, -DPS), y = DPS, label = DPS)) +
  geom_bar(stat = "identity", fill = "orange") +
  geom_text(position = position_dodge(width = 0.9), vjust = -0.5, size = 3) +
  labs(title = "Top 10 héros avec le plus de DPS (Hors capacités)",
       x = "Héros",
       y = "DPS") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        plot.background = element_blank())
```

Top 10 héros avec le plus de DPS (Hors capacités)



Observations : A l'exception de Zarya (Tank), et Zenyatta (Soin), nous retrouvons 8 héros de type Dégât dans le top 10, voilà qui confirme nos suppositions. On remarque par contre qu'un héro réalise beaucoup plus de DPS que n'importe quel autre. On pourrait, à première vue, penser que le jeu n'est pas bien équilibré, mais cela s'explique par le fait que Bastion possède deux formes, dont la forme tourelle (Bastion Turret), qui fait de lui une tourelle, assez large et immobile qui tire avec une énorme cadence. Certes le personnage fait beaucoup de dégâts par seconde, mais il est gros et très facilement touchable en contrepartie.

2) Quel rôle de héros provoque le plus de dégâts en jeu ?

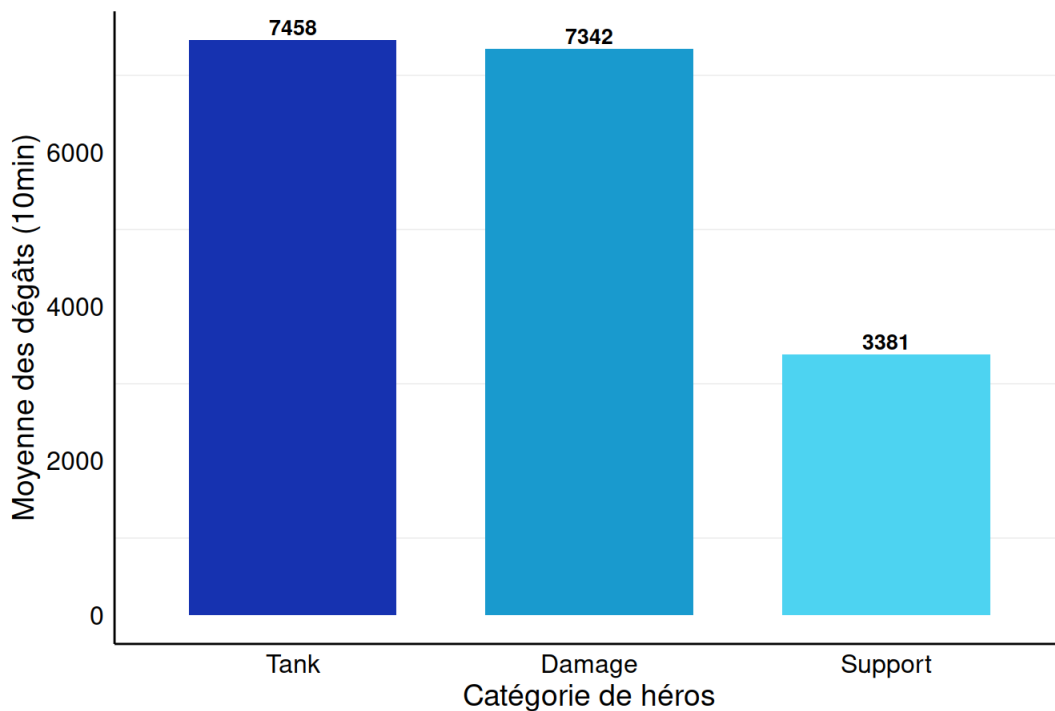
Pour répondre à cette question, nous nous appuyerons sur le nombre de dégâts infligés par les joueurs en 10 minutes, en fonction du rôle de leur héros. On s'attend ici à trouver le rôle Damage en première place ; c'est en effet son but principal de provoquer des dégâts. Au contraire, les dégâts des héros Support devraient être bien inférieurs aux autres, leur but étant de rester à l'arrière du combat, principalement pour soigner leurs coéquipiers.

Nous utiliserons un diagramme simple en colonne, pour représenter la moyenne des dégâts infligés en 10 min pour chaque rôle

```
#Calcul de la moyenne
AverageDamage <- data %>%
  group_by(Role) %>%
  summarize(AverageDamage = mean(`Damage / 10min`))

# Graphique
ggplot(AverageDamage, aes(x = reorder(Role, -AverageDamage), y = AverageDamage, fill = Role)) +
  geom_col(width = 0.7, color = NA, position = "dodge", fill = c("#199ace", "#4dd3f1", "#1632b0")) +
  labs(title = "Moyenne des dégâts par catégorie de héros", # Titres
       x = "Catégorie de héros",
       y = "Moyenne des dégâts (10min)") +
  theme_minimal() + # Thème minimal
  theme(legend.position = "none",
        panel.grid.major = element_blank(), # Pas de lignes de grille majeures
        axis.line = element_line(color = "black"), # Couleur de la ligne d'axe
        axis.text = element_text(size = 12, color = "black"), # Taille et couleur du texte de l'axe
        axis.title = element_text(size = 14, color = "black"), # Taille et couleur du titre de l'axe
        plot.title = element_text(size = 16, hjust = 0.5, color = "black")) + # Taille et couleur du titre du graphique
  geom_text(aes(label = round(AverageDamage)), vjust = -0.3, size = 3.5, color = "black", fontface = "bold")
```

Moyenne des dégâts par catégorie de héros



Plusieurs observations peuvent être faites grâce à ce graphique :

Tout d'abord, les héros de type support ont en effet une moyenne de dégâts bien moins élevées que celle des autres rôles. C'est un résultat qui est attendu et sans surprise.

Par contre, contrairement à nos hypothèses, on peut remarquer le rôle Damage ne se démarque pas du rôle Tank. C'est même ce dernier qui à la moyenne la plus haute. Ce résultat peut paraître étonnant si on regarde les résultats de la question précédente ; en effet, au niveau des dégâts par seconde, les héros de type Damage sont les plus puissants.

Mais la question précédente prend en compte seulement les statistiques théoriques des héros. Dans le cas de ce diagramme, les données sont des statistiques réelles, prises en jeu. Alors même si les héros de rôle Damage sont plus puissant en théorie, les type Tank sont toujours au front. Ils ont plus de vie, et donc plus de temps pour mettre des dégâts aux autres joueurs avant de mourir, même si leurs attaques sont moins puissantes. Ceci explique donc l'équilibre entre les dégâts des Tank et des Damage sur 10 minutes de jeu.

3) Est-ce que les joueurs qui joue le même rôle ont des sensibilités similaires ?

On peut effectivement s'attendre à ce que les mêmes rôles partagent globalement la même façon de jouer et ont donc des sensibilités similaires.

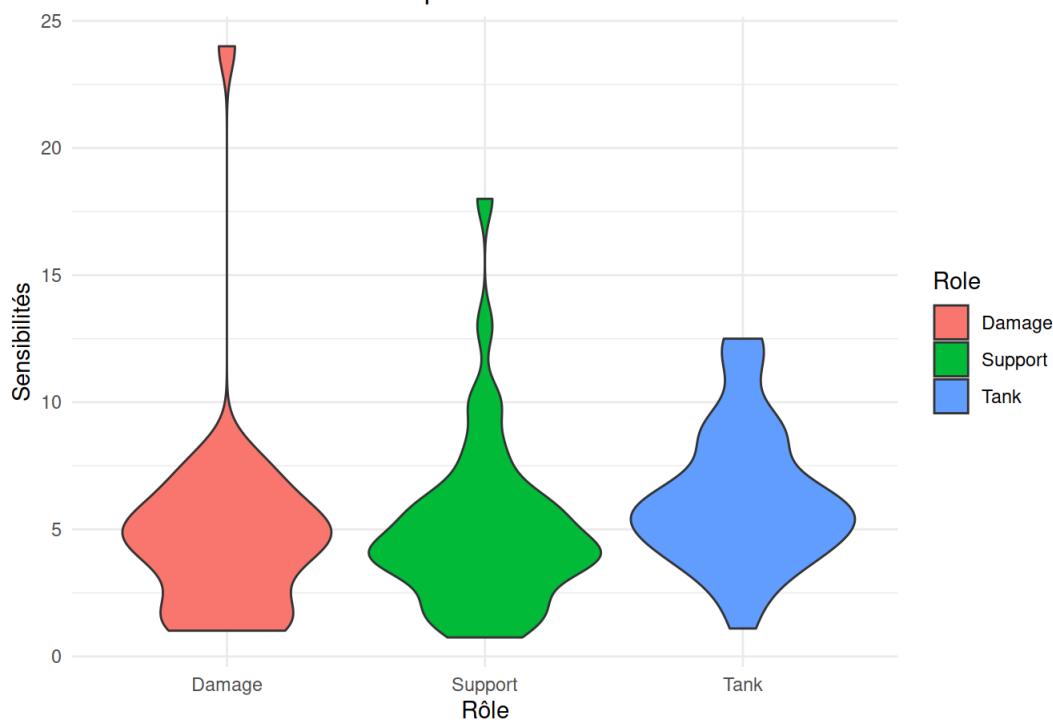
```
dpi_reference <- 800

player_data$Sensitivity <- as.numeric(player_data$Sensitivity)
player_data$DPI <- as.numeric(player_data$DPI)

player_data <- player_data %>%
  mutate(ratio_sensibilite = Sensitivity / DPI,
         sensibilite_standardisee = ratio_sensibilite * dpi_reference)

damage_data <- player_data %>% filter(Role == "Damage")
tank_data <- player_data %>% filter(Role == "Tank")
support_data <- player_data %>% filter(Role == "Support")
ggplot(player_data, aes(x = Role, y = sensibilite_standardisee, fill = Role)) +
  geom_violin() +
  labs(title = "Distribution des sensibilités par rôle",
       x = "Rôle", y = "Sensibilités") +
  theme_minimal()
```

Distribution des sensibilités par rôle



Tout d'abord il faut préciser que dans le jeu de données nous avons 2 informations sur la sensibilité des joueurs. On a la DPI de leur souris et leur sensibilité dans le jeu. Le problème est que ces 2 valeurs n'ont pas du tout le même ordre de grandeur et donc on ne pouvait pas les comparer même en les multipliant par exemple. On a donc réfléchi à comment avoir des valeurs comparables et on a fait un ratio en divisant la sensibilité par la DPI et en multipliant par une DPI de référence ici 800.

Pour le rôle Damage, les sensibilités sont étendues mais reste tout de même entre 2 et 8-9. Il y a un ou deux joueurs qui ont une sensibilité vraiment différente et bien plus élevée.

Pour le rôle Support, les sensibilités sont déjà plus étendues allant de 2 à 11 avec quelques exceptions au-dessus de 11 jusqu'à 16.

Et pour le rôle Tank, elles sont aussi relativement étendues allant de 2 à 15 avec tout de même beaucoup de joueurs aux alentours de 6.

Pour conclure les joueurs qui jouent le même rôle n'ont pas forcément la même sensibilité car on peut observer une large gamme dans les trois rôles. On observe quand même une sensibilité avec une fréquence élevée pour les rôles Support et Tank à 4 et 6 respectivement.

4) Quelle est la moyenne des gains des joueurs en fonction de leur nationalité ?

Pour répondre à cette question, nous allons d'abord fusionner les données des joueurs avec les données des pays pour obtenir le nom complet des pays. Ensuite, nous allons agréger les gains des joueurs par nationalité et calculer la moyenne des gains pour chaque pays. Enfin, nous allons créer un graphique à barres horizontales pour visualiser les moyennes des gains par nationalité. Nous nous attendons à voir des disparités significatives dans les gains moyens des joueurs en fonction de leur nationalité. Les pays avec une industrie de l'e-sport plus développée ou des opportunités de tournois plus nombreuses pourraient avoir des gains moyens plus élevés. À l'inverse, les pays avec moins de soutien financier ou d'infrastructures pourraient avoir des gains moyens plus bas.

```
overwatch_players <- players[players$Game == "Overwatch", ]
players$CountryCode <- tolower(overwatch_players$CountryCode)
country_codes$Country_Name <- sub(".*", "", country_codes$Country_Name)
```

```
head(country_codes$Country_Name)
```

```
## [1] "Afghanistan"
## [2] "Albania"
## [3] "Antarctica (the territory South of 60 deg S)"
## [4] "Algeria"
## [5] "American Samoa"
## [6] "Andorra"
```



```

country_codes$Two_Letter_Country_Code <- tolower(country_codes$Two_Letter_Country_Code)

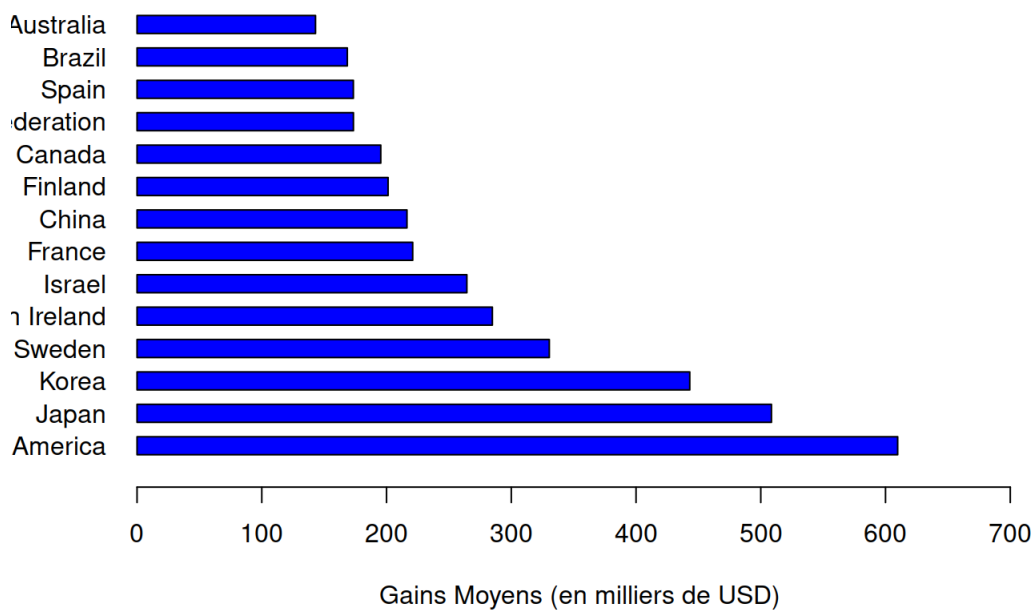
merged_data <- merge(players, country_codes, by.x = "CountryCode", by.y = "Two_Letter_Country_Code", all.x = TRUE)

average_earnings <- aggregate(TotalUSDPrize ~ Country_Name, data = merged_data, FUN = mean, na.rm = TRUE)
average_earnings <- average_earnings[order(-average_earnings$TotalUSDPrize),]

barplot(average_earnings$TotalUSDPrize / 1000,
        names.arg = average_earnings$Country_Name,
        las = 1,
        main = "Moyenne des Gains par Nationalité des Joueurs d'Overwatch",
        xlab = "Gains Moyens (en milliers de USD)",
        col = "blue",
        horiz = TRUE,
        space = 0.8,
        xlim = c(0, max(average_earnings$TotalUSDPrize / 1000) + 100),
        ylim = c(0, 25),
        cex.names = 1)

```

Moyenne des Gains par Nationalité des Joueurs d'Overwatch



Observations : On peut observer que les gains moyens des joueurs d'Overwatch varient considérablement en fonction de leur nationalité. Les États-Unis, la Corée du Sud et la Chine sont les pays avec les gains moyens les plus élevés, tandis que les pays européens et sud-américains ont des gains moyens plus bas. Cela peut refléter les différences dans les infrastructures de l'e-sport, les opportunités de tournois et le soutien financier pour les joueurs professionnels dans ces pays.