

Livrable 1 - Pokemon

Musset Aurélien
Arulraj Sinthujan
Schummer Lucas
Khuu Sophie

2025-04-20

1. Visualisation des différents combos de types de Pokémon

Objectif du graphique :

Ici, nous allons répondre à la question suivante : **Quel est le nombre de Pokémon par combinaison de types ?** En effet, les pokémons possèdent un type principal (type1) et un type secondaire (type2). Le but est de voir ici, si ces types sont équitablement répartis, où si on peut observer des tendances. Nous allons donc créer une heatmap permettant de visualiser les combinaisons de types de Pokémon.

Présentation :

Pour réaliser ce graphique, nous avons utilisé les données du dataset “Pokemon_data”. Nous avons filtré les Pokémon qui n’ont pas de type secondaire (type2 = “None”). Nous avons ensuite compté le nombre de Pokémon par combinaison de types, puis nous avons ajouté des totaux pour chaque type principal et secondaire.

Code et Visualisation :

```
library(ggplot2)
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(tidyr)

#Ouverture du dataset
df <- read.csv("data/Pokemon_data.csv", sep = ",")

#On ne garde que les colonnes nécessaires : name, type1 et type2
df <- df %>%
```

```

select(name, type1, type2) %>%
mutate(type1 = as.character(type1), type2 = as.character(type2))

# Comptage
df_filtered <- df %>%
  filter(type2 != "None") %>%
  count(type1, type2)

# Totaux
row_totals <- df_filtered %>%
  group_by(type1) %>%
  summarise(n = sum(n)) %>%
  mutate(type2 = "Total")

col_totals <- df_filtered %>%
  group_by(type2) %>%
  summarise(n = sum(n)) %>%
  mutate(type1 = "Total")

# Fusion
df_augmented <- bind_rows(df_filtered, row_totals, col_totals)

# Marquage des cellules de total
df_augmented <- df_augmented %>%
  mutate(
    is_total = type1 == "Total" | type2 == "Total",
    fill_value = ifelse(is_total, NA, n),
    text_color = ifelse(is_total, "black", "gray10"),
    border_color = ifelse(is_total, "black", "white")
  )

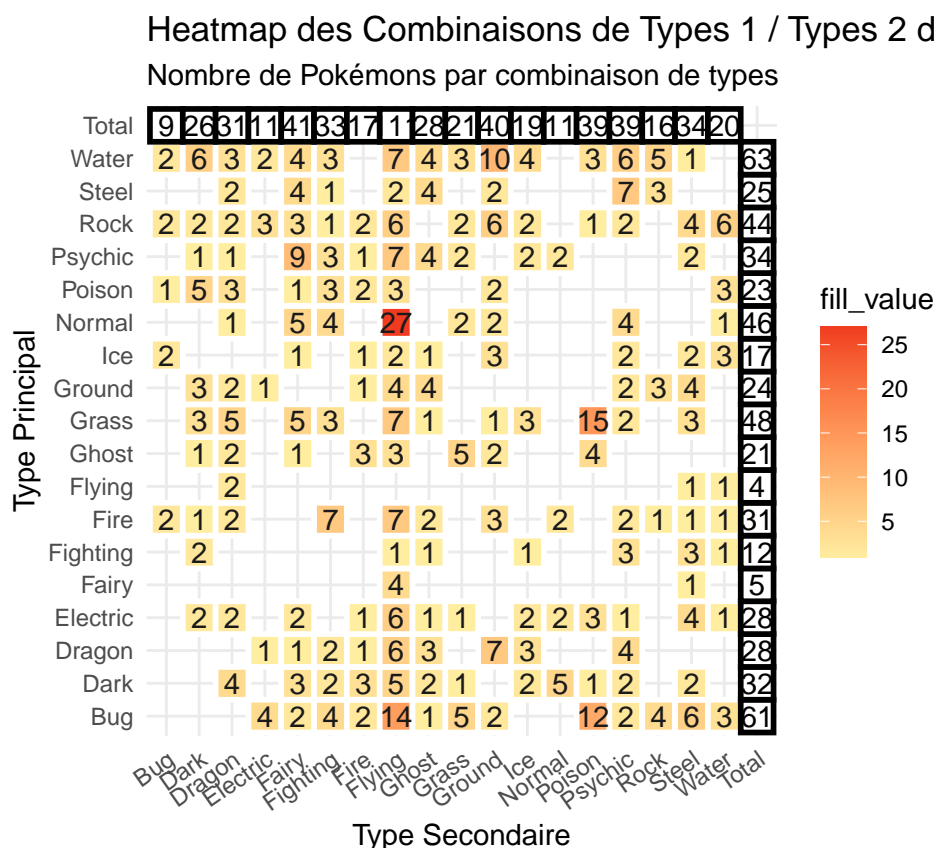
# Forcer l'ordre pour mettre "Total" à droite / en haut
type1_levels <- c(sort(unique(df_augmented$type1[df_augmented$type1 != "Total"])), "Total")
type2_levels <- c(sort(unique(df_augmented$type2[df_augmented$type2 != "Total"])), "Total")

df_augmented <- df_augmented %>%
  mutate(
    type1 = factor(type1, levels = type1_levels),
    type2 = factor(type2, levels = type2_levels)
  )

# Heatmap finale
ggplot(df_augmented, aes(x = type2, y = type1)) +
  geom_tile(aes(fill = fill_value), color = df_augmented$border_color, size = 1) +
  geom_text(aes(label = n, color = text_color), show.legend = FALSE) +
  scale_fill_gradient(low = "#ffeda0", high = "#f03b20", na.value = "white") +
  scale_color_identity() +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1)) +
  labs(
    title = "Heatmap des Combinaisons de Types 1 / Types 2 des Pokémon avec un Type Secondaire",
    subtitle = "Nombre de Pokémon par combinaison de types",
    x = "Type Secondaire",

```

```
y = "Type Principal"
) +
coord_fixed()
```



Analyse :

Nous pouvons observer que la majorité des Pokémon possèdent un type principal de type “Eau” (Water) et un type secondaire de type “Vol” (Flying). Nous pouvons également observer que les Pokémon de type “Normal” (Normal) possèdent le plus fréquemment un type secondaire de type “Vol” (Flying) et que les Pokémon de type “Plante” (Grass) possèdent souvent un type secondaire de type “Poison” (Poison). **Ce sont les combos les plus fréquents.** On peut également remarquer que de nombreux combos n'existent pas, comme par exemple “Poison” (Poison) et “Acier” (Steel), ou encore “Fée” (Fairy) et “Combat” (Fighting). Cela peut sembler logique pour certaines combinaisons (par exemple “Feu” (Fire) et “Eau” (Water)), mais il est intéressant de le noter. Malgré tout, on peut remarquer que la plupart des combinaisons sont représentées par la même échelle de couleur, ce qui signifie qu'il n'y a pas de combo de type qui se démarque vraiment des autres (hormis ceux notés au début).

2. Visualisation de la popularité des jeux vidéos Pokémon à l'aide du nombre de ventes de 1996 à 2014

Objectif du graphique :

Ici, nous allons répondre aux questions suivantes : **À quel point les jeux vidéos Pokémon sont-ils populaires ? Les jeux récents sont-ils plus populaires que les anciens ?** Les jeux vidéos Pokémon existent depuis très longtemps et de nouveaux jeux ne cessent de sortir. À l'aide de ce graphique, le but est d'observer si les jeux vidéos Pokémon occupent une place importante sur le marché des jeux vidéos et de

voir comment évolue cette place au cours du temps. Nous avons créé un barchart permettant de comparer chaque jeu vidéo Pokémon avec le jeu le plus vendu de la même année.

Présentation :

Pour réaliser ce graphique, nous avons utilisé les données du dataset “vgsales.csv”. Nous avons conservé uniquement les jeux Pokémon qui ont des ventes globales supérieures à 6 millions afin de conserver uniquement les jeux principaux de Pokémon et éliminer tous les spin-off. Nous avons ensuite comparé à l’aide d’un barchart le nombre de ventes des jeux Pokémon avec le jeu vidéo le plus vendu de la même année. On peut voir également la position du jeu Pokémon parmi les ventes globales de l’année.

Code et Visualisation :

```
# Charger les bibliothèques
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats   1.0.0      v readr     2.1.5
## v lubridate 1.9.4      v stringr  1.5.1
## v purrr     1.0.2      v tibble   3.2.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Charger le dataset
df <- read_csv("data/vgsales.csv")
```

```
## Rows: 16598 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (5): Name, Platform, Year, Genre, Publisher
## dbl (6): Rank, NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales
##
## 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. Filtrer les jeux Pokémon avec plus de 6 millions de ventes
pokemon_games <- df %>%
  filter(str_detect(Name, "(?i)pokemon") & Global_Sales > 6) %>%
  drop_na(Year, Global_Sales) %>%
  filter(is.finite(Global_Sales))
```

```
# 2. Identifier le jeu le plus vendu par année
top_games_by_year <- df %>%
  drop_na(Year, Global_Sales) %>%
  filter(is.finite(Global_Sales)) %>%
  group_by(Year) %>%
  slice_max(order_by = Global_Sales, n = 1) %>%
  ungroup()
```

```
# 3. Calculer le rang de chaque jeu Pokémon par année
df_with_rank <- df %>%
  drop_na(Year, Global_Sales) %>%
  filter(is.finite(Global_Sales)) %>%
```

```

group_by(Year) %>%
mutate(Rank = rank(-Global_Sales)) %>%
ungroup()

# 4. Associer chaque jeu Pokémon à son jeu top de l'année et ajouter le rang
comparaison_df <- pokemon_games %>%
  left_join(top_games_by_year, by = "Year", suffix = c("_pokemon", "_top")) %>%
  left_join(df_with_rank %>% filter(str_detect(Name, "(?i)pokemon")), by = c("Year", "Name_pokemon" = "Name_top"))
select(Year,
       Pokemon_Game = Name_pokemon,
       Pokemon_Sales = Global_Sales_pokemon,
       Top_Game = Name_top,
       Top_Game_Sales = Global_Sales_top,
       Pokemon_Rank = Rank)

# 5. Préparer les données pour un bar chart
plot_data <- comparaison_df %>%
  drop_na(Year, Pokemon_Sales, Top_Game_Sales) %>%
  filter(is.finite(Pokemon_Sales), is.finite(Top_Game_Sales)) %>%
  pivot_longer(cols = c(Pokemon_Sales, Top_Game_Sales),
               names_to = "Type",
               values_to = "Sales") %>%
  mutate(
    Game_Name = ifelse(Type == "Pokemon_Sales", Pokemon_Game, Top_Game),
    Color = ifelse(Type == "Pokemon_Sales", "Pokemon_Game", "Top_Game"),
    Label = gsub("(?i)pokemon", "", Game_Name, ignore.case = TRUE),
    Rank_Label = ifelse(Type == "Pokemon_Sales", paste(Pokemon_Rank), NA)
  )

# 6. Faire un bar chart avec un thème sombre personnalisé
ggplot(plot_data, aes(x = factor(Year), y = Sales, fill = Color)) +
  geom_bar(stat = "identity", position = position_dodge2(width = 0.9, preserve = "single"),
           width = 0.7) +
  labs(title = "Comparaison des ventes de jeux vidéos: Pokémon vs Top Vente par Année",
       x = "Année",
       y = "Ventes Globales (Millions)") +
  scale_fill_manual(values = c("Pokemon_Game" = "#1f77b4", "Top_Game" = "#ff7f0e"),
                    labels = c("Jeu Pokémon", "Top Jeu Année"),
                    name = "Type de Jeu") +
  scale_y_continuous(limits = c(0, 100)) +
  theme_minimal(base_family = "sans") + # Changement de thème
  theme(
    axis.text.x = element_text(angle = 70, hjust = 1, size = 10, face = "bold", color = "black"),
    axis.text.y = element_text(size = 10, face = "bold", color = "black"),
    axis.title.x = element_text(face = "bold", size = 12, color = "black"),
    axis.title.y = element_text(face = "bold", size = 12, color = "black"),
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5, color = "black"),
    legend.title = element_text(face = "bold", color = "black"),
    legend.text = element_text(face = "bold", color = "black"),
    axis.ticks = element_line(color = "black", size = 0.7),
    axis.line = element_line(color = "black", size = 0.7),
    plot.margin = margin(t = 30, r = 10, b = 10, l = 10)
  ) +

```

```

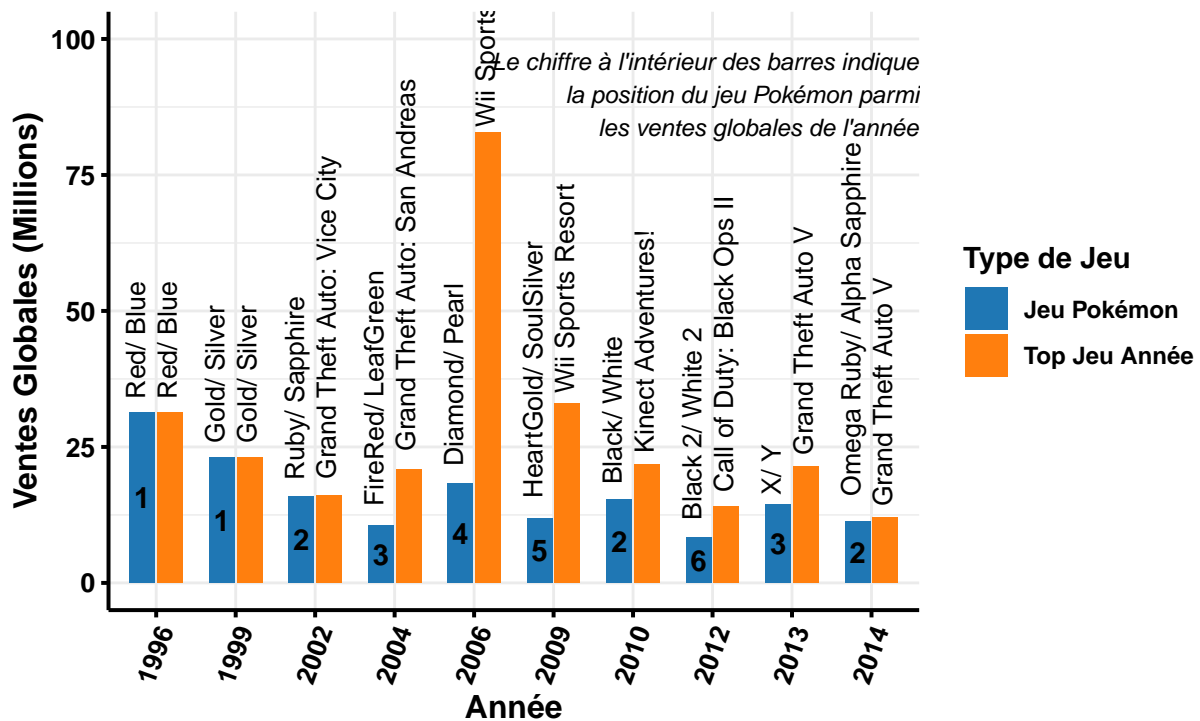
geom_text(aes(label = Label),
  position = position_dodge(width = 0.8),
  vjust = 0.25,
  hjust = -0.05,
  size = 3.5,
  colour = "black",
  angle = 90) +

geom_text(aes(label = Rank_Label, y = Sales / 2),
  position = position_dodge2(width = 0.9, preserve = "single"),
  vjust = 0.5,
  hjust = 0.25,
  size = 4,
  colour = "black",
  fontface = "bold") +

annotate("text",
  x = Inf, y = Inf,
  label = "Le chiffre à l'intérieur des barres indique\nla position du jeu Pokémon parmi\nles ventes globales de l'année",
  hjust = 1.0, vjust = 1.5,
  size = 3,
  color = "black",
  fontface = "italic")

```

Comparaison des ventes de jeux vidéo: Pokémon vs Top Vente par Année



Analyse :

On peut observer que les jeux Pokémon récents se vendent globalement moins bien que les anciens. On peut remarquer cela car les 2 jeux Pokémon les plus vendus de l'histoire sont sortis en **1996 et 1999** avec respectivement **plus de 30 et 20 millions** de ventes, tandis que les plus récents peinent à atteindre les **20 millions** de ventes.

Cependant à l'aide des rangs des jeux Pokémon parmi les ventes globales de chaque année, on peut voir qu'à chaque sortie d'un nouveau jeu Pokémon, ce dernier figure presque toujours parmi le **TOP 5** des ventes de l'année. Récemment, il est rare de voir des jeux Pokémon en première position comme on avait pu le voir en **1996 et 1999**. Mais on peut conclure que les jeux Pokémon occupent toujours une place très importante sur le marché du jeu vidéo. On peut expliquer cette baisse des ventes par le fait qu'il y'a de plus en plus de concurrence et un manque de nouveauté dans le contenu.

3. Popularité des type de Pokémon

Objectif du graphique :

Ici, nous allons répondre à la question suivante : **Quels sont les types de Pokémon préférés de la communauté ?**

Présentation :

Pour réaliser ce graphique, nous avons combiné deux jeux de données :

- **The complete Pokemon dataset** : Statistiques complètes des Pokémon des 8 premières générations (notamment le type)
- **Sondage de popularité** : Sondage réalisé en 2019 sur un forum de la communauté Pokémon recueillant le Pokémon préféré de 52000 votants

Nous considérons ici seulement le premier type de chaque Pokémon (certains Pokémon ont un deuxième type, mais on peut considérer qu'il est moins caractérisant que le premier).

On affiche ici, pour chacun des types, le pourcentage de vote reçu par l'ensemble des Pokémon de ce type. Pour mettre en perspective ces résultats, on affiche également la distribution des Pokémon par type. Les Pokémon dessinés sont les plus populaires pour chacun des types.

Code et Visualisation :

```
library(tidyverse)
library(ggimage)

# Load data
popularity <- tibble(read.table("data/popularity.txt", header = TRUE, sep = ","))
pok_data <- tibble(read.csv2("data/Pokemon_data.csv", header = TRUE, sep = ','))

# Join datasets by Pokemon name
merged <- popularity %>% left_join(pok_data %>% select(name, type1), by = c("Pokemon" = "name"))

# Create aggregated data by Pokemon type
type_summary <- merged %>%
  filter(!is.na(type1)) %>%
  group_by(type1) %>%
  summarise(TotalVotes = sum(Number.of.votes), Count = n()) %>%
  mutate(AvgVotes = TotalVotes / Count) %>%
  arrange(desc(AvgVotes)) %>%
  left_join(
    # Select the most popular Pokemon for each type
```

```

merged %>%
  filter(!is.na(type1)) %>%
  group_by(type1) %>%
  slice_max(Number.of.votes, n = 1) %>%
  select(type1, Pokemon, Number.of.votes),
  by = "type1"
) %>%
mutate(img_path = paste0("data/images/", tolower(Pokemon), ".png"))

# Define colors of each type for graphical purpose
type_colors <-c(
  'Normal' = '#A8A77A',
  'Fire' = '#EE8130',
  'Water' = '#6390F0',
  'Electric' = '#F7D02C',
  'Grass' = '#7AC74C',
  'Ice' = '#96D9D6',
  'Fighting' = '#C22E28',
  'Poison' = '#A33EA1',
  'Ground' = '#E2BF65',
  'Flying' = '#A98FF3',
  'Psychic' = '#F95587',
  'Bug' = '#A6B91A',
  'Rock' = '#B6A136',
  'Ghost' = '#735797',
  'Dragon' = '#6F35FC',
  'Dark' = '#705746',
  'Steel' = '#B7B7CE',
  'Fairy' = '#D685AD'
)

# Compute the average number of votes
avg_votes <- mean(type_summary$TotalVotes) / sum(type_summary$TotalVotes) * 100

# Draw graph
type_summary %>%
  ggplot(aes(x = reorder(type1, -TotalVotes),
    y = TotalVotes / sum(TotalVotes) * 100,
    fill=type1)) +
  # Colored bars
  geom_col() +
  geom_hline(yintercept = avg_votes, linetype = "dashed", color = "red", linewidth = 1, alpha=.5) +
  # Black bars
  geom_col(aes(y = Count / sum(Count) * 100, fill = "Nombre de Pokémons"), alpha = 0.3, width = 0.6) +
  # Fake invisible layer to create "Nombre de Votes" legend
  geom_col(aes(y = 0, fill = "Nombre de Votes"), show.legend = TRUE) +
  scale_fill_manual(
    name = NULL,
    values = c(type_colors, "Nombre de Pokémons" = "black", "Nombre de Votes" = "#6390F0"),
    breaks = c("Nombre de Votes", "Nombre de Pokémons")
  )

```



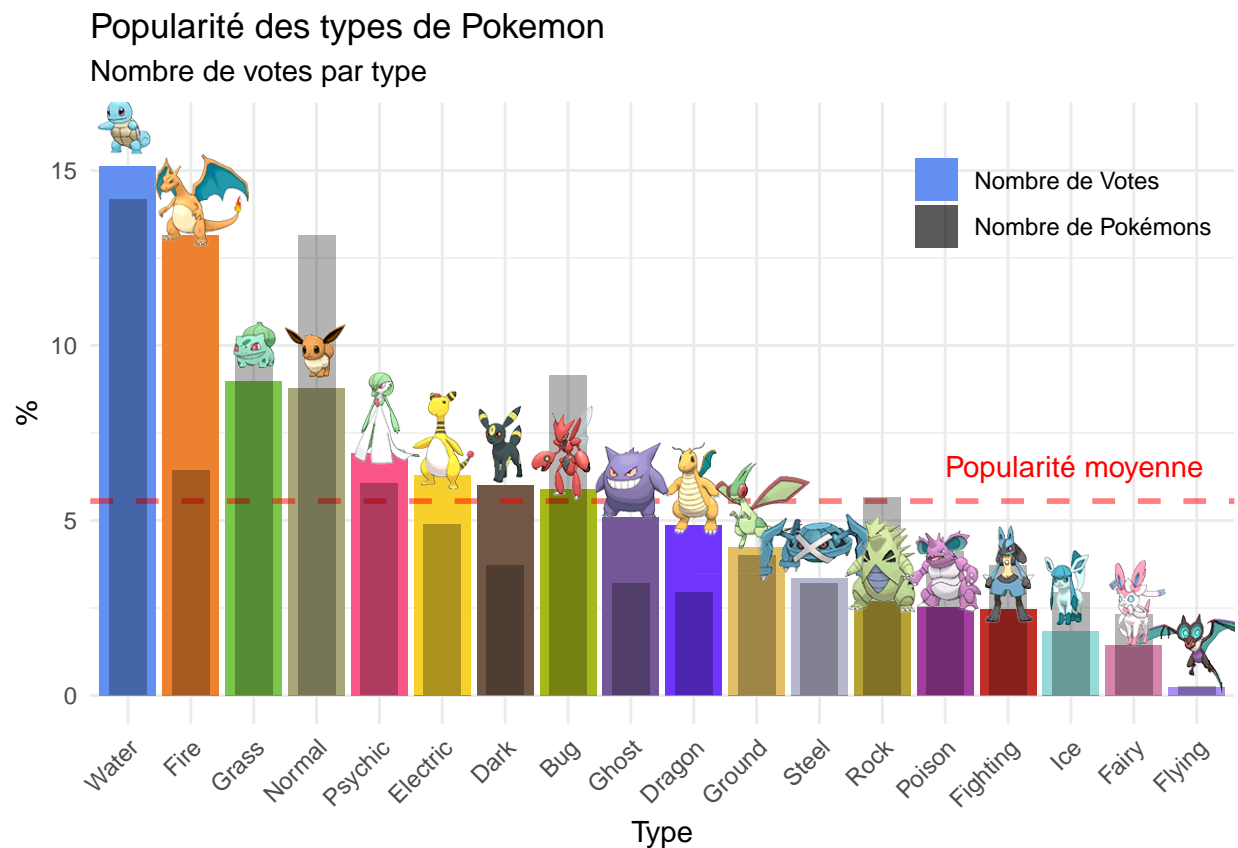
```

) +
guides(
  fill = guide_legend(override.aes = list(alpha = c(1, 0.3)))
) +

annotate("text", x = 14, y = avg_votes + 1, label = "Popularité moyenne", color = "red", size = 4, hjust = 0)
labs(title = "Popularité des types de Pokemon", subtitle = "Nombre de votes par type", x = "Type", y = "%")
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = c(0.85, 0.85)) +

geom_image(aes(image=img_path, y = TotalVotes / sum(TotalVotes) * 100 + 1), size=.2)

```



Analyse :

Tout d'abord, il est intéressant d'observer que les 3 types ayant reçu le plus grand nombre de vote sont les 3 types de base (**Eau**, **Feu** et **Plante**). Presque 40% des votes se sont tournés vers ces types, notamment grâce aux Pokémon **starters** (Pokémons de départ) qui sont très populaires. Bien que le type **Eau** arrive en tête du trio, les Pokémon **Eau** étant assez largement les plus représentés, ils sont, en proportion, bien moins appréciés que les Pokémon **Feu**.

De même, le type **Normal** arrive en quatrième position, mais cela semble surtout s'expliquer par le nombre très important de Pokémon de ce type. Les types **Insecte** et **Roche** ont également reçu un nombre de vote très faible en comparaison de leur nombre.

A l'inverse, les types **Ténèbres**, **Spectre** et **Dragon** ont été beaucoup cités malgré le nombre de Pokémon assez modeste.

Le grand gagnant de ce sondage apparaît assez clairement : il s'agit du type **Feu**

4. Comparaison des base stats des Pokémons légendaires et des Pokémons normaux

Objectif du graphique : Ici, nous allons répondre à la question suivante : **Les Pokémons légendaires sont-ils les plus forts ?**

Parmi tous les Pokémon existants, certains Pokémon sont considérés comme étant spéciaux : les Pokémon légendaires, iconiques à chaque jeu, accompagnés des Pokémon fabuleux. Introduits dans les jeux Pokémon X et Y, les méga-évolutions constituent également une catégorie à part.

Présentation : Afin de savoir si ces Pokémon sont réellement plus puissants que les autres, nous allons observer la répartition des totaux des statistiques de base (abrégé “base stats”) pour tous les Pokémon, en différenciant les légendaires des autres.

Nous allons utiliser le dataset “The Complete Pokemon Dataset”.

Les conditions des Pokémon légendaires, mythiques (“mythical” en anglais) et méga-évolutions sont dispersées dans 3 colonnes différentes que nous allons fusionner en une seule colonne “group”. Enfin, nous affichons l’histogramme du total de base stats, en différenciant les Pokémon spéciaux par la couleur.

Code et Visualisation :

```
# importations des bibliothèques tidyverse
library(ggplot2)
library(dplyr)
library(readr)
```

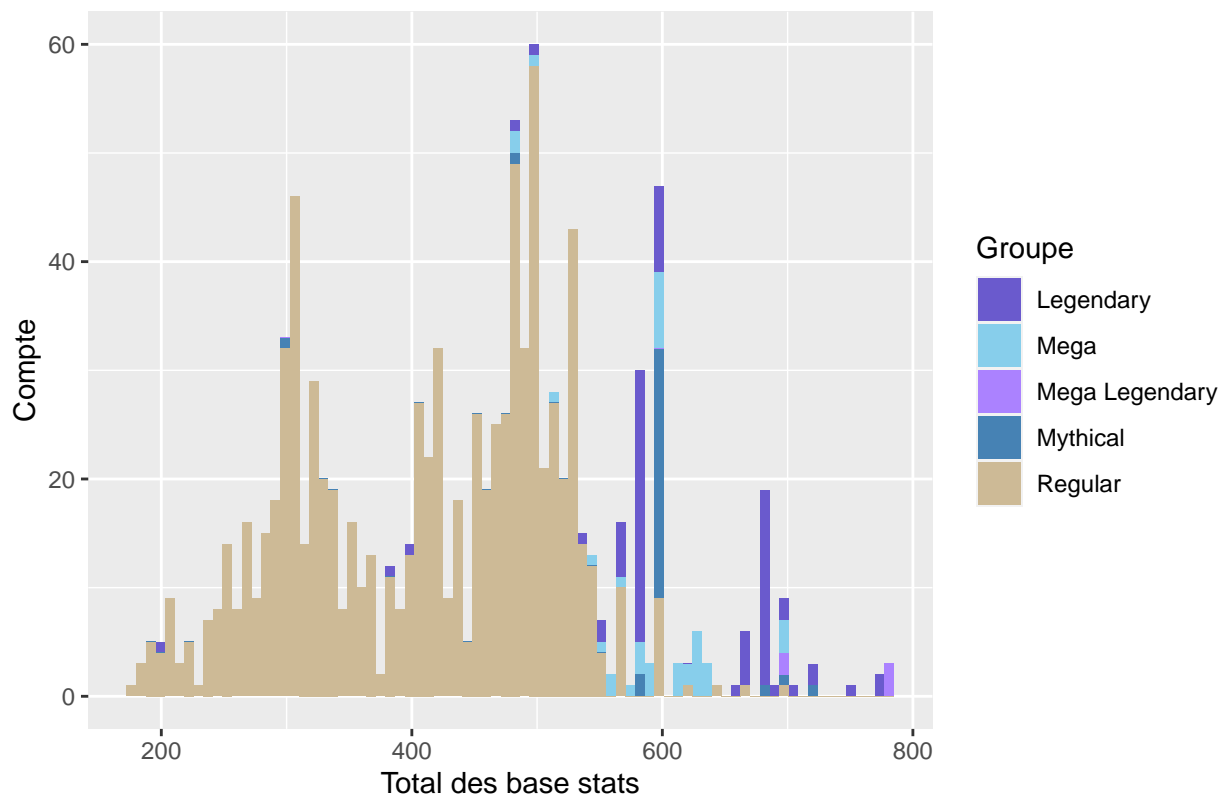
```
# ouverture du dataset
df = read_csv("data/Pokemon_data.csv")
```

```
## Rows: 1033 Columns: 40
## -- Column specification -----
## Delimiter: ","
## chr (6): name, classification, abilities, type1, type2, base_egg_steps
## dbl (34): pokedex_number, generation, height_m, weight_kg, base_total, hp, a...
##
## 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.
```

```
# création d'une nouvelle colonne "group", possédant comme valeur "legendary" si le pokémon est légendaire
df <- df %>% mutate(group = case_when(
  (is_legendary == TRUE & is_mega == TRUE) ~ "Mega Legendary",
  is_legendary == TRUE ~ "Legendary",
  is_mythical == TRUE ~ "Mythical",
  is_mega == TRUE ~ "Mega",
  TRUE ~ "Regular"
))
```

```
# affichage de l'histogramme
ggplot(df, aes(base_total)) +
  geom_histogram(bins=80, aes(fill = group)) +
  scale_fill_manual(values = alpha(c("slateblue", "skyblue", "mediumpurple1", "steelblue", "wheat3"))) +
  labs(
    title = "Distribution des Pokémons selon le total de leurs base stats et leur groupe",
    x = "Total des base stats",
    y = "Compte",
    fill = "Groupe")
```

Distribution des Pokémons selon le total de leurs base stats et leur groupe



Analyse :

On peut observer que les Pokémons légendaires, fabuleux et méga-évolués possèdent bien un total de statistiques plus élevé que les Pokémons normaux. Ils constituent la quasi majorité de tous les Pokémons ayant 580 ou plus de base stats. On observe également que la plupart des Pokémons spéciaux ont le même montant de base stats : 600, 580 ou 680. De plus, les Pokémons les plus “puissants”, possédant le montant le plus élevé de base stats font partie d’un groupe rare : ils sont à la fois des méga-évolutions et des Pokémons légendaires (Mega Mewtwo X, Mega Mewtwo Y et méga Rayquaza). Il existe tout de même quelques cas de Pokémons spéciaux avec des statistiques moyennes voire faibles. On peut penser à Cosmog (200 de base stats), qui est un cas particulier parmi les Pokémons légendaires car il peut évoluer : sa dernière évolution, Solgaleo, possède lui 680 de base stats.