

# Analyse approfondie des données Pokémon

Projet R

2025-01-12

## Introduction

Ce rapport propose une analyse approfondie des caractéristiques des Pokémon, comprenant des statistiques descriptives, des analyses croisées et des visualisations graphiques détaillées.

## Description des données

### Aperçu général

```
# Aperçu des données
glimpse(pokemon_data)
```

```
## Rows: 800
## Columns: 14
## $ `#`      <dbl> 1, 2, 3, 3, 4, 5, 6, 6, 6, 7, 8, 9, 9, 10, 11, 12, 13, 14, ~
## $ Name     <chr> "Bulbasaur", "Ivysaur", "Venusaur", "VenusaurMega Venusaur"~
## $ `Type 1` <chr> "Grass", "Grass", "Grass", "Grass", "Fire", "Fire", "Fire",~
## $ `Type 2` <chr> "Poison", "Poison", "Poison", "Poison", NA, NA, "Flying", "~
## $ Total    <dbl> 318, 405, 525, 625, 309, 405, 534, 634, 634, 314, 405, 530,~
## $ HP       <dbl> 45, 60, 80, 80, 39, 58, 78, 78, 78, 44, 59, 79, 79, 45, 50,~
## $ Attack   <dbl> 49, 62, 82, 100, 52, 64, 84, 130, 104, 48, 63, 83, 103, 30,~
## $ Defense  <dbl> 49, 63, 83, 123, 43, 58, 78, 111, 78, 65, 80, 100, 120, 35,~
## $ `Sp. Atk` <dbl> 65, 80, 100, 122, 60, 80, 109, 130, 159, 50, 65, 85, 135, 2~
## $ `Sp. Def` <dbl> 65, 80, 100, 120, 50, 65, 85, 85, 115, 64, 80, 105, 115, 20~
## $ Speed    <dbl> 45, 60, 80, 80, 65, 80, 100, 100, 100, 43, 58, 78, 78, 45, ~
## $ Generation <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ Legendary <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FAL~
## $ TotalStats <dbl> 318, 405, 525, 625, 309, 405, 534, 634, 634, 314, 405, 530,~
```

```
# Résumé statistique
summary(pokemon_data)
```

```
##           #           Name           Type 1           Type 2
## Min.      : 1.0      Length:800      Length:800      Length:800
## 1st Qu.:184.8      Class :character  Class :character  Class :character
## Median   :364.5      Mode  :character  Mode  :character  Mode  :character
## Mean     :362.8
## 3rd Qu.:539.2
## Max.     :721.0
##           Total           HP           Attack           Defense
## Min.      :180.0      Min.      : 1.00      Min.      : 5      Min.      : 5.00
## 1st Qu.:330.0      1st Qu.: 50.00      1st Qu.: 55      1st Qu.: 50.00
```

```
## Median :450.0 Median : 65.00 Median : 75 Median : 70.00
## Mean :435.1 Mean : 69.26 Mean : 79 Mean : 73.84
## 3rd Qu.:515.0 3rd Qu.: 80.00 3rd Qu.:100 3rd Qu.: 90.00
## Max. :780.0 Max. :255.00 Max. :190 Max. :230.00
## Sp. Atk Sp. Def Speed Generation
## Min. : 10.00 Min. : 20.0 Min. : 5.00 Min. :1.000
## 1st Qu.: 49.75 1st Qu.: 50.0 1st Qu.: 45.00 1st Qu.:2.000
## Median : 65.00 Median : 70.0 Median : 65.00 Median :3.000
## Mean : 72.82 Mean : 71.9 Mean : 68.28 Mean :3.324
## 3rd Qu.: 95.00 3rd Qu.: 90.0 3rd Qu.: 90.00 3rd Qu.:5.000
## Max. :194.00 Max. :230.0 Max. :180.00 Max. :6.000
## Legendary TotalStats
## Mode :logical Min. :180.0
## FALSE:735 1st Qu.:330.0
## TRUE :65 Median :450.0
## Mean :435.1
## 3rd Qu.:515.0
## Max. :780.0
```

```
# Valeurs manquantes
```

```
missing_values
```

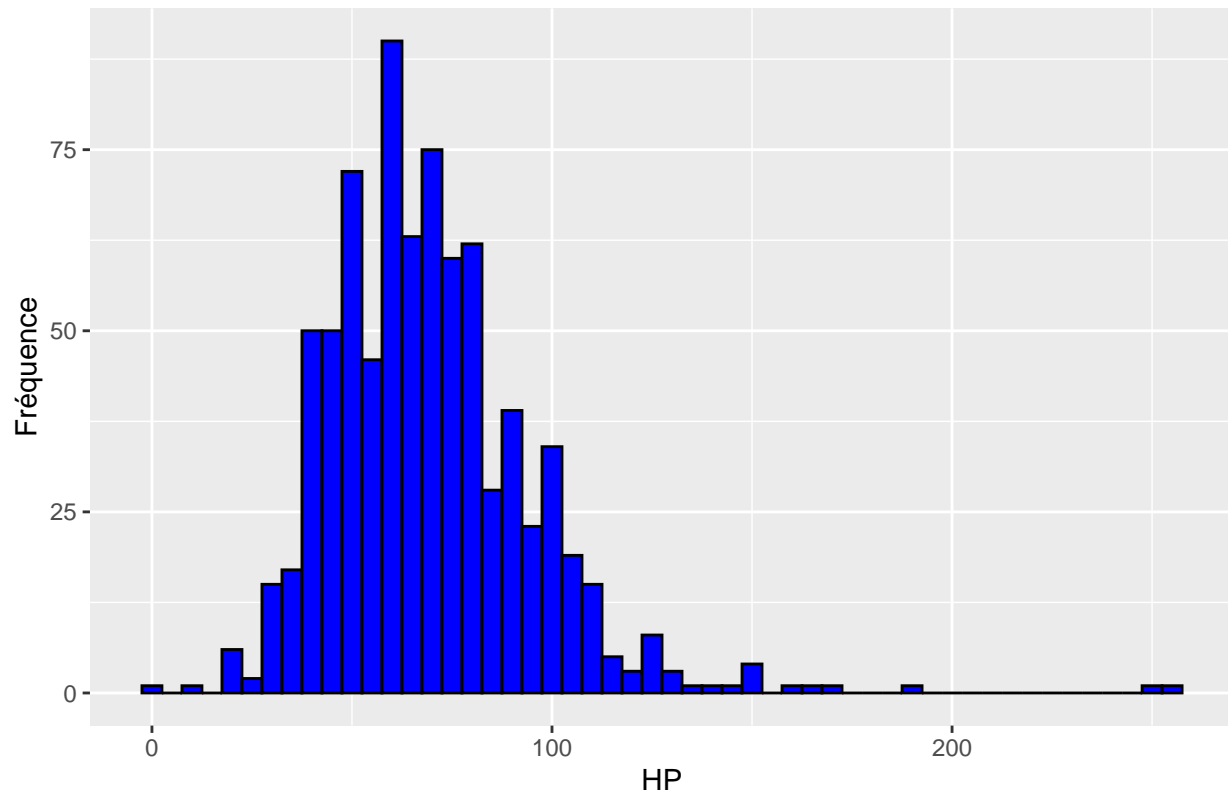
```
## # Name Type 1 Type 2 Total HP Attack
## 0 0 0 386 0 0 0
## Defense Sp. Atk Sp. Def Speed Generation Legendary
## 0 0 0 0 0 0
```

## Distribution des variables principales

```
# Distribution des points de vie (HP)
```

```
ggplot(pokemon_data, aes(x = HP)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  labs(title = "Distribution des points de vie (HP)", x = "HP", y = "Fréquence")
```

Distribution des points de vie (HP)

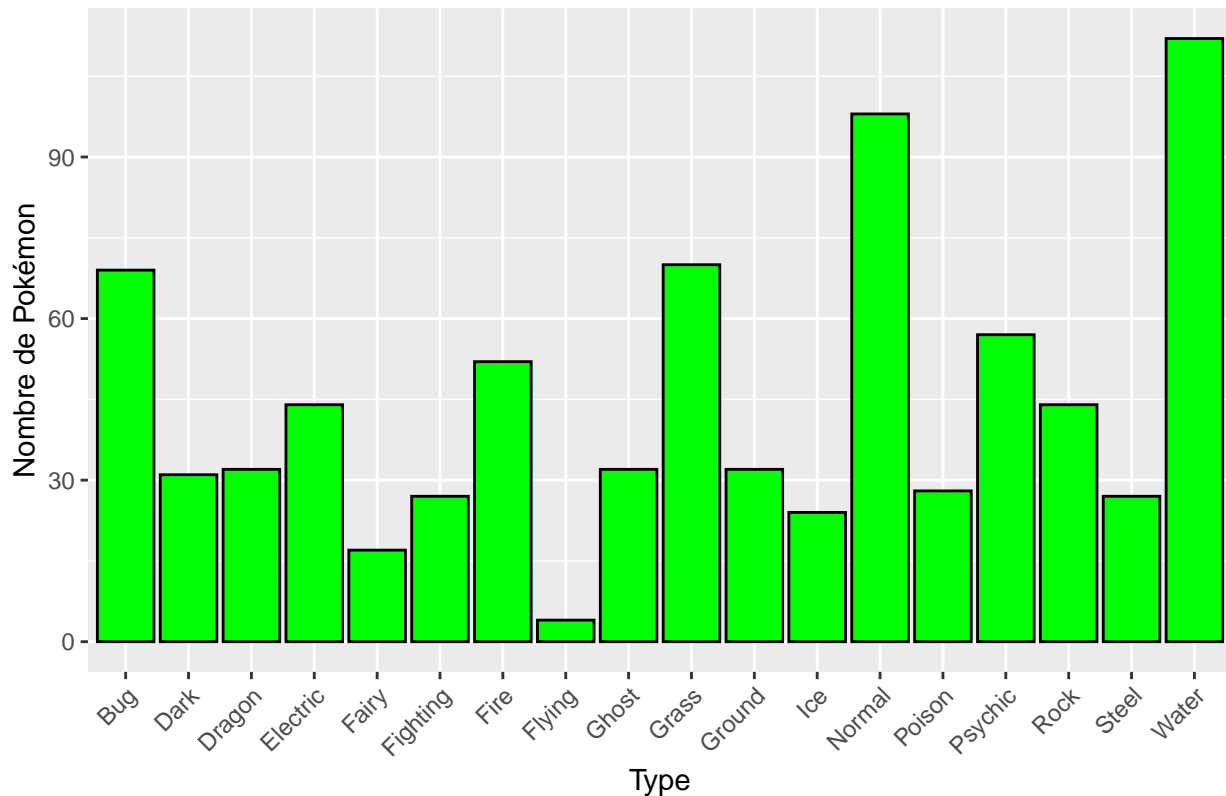


*# Analyse : Cette distribution montre que la majorité des Pokémon ont des points de vie dans une plage de 50 à 100.*

*# Répartition des types principaux*

```
ggplot(pokemon_data, aes(x = `Type 1`)) +  
  geom_bar(fill = "green", color = "black") +  
  labs(title = "Répartition des types principaux", x = "Type", y = "Nombre de Pokémon") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

## Répartition des types principaux



*# Analyse : On observe une répartition inégale des types principaux, avec certains types comme Water et*

## Analyse descriptive

### Comparaison des statistiques par Type principal

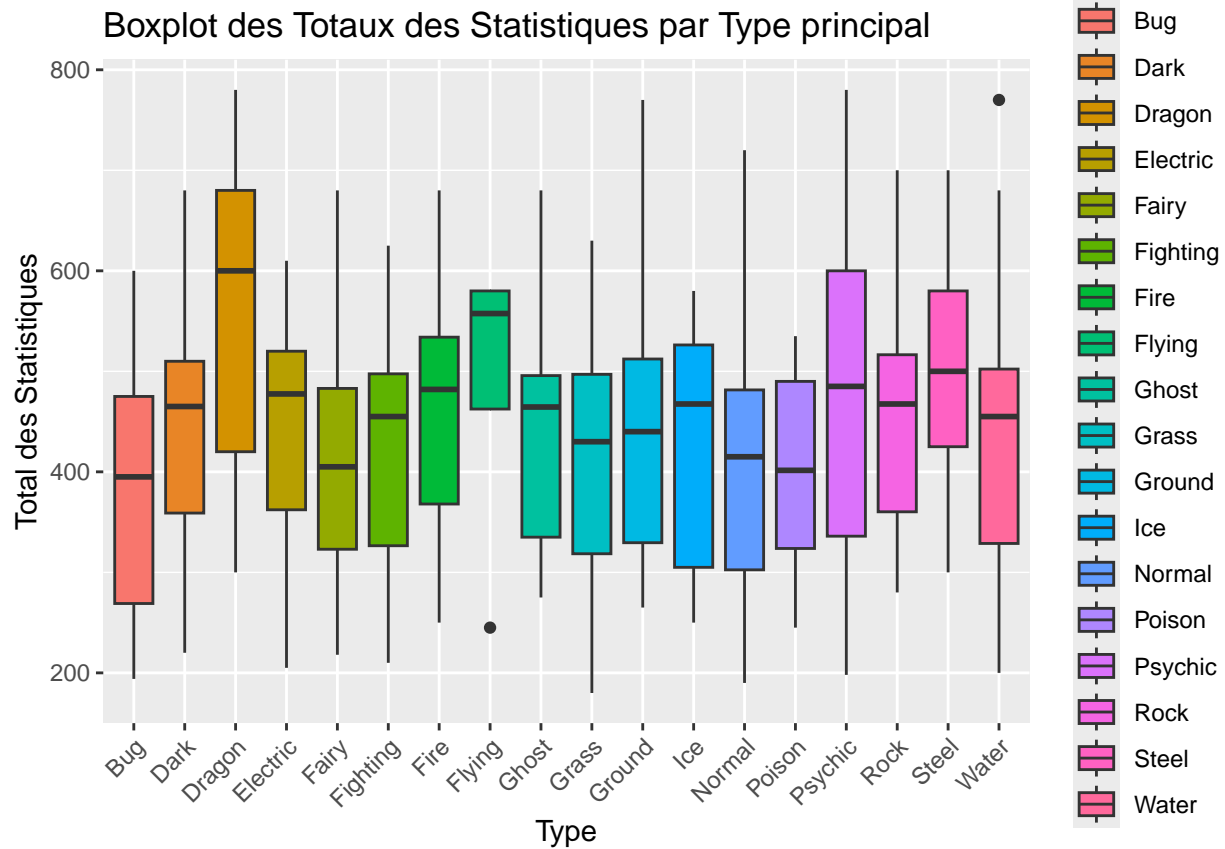
*# Résumés statistiques par Type principal*

```
print(type_summary)
```

```
## # A tibble: 1 x 5
##   `Type 1` AvgAttack AvgDefense AvgSpeed TotalCount
##   <chr>      <dbl>      <dbl>      <dbl>      <int>
## 1 Type 1      79.0      73.8      68.3      800
```

*# Boxplot des Totaux des Statistiques par Type principal*

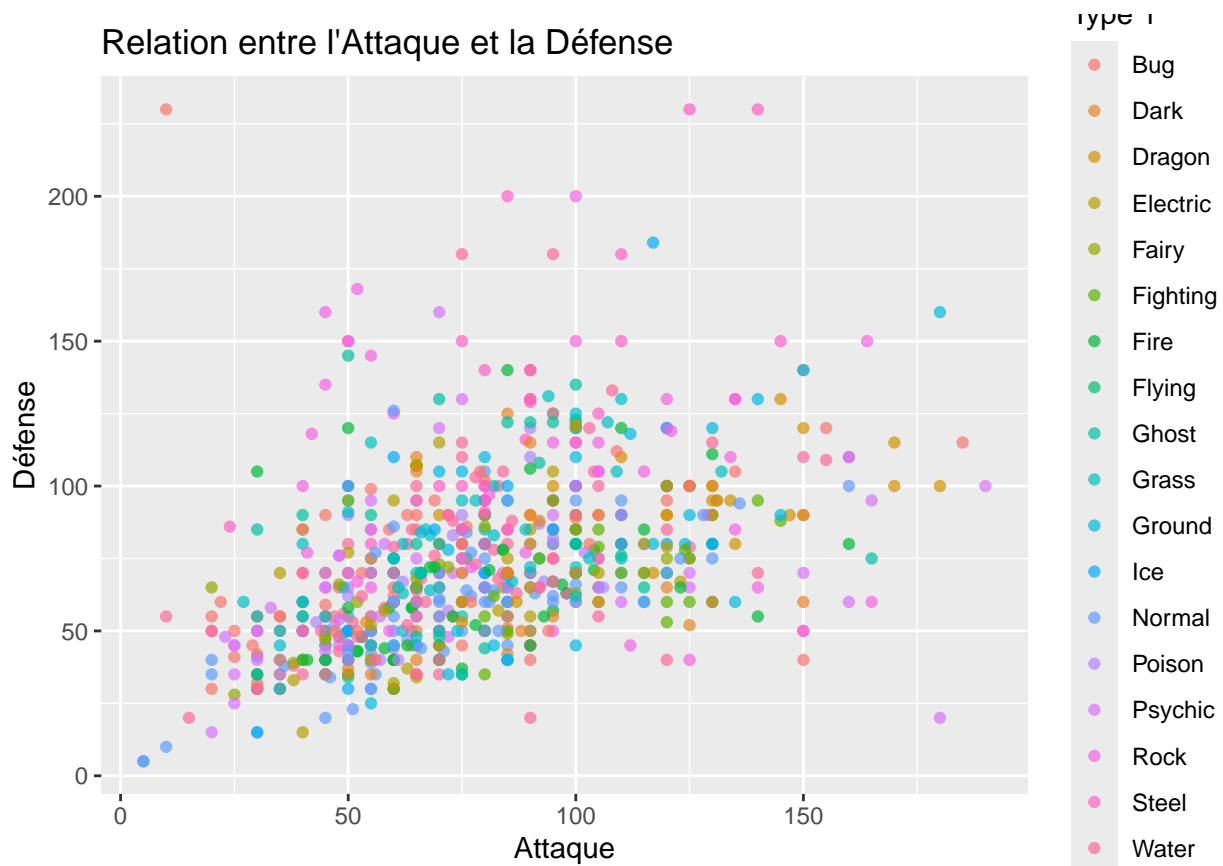
```
ggplot(pokemon_data, aes(x = `Type 1`, y = TotalStats, fill = `Type 1`)) +
  geom_boxplot() +
  labs(title = "Boxplot des Totaux des Statistiques par Type principal", x = "Type", y = "Total des Sta
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



*# Analyse : Les types Dragon et Psychic semblent avoir des Totaux des Statistiques plus élevés en moyenne*

### Relation entre les variables

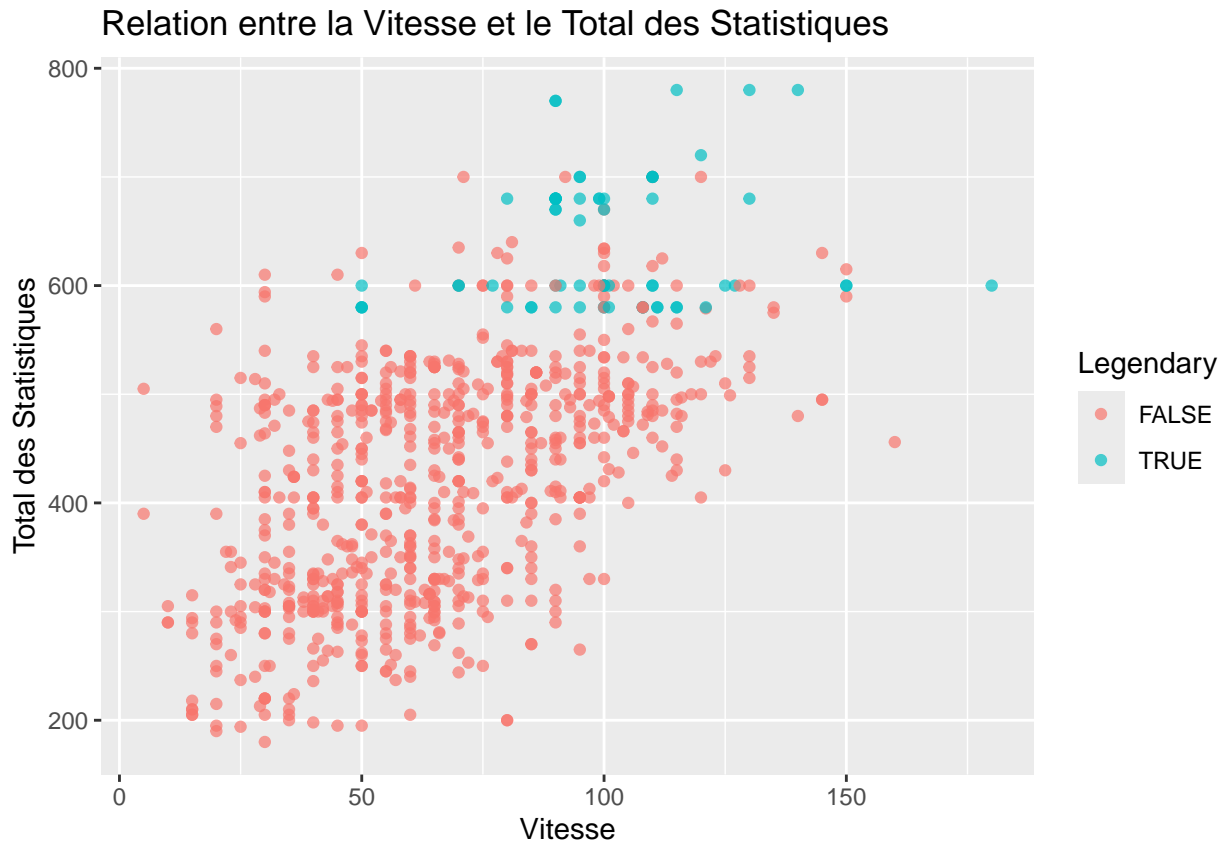
```
# Relation entre Attack et Defense
ggplot(pokemon_data, aes(x = Attack, y = Defense, color = `Type 1`)) +
  geom_point(alpha = 0.7) +
  labs(title = "Relation entre l'Attaque et la Défense", x = "Attaque", y = "Défense")
```



*# Analyse : Il existe une corrélation modérée entre l'attaque et la défense. Certains types comme Rock*

*# Relation entre Speed et TotalStats*

```
ggplot(pokemon_data, aes(x = Speed, y = TotalStats, color = Legendary)) +
  geom_point(alpha = 0.7) +
  labs(title = "Relation entre la Vitesse et le Total des Statistiques", x = "Vitesse", y = "Total des Statistiques")
```



*# Analyse : Les Pokémon légendaires ont généralement des Totaux des Statistiques élevés indépendamment*

## Analyse par sous-groupes

### Pokémon légendaires vs non légendaires

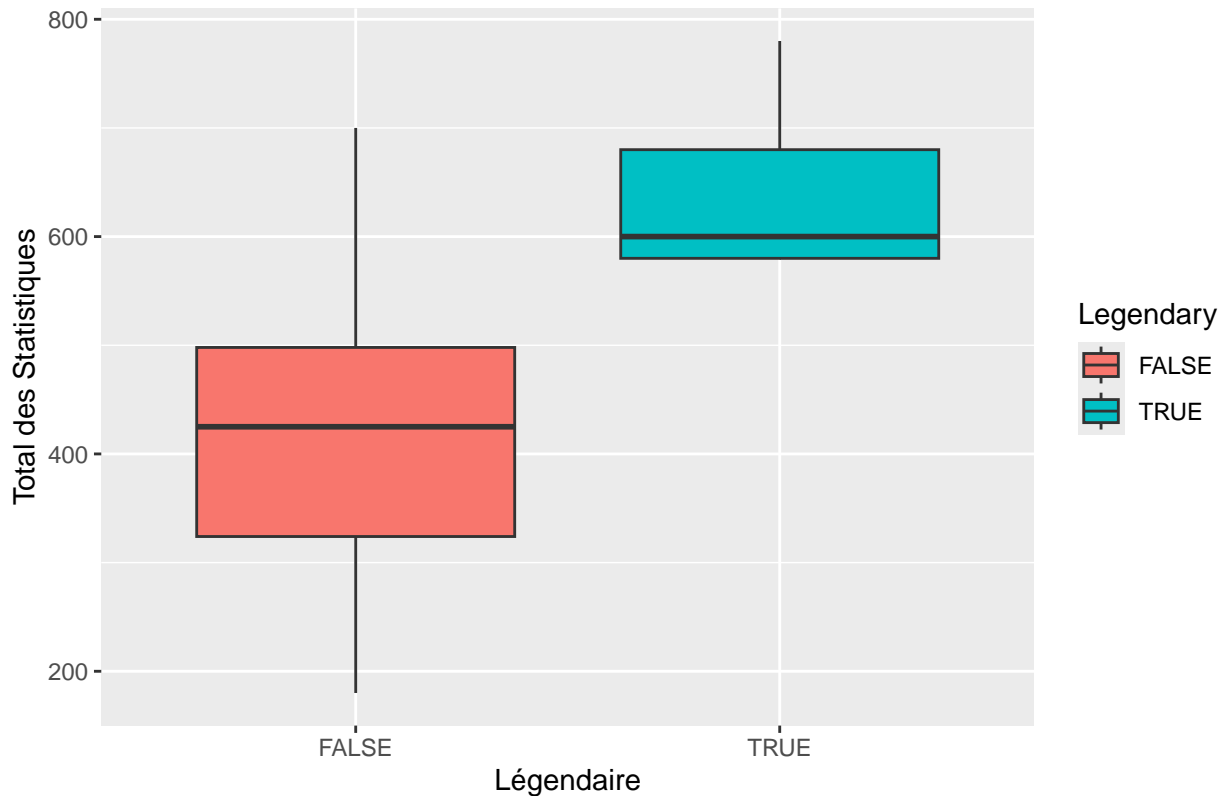
```
legendary_summary <- pokemon_data %>%
  group_by(Legendary) %>%
  summarise(
    AvgTotalStats = mean(TotalStats, na.rm = TRUE),
    AvgAttack = mean(Attack, na.rm = TRUE),
    AvgDefense = mean(Defense, na.rm = TRUE),
    Count = n()
  )
print(legendary_summary)
```

```
## # A tibble: 2 x 5
##   Legendary AvgTotalStats AvgAttack AvgDefense Count
##   <lgl>      <dbl>      <dbl>      <dbl> <int>
## 1 FALSE      417.        75.7        71.6   735
## 2 TRUE       637.       117.        99.7    65
```

*# Comparaison visuelle*

```
ggplot(pokemon_data, aes(x = Legendary, y = TotalStats, fill = Legendary)) +
  geom_boxplot() +
  labs(title = "Comparaison des Totaux des Statistiques: Légendaires vs Non-légendaires", x = "Légendaires")
```

## Comparaison des Totaux des Statistiques: Légendaires vs Non-légendaire



*# Analyse : Les Pokémon légendaires ont des Totaux des Statistiques nettement plus élevés que les non-l*

## Analyse des Pokémon par génération

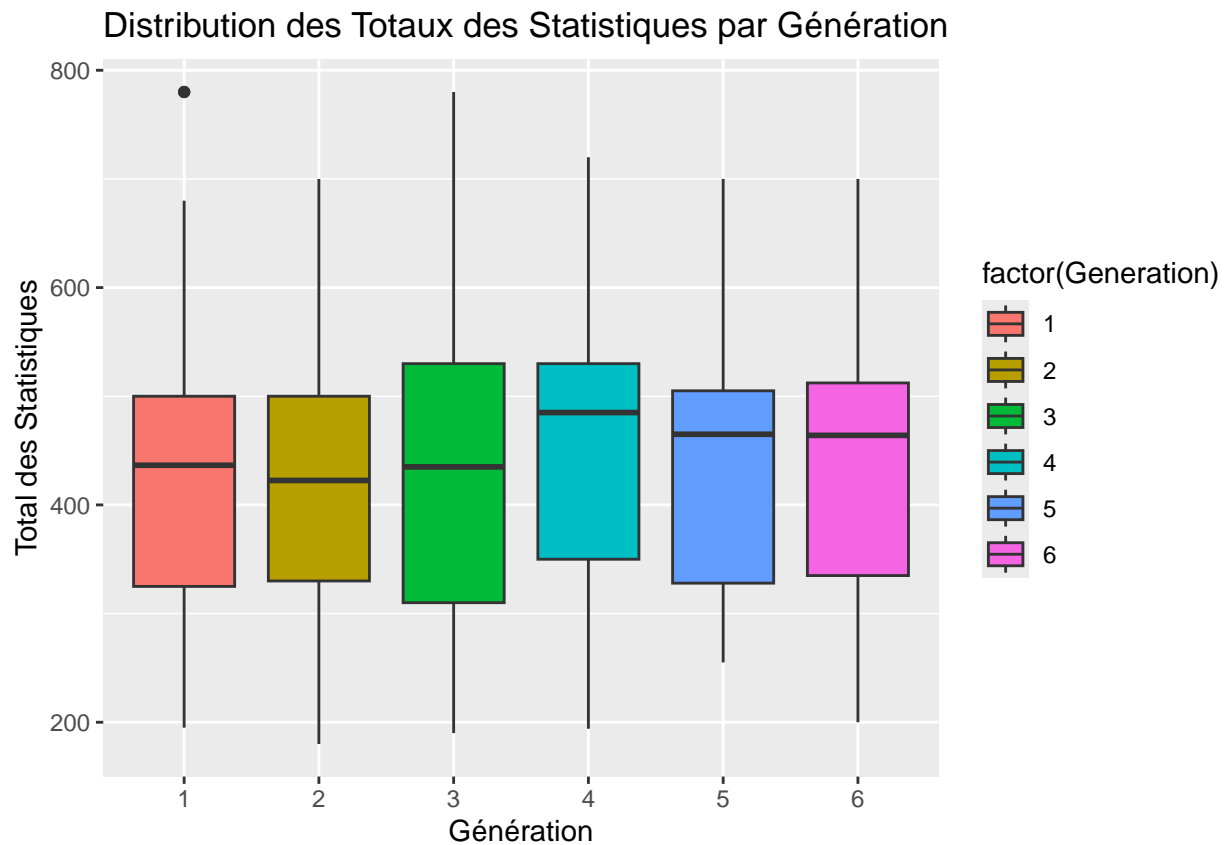
```
generation_summary <- pokemon_data %>%
  group_by(Generation) %>%
  summarise(
    AvgTotalStats = mean(TotalStats, na.rm = TRUE),
    AvgHP = mean(HP, na.rm = TRUE),
    Count = n()
  )
print(generation_summary)
```

```
## # A tibble: 6 x 4
##   Generation AvgTotalStats AvgHP Count
##   <dbl>      <dbl> <dbl> <int>
## 1         1         427.   65.8   166
## 2         2         418.   71.2   106
## 3         3         436.   66.5   160
## 4         4         459.   73.1   121
## 5         5         435.   71.8   165
## 6         6         436.   68.3    82
```

*# Visualisation par génération*

```
ggplot(pokemon_data, aes(x = factor(Generation), y = TotalStats, fill = factor(Generation))) +
  geom_boxplot() +
  labs(title = "Distribution des Totaux des Statistiques par Génération", x = "Génération", y = "Total Stats")
```





*# Analyse : Les générations plus récentes semblent avoir des Totaux des Statistiques légèrement supérieures.*

## Conclusion

Cette analyse approfondie met en évidence les relations complexes entre les caractéristiques des Pokémon. Les graphiques et résumés statistiques illustrent des tendances claires entre les types, les générations et le statut légendaire des Pokémon.