

TP N°3 Bendris Nael / Benkheira Lilya Groupe E

Problématique

Où a commencé la mode des jeux vidéo ? Quels sont les jeux qui ont le plus marqué ce monde ? Y a-t-il des jeux propres à des zones géographiques ?

Introduction

Dans l'ère moderne du divertissement, les jeux vidéo ont pris une place prépondérante. Leur popularité est telle qu'ils sont désormais une industrie majeure à l'échelle mondiale. Dans cette analyse, nous nous intéressons à l'origine de cette tendance, aux jeux vidéo les plus marquants et à toute éventuelle spécificité géographique dans les préférences de jeux. Pour cela, nous utilisons un jeu de données regroupant des informations sur les ventes de jeux vidéo à travers le monde.

Description du jeu de données

Ce jeu de données provient du site suivant : <https://www.kaggle.com/datasets/gregorut/videogamesales> et du site <https://www.vgchartz.com/>; vous constaterez peut être certaines erreurs mais prenons les données comme elles sont et avec précaution !

Dans ce jeu de données vous trouverez pour certains jeux vidéos : leur nom, leur année de sortie, leur plateforme, leur genre, l'éditeur et leurs ventes en Amérique du Nord (NA_Sales), en Europe (EU_Sales), au Japon (JP_Sales), dans le reste du monde et enfin le total. Certains jeux peuvent apparaître plusieurs fois (voir tableau ??)

Rank-> Dans ce jeu de données, la variable Rank rentre en lien avec la variable Sales qui attribue au jeu un rank plus élevé en fonction d'exemplaires vendus. Il s'agit d'une variable qualitative ordinale.

```
head(vgsales,6)
```

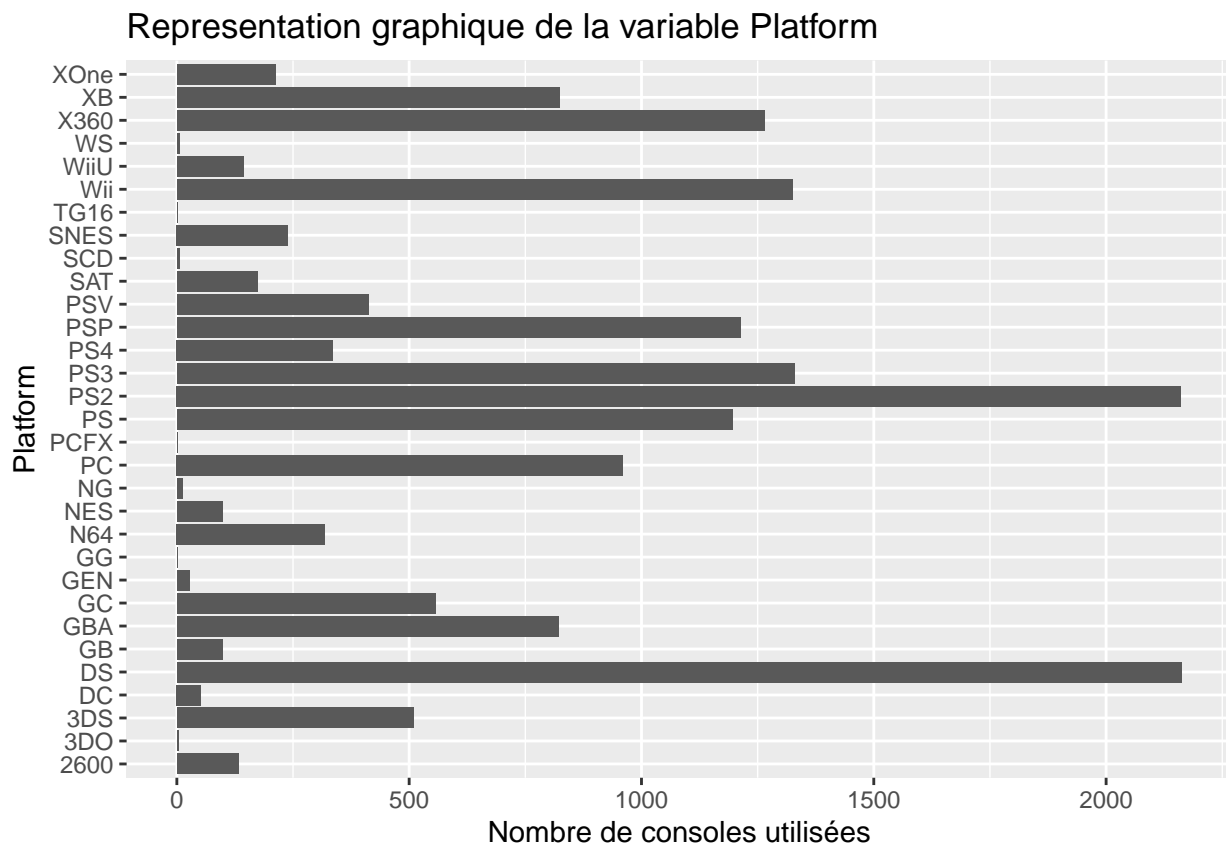
```
## # A tibble: 6 x 11
##   Rank Name      Platf~1 Year  Genre Publi~2 NA_Sa~3 EU_Sa~4 JP_Sa~5 Other~6
##   <dbl> <chr>      <chr>  <chr> <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1     1   Wii Sports   Wii    2006 Spor~ Ninten~   41.5    29.0     3.77     8.46
## 2     2   Super Mario~ NES    1985 Plat~ Ninten~   29.1     3.58     6.81     0.77
## 3     3   Mario Kart ~ Wii    2008 Raci~ Ninten~   15.8    12.9     3.79     3.31
## 4     4   Wii Sports ~ Wii    2009 Spor~ Ninten~   15.8    11.0     3.28     2.96
## 5     5   Pokemon Red~ GB     1996 Role~ Ninten~   11.3     8.89    10.2      1
## 6     6   Tetris       GB     1989 Puzz~ Ninten~   23.2     2.26     4.22     0.58
## # ... with 1 more variable: Global_Sales <dbl>, and abbreviated variable names
## #   1: Platform, 2: Publisher, 3: NA_Sales, 4: EU_Sales, 5: JP_Sales,
## #   6: Other_Sales
```

Ici, nous avons décider d'afficher les 6 premiers jeux représenté par le plus de vente situées sur différentes zones géographiques tels que L'amérique du Nord, l'Europe et le Japon.

Name -> La variable Name représente le nom du jeu associé, elle est caractérisé par une variable qualitative nominale.

Platform -> La variable Platform représente la console de jeu utilisé pour un jeu donnée, elle est caractérisé par une variable qualitative nominale.

```
ggplot(vgsales, aes(y = Platform)) +  
  geom_bar()+labs(title = "Représentation graphique de la variable Platform", x = "Nombre de consoles ut.
```



Year -> La variable Year représente la date à laquelle le jeu est sorti, cette variable est caractérisé par le type qualitative ordinaire. Pour plus de simplicité le type de cette variable va être modifié. Cette variable va devenir de type quantitative continue.

```
vgsales$An <- as.numeric(vgsales$Year)
```

```
## Warning: NAs introduced by coercion
```

Voici un sommaire graphique qui représente la variable Year :

```
vgsales$An <- as.numeric(vgsales$Year)
```

```
## Warning: NAs introduced by coercion
```

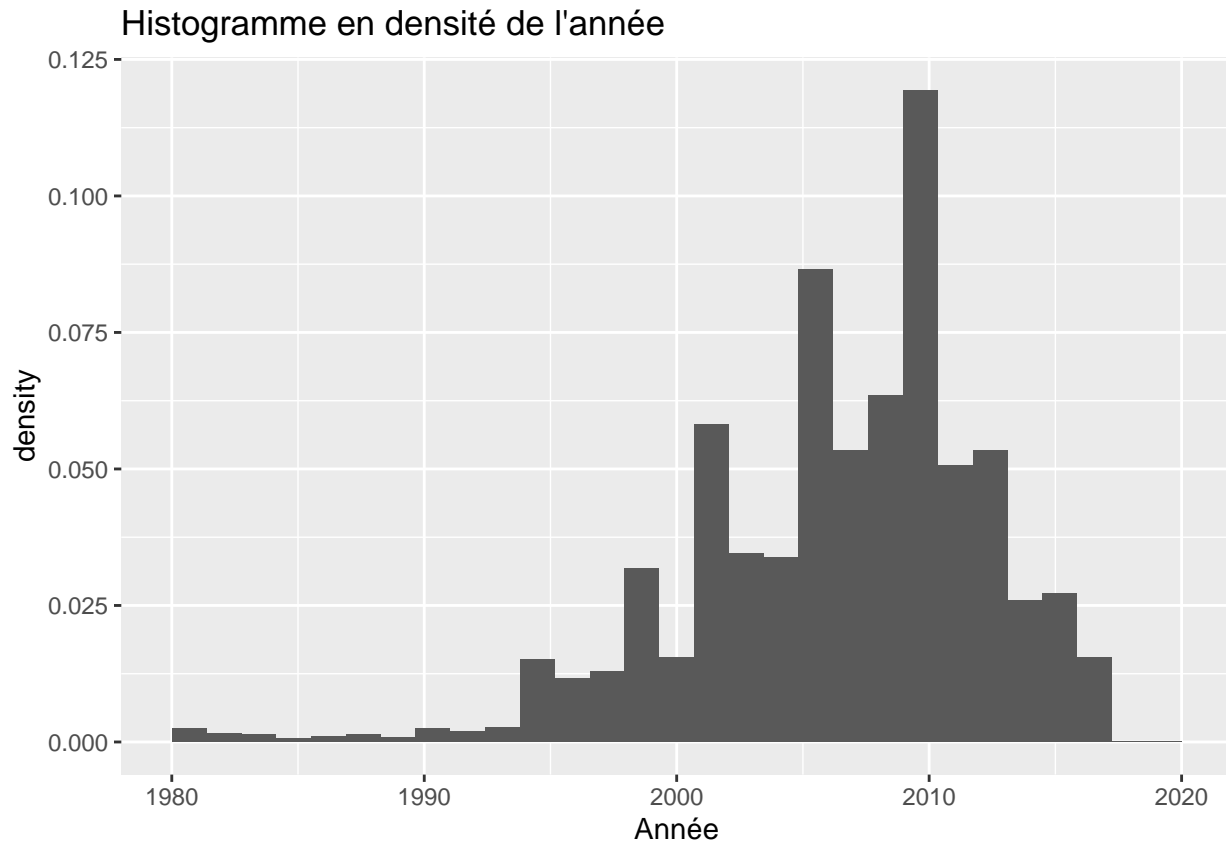
```
summary(vgsales$An)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      1980    2003    2007    2006    2010    2020      264
```

```
ggplot(vgsales, aes(x = vgsales$An)) +  
  geom_histogram(aes(y = after_stat(density)))+labs(title = "Histogramme en densité de l'année", x = "Année")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 264 rows containing non-finite values (`stat_bin()`).
```

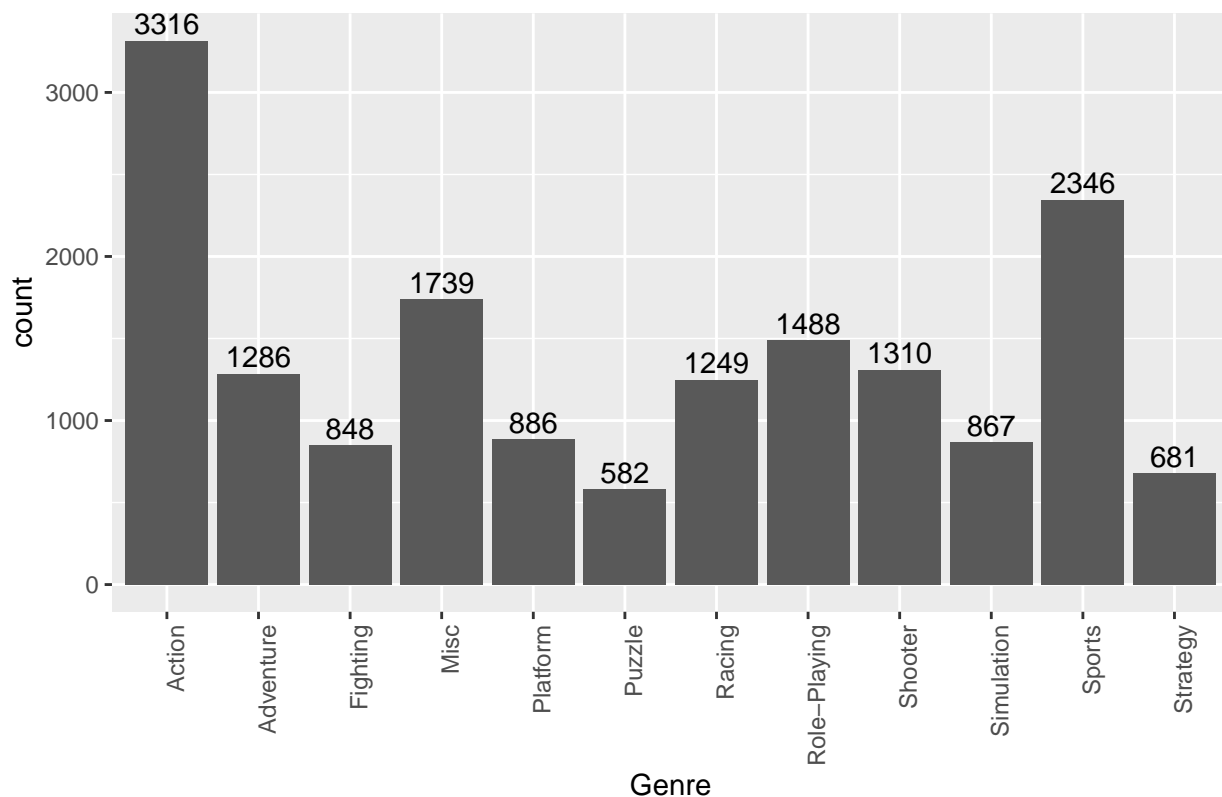


Ce graphique ci-dessus représenté caractérise grâce a un histogramme en densité, la densité ou autrement dit la présence des jeux-videos au fil des années. A travers ce graphique nous pouvons constater une augmentation significative de la densité des jeux vidéo pendant la période de 2010. Cela peut indiquer une augmentation marquée de la sortie de jeux vidéo, une popularité croissante des jeux vidéo.

Genre -> Le Genre représente le genre du jeu vidéo concernée. Elle est caractérisée par une variable de type qualitative nominale.

```
ggplot(vgsales, aes(x = Genre)) + geom_text(stat = 'count', aes(label=after_stat(count)), vjust = -0.3) +
  geom_bar() + labs(title = "Graphique representant la variable Genre") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Graphique représentant la variable Genre



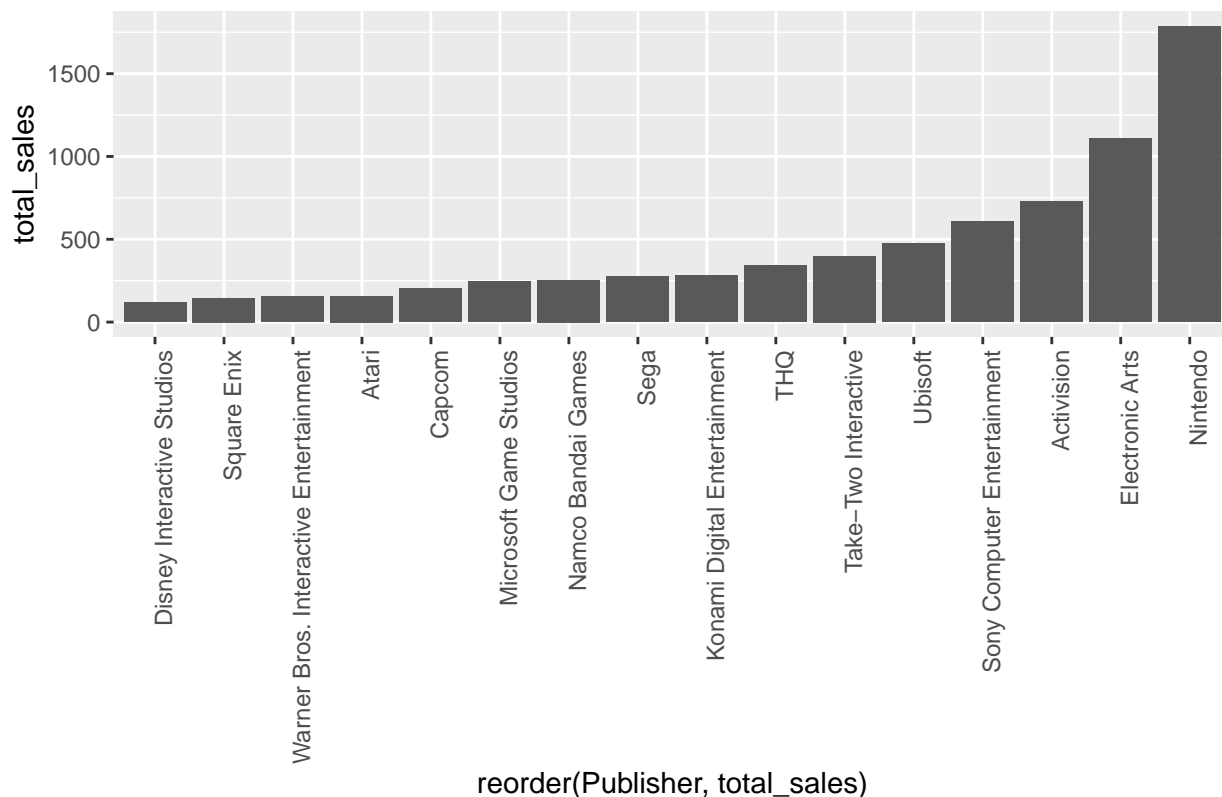
Le graphique ci-dessus illustre le nombre de jeux par genre, exprimé en milliers. Il est clair dans cette représentation que le genre 'Action' détient la plus grande quantité de jeux parmi tous les genres.

Publisher -> Publisher représente l'entreprise qui a publié le jeu. Elle est caractérisée par une variable de type qualitative nominale.

```
# Filtrer les données pour inclure uniquement les éditeurs avec plus de 100 ventes
filtered_data <- vgsales %>%
  group_by(Publisher) %>%
  summarise(total_sales = sum(Global_Sales)) %>%
  filter(total_sales > 100)

# Créer le graphique en utilisant les données filtrées
ggplot(filtered_data, aes(x = reorder(Publisher, total_sales), y = total_sales)) +
  geom_bar(stat = "identity") +
  labs(title = "Les éditeurs avec plus de 100 Millions de ventes dans le monde") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Les éditeurs avec plus de 100 Millions de ventes dans le monde



reorder(Publisher, total_sales)

Ce graphique présente les éditeurs ayant réalisé des ventes mondiales de plus de 100 millions, sans tenir compte des distinctions géographiques de territoire. On remarque que Nintendo est de loin le développeur qui vend le plus.

```
vg-sales %>%
  summarise(minimum = min(NA_Sales), moyenne = round(mean(NA_Sales), 3), maximum = max(NA_Sales), Q1 = quantile(NA_Sales, 0.25), Q3 = quantile(NA_Sales, 0.75))
```

```
## # A tibble: 1 x 5
##   minimum moyenne maximum    Q1    Q3
##   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1      0  0.265    41.5     0  0.24
```

```
vg-sales %>%
  summarise(minimum = min(EU_Sales), moyenne = round(mean(EU_Sales), 3), maximum = max(EU_Sales), Q1 = quantile(EU_Sales, 0.25), Q3 = quantile(EU_Sales, 0.75))
```

```
## # A tibble: 1 x 5
##   minimum moyenne maximum    Q1    Q3
##   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1      0  0.147    29.0     0  0.11
```

```
vg-sales %>%
  summarise(minimum = min(JP_Sales), moyenne = round(mean(JP_Sales), 3), maximum = max(JP_Sales), Q1 = quantile(JP_Sales, 0.25), Q3 = quantile(JP_Sales, 0.75))
```

```
## # A tibble: 1 x 5
##   minimum moyenne maximum    Q1    Q3
##   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1      0  0.078    10.2     0  0.04
```

```
vg-sales %>%
  summarise(minimum = min(Other_Sales), moyenne = round(mean(Other_Sales), 3), maximum = max(Other_Sales), Q1 = quantile(Other_Sales, 0.25), Q3 = quantile(Other_Sales, 0.75))
```

```
## # A tibble: 1 x 5
##   minimum moyenne maximum    Q1    Q3
##   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1      0    0.048    10.6     0    0.04
```

```
vg-sales %>%
```

```
summarise(minimum = min(Global_Sales), moyenne = round(mean(Global_Sales), 3), maximum = max(Global_Sales))
```

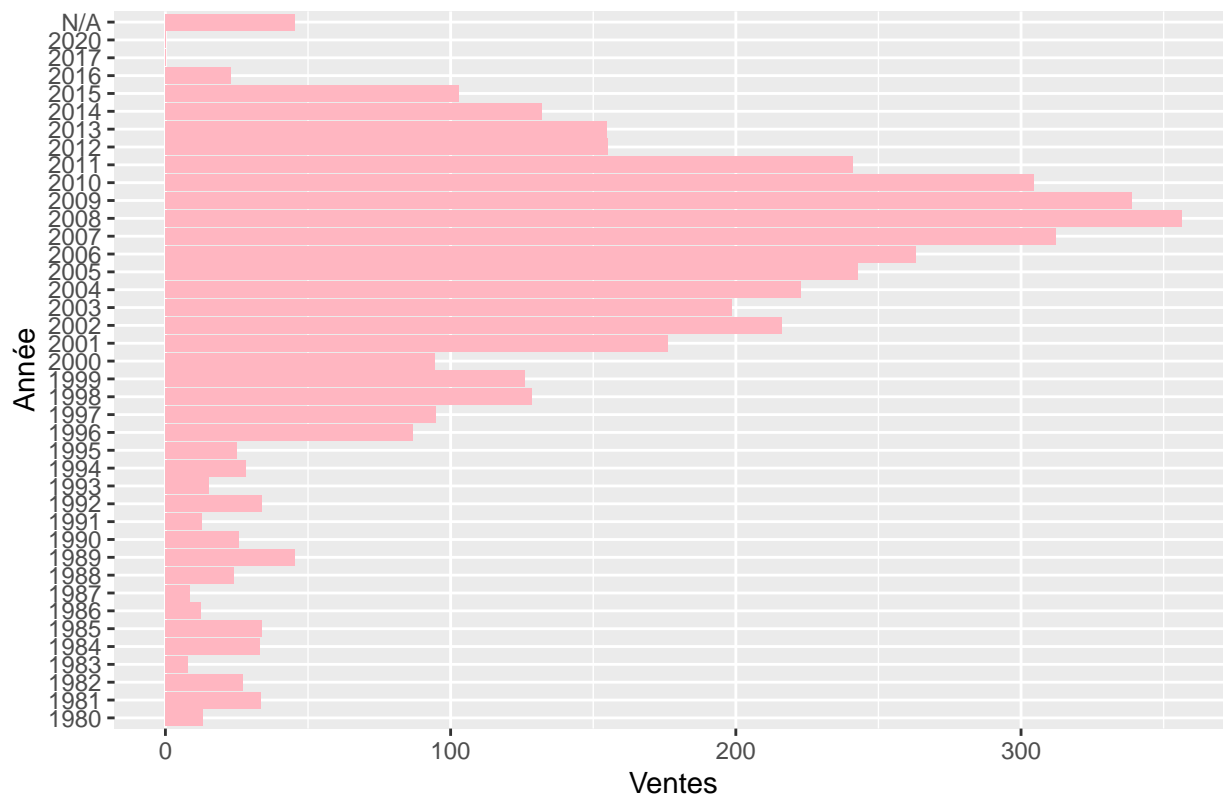
```
## # A tibble: 1 x 5
##   minimum moyenne maximum    Q1    Q3
##   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1    0.01    0.537    82.7  0.06  0.47
```

Parmi ces résumés de données, il est clairement indiqué que l'Amérique du Nord est le territoire géographique où les ventes de jeux vidéo sont les plus élevées, suivies dans l'ordre décroissant par l'Europe et le Japon. Ces données sont corrélées à la population ; ainsi, il est évident que les territoires géographiques avec une population plus faible vendent moins de jeux vidéo. Cependant l'Europe détient une plus grande population comparé à l'Amérique du Nord de ce fait on peut conclure une culture du jeux-vidéo plus accrue en Amérique du Nord. Population en Amérique du Nord → 579 millions Population en Europe → 746,4 millions Population au Japon → 125,7 millions

NA_Sales → Corresponds aux nombres de ventes en Amérique du Nord. C'est une variable quantitative continue.

```
ggplot(vg-sales, aes(y = Year, x = NA_Sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Comparaison du nombre de ventes en Amérique du Nord par année",
       y = "Année",
       x = "Ventes")
```

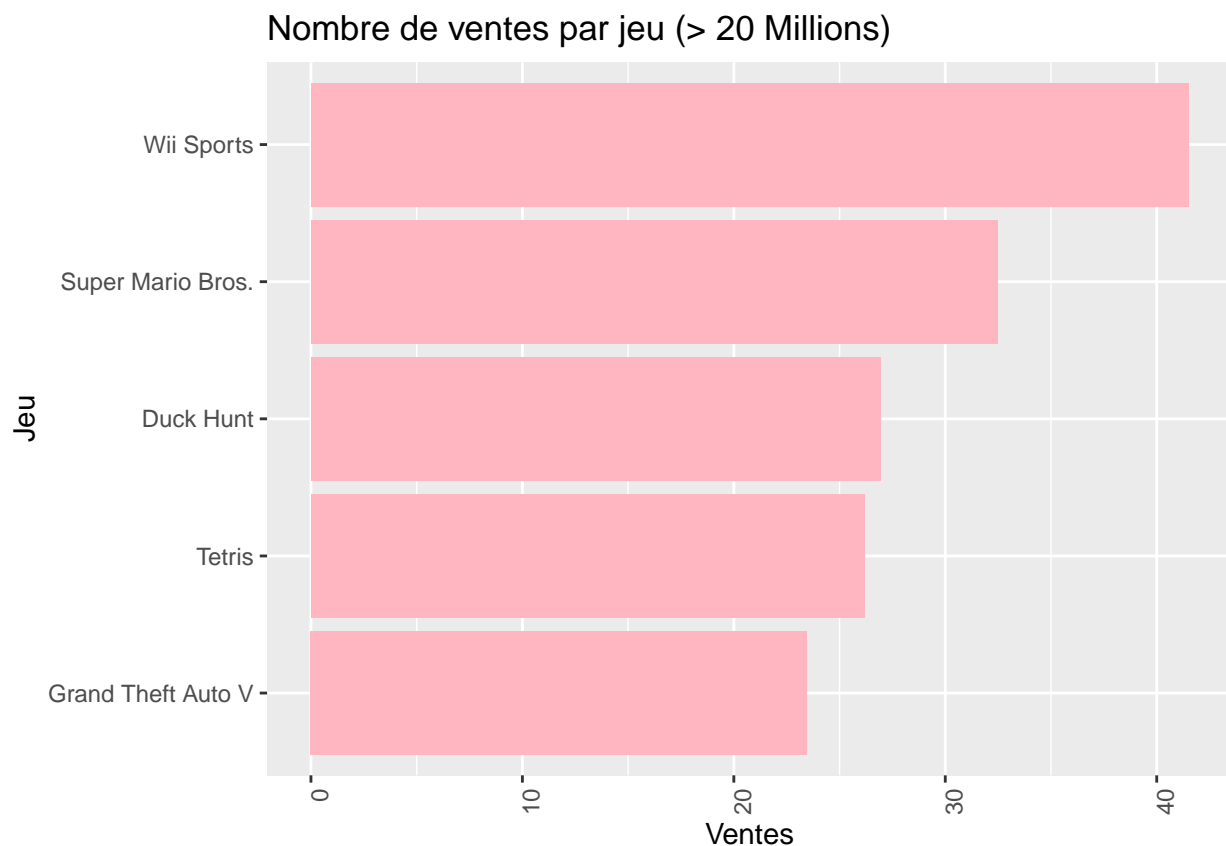
Comparaison du nombre de ventes en Amérique du Nord par année



Le graphique présenté ci-dessus illustre les ventes de jeux vidéo par année, sur le territoire géographique spécifique de l'Amérique du Nord. Une forte croissance des ventes est observée à partir de 2005 et aux alentours de l'année 2010.

```
# Filtrer les jeux avec plus de 20 ventes
filtered_data6 <- vgsales %>%
  group_by(Name) %>%
  summarise(total_sales = sum(NA_Sales)) %>%
  filter(total_sales > 20)

# Création du diagramme adapté
ggplot(filtered_data6, aes(x = reorder(Name, total_sales), y = total_sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Nombre de ventes par jeu (> 20 Millions)",
       x = "Jeu",
       y = "Ventes") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```

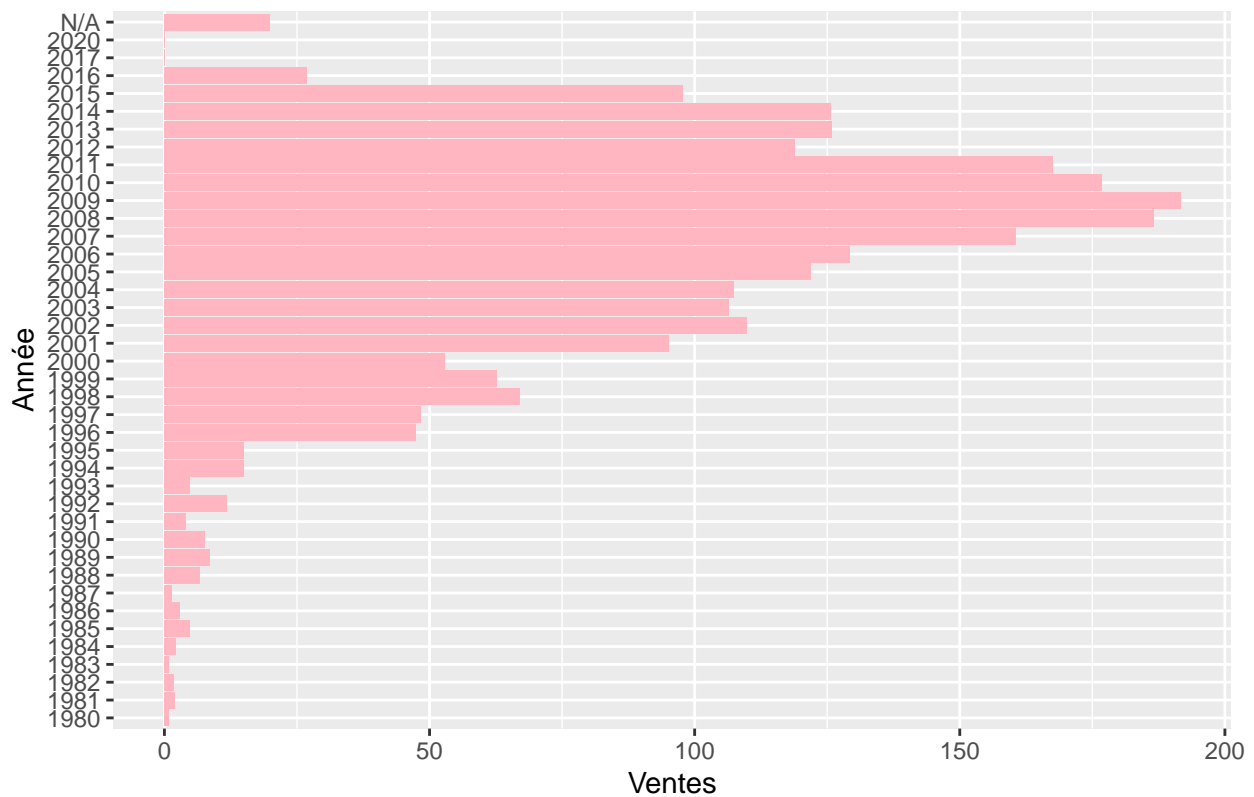


Ce deuxième graphique représente les jeux vidéo qui on le plus réussis. Pour simplifier, nous avons inclus uniquement les jeux ayant totalisé plus de 20 millions de ventes.

EU_Sales -> Corresponds aux nombres de ventes en europe. C'est une variable quantitative continu.

```
ggplot(vgsales, aes(y = Year, x = EU_Sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Comparaison du nombre de ventes en Europe par année",
       y = "Année",
       x = "Ventes")
```

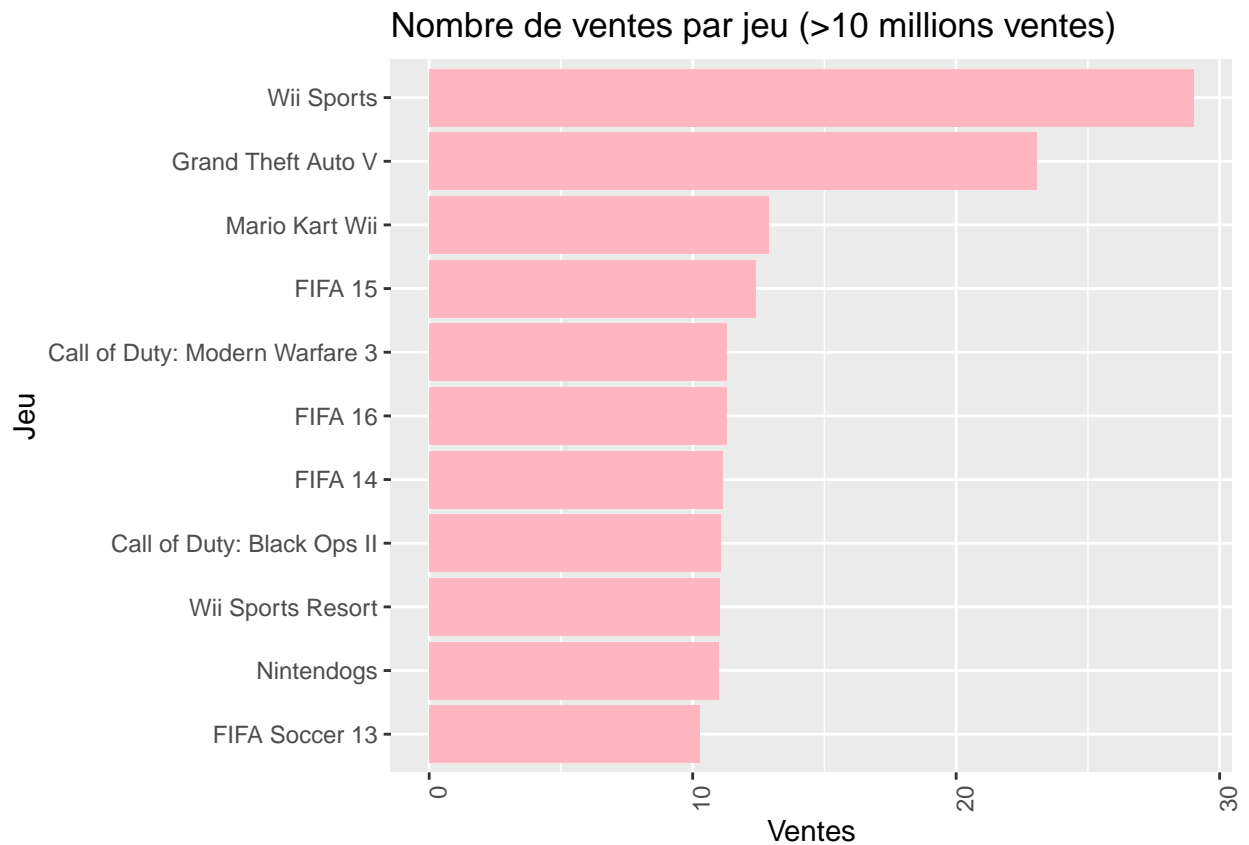
Comparaison du nombre de ventes en Europe par année



Le graphique présenté ci-dessus illustre les ventes de jeux vidéo par année, sur le territoire géographique spécifique de l'Europe. Une forte croissance des ventes est observée à partir de 2005 et aux alentours de l'année 2008.

```
# Filtrer les jeux avec plus de 10 ventes
filtered_data5 <- vgsales %>%
  group_by(Name) %>%
  summarise(total_sales = sum(EU_Sales)) %>%
  filter(total_sales > 10)

# Création du diagramme adapté
ggplot(filtered_data5, aes(x = reorder(Name, total_sales), y = total_sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Nombre de ventes par jeu (>10 millions ventes)",
        x = "Jeu",
        y = "Ventes") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```

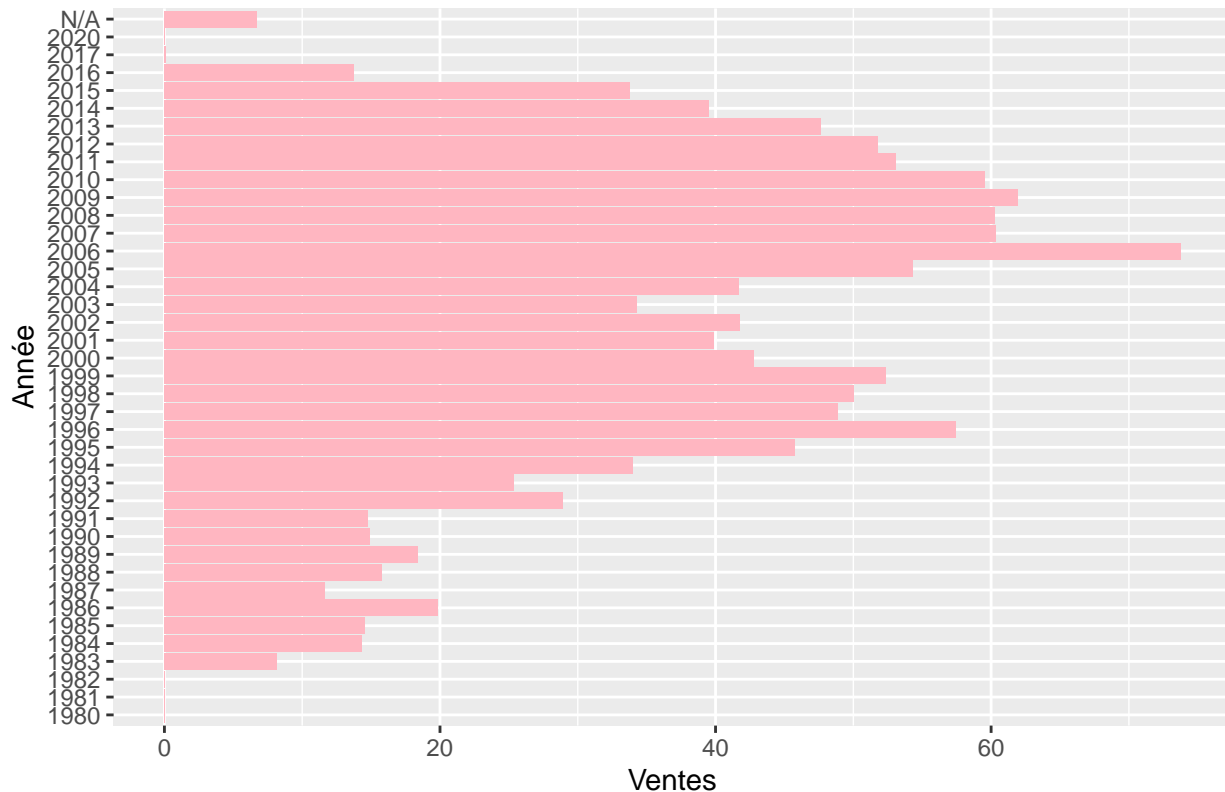



Ce deuxième graphique représente les jeux vidéo qui on le plus réussis. Pour simplifier, nous avons inclus uniquement les jeux ayant totalisé plus de 10 millions de ventes. On remarque que les jeux les plus vendus en Europe sont similaires aux jeux-vidéos vendus en Amérique du Nord, cependant on observe au niveau des genres des jeux-vidéos present en Europe l'ajout de jeux-vidéos "sport".

JP_Sales -> Corresponds aux nombres de ventes au Japon. C'est une variable quantitative continu.

```
ggplot(vgsales, aes(y = Year, x = JP_Sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Comparaison du nombre de ventes au Japon par année",
       y = "Année",
       x = "Ventes")
```

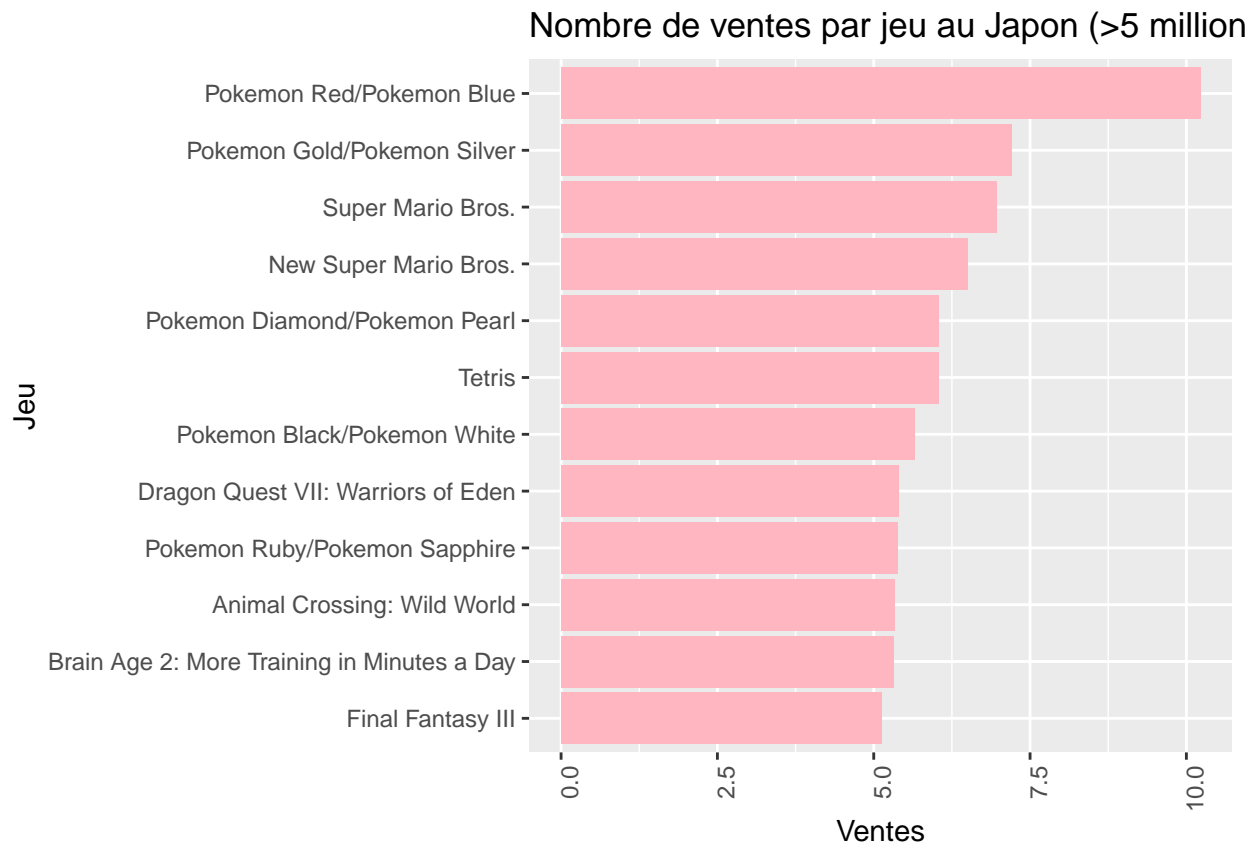
Comparaison du nombre de ventes au Japon par année



Le graphique présenté ci-dessus illustre les ventes de jeux vidéo par année, sur le territoire géographique spécifique du Japon. Une forte croissance des ventes est observée à partir de 1994 et aux alentours de l'année 2005. On remarque que la popularité des jeux vidéos au Japon a commencé plus tôt que l'Europe et l'Amérique du Nord.

```
# Filtrer les jeux avec plus de 20 ventes
filtered_data4 <- vgsales %>%
  group_by(Name) %>%
  summarise(total_sales = sum(JP_Sales)) %>%
  filter(total_sales > 5)

# Création du diagramme adapté
ggplot(filtered_data4, aes(x = reorder(Name, total_sales), y = total_sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Nombre de ventes par jeu au Japon (>5 millions ventes)",
        x = "Jeu",
        y = "Ventes") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```

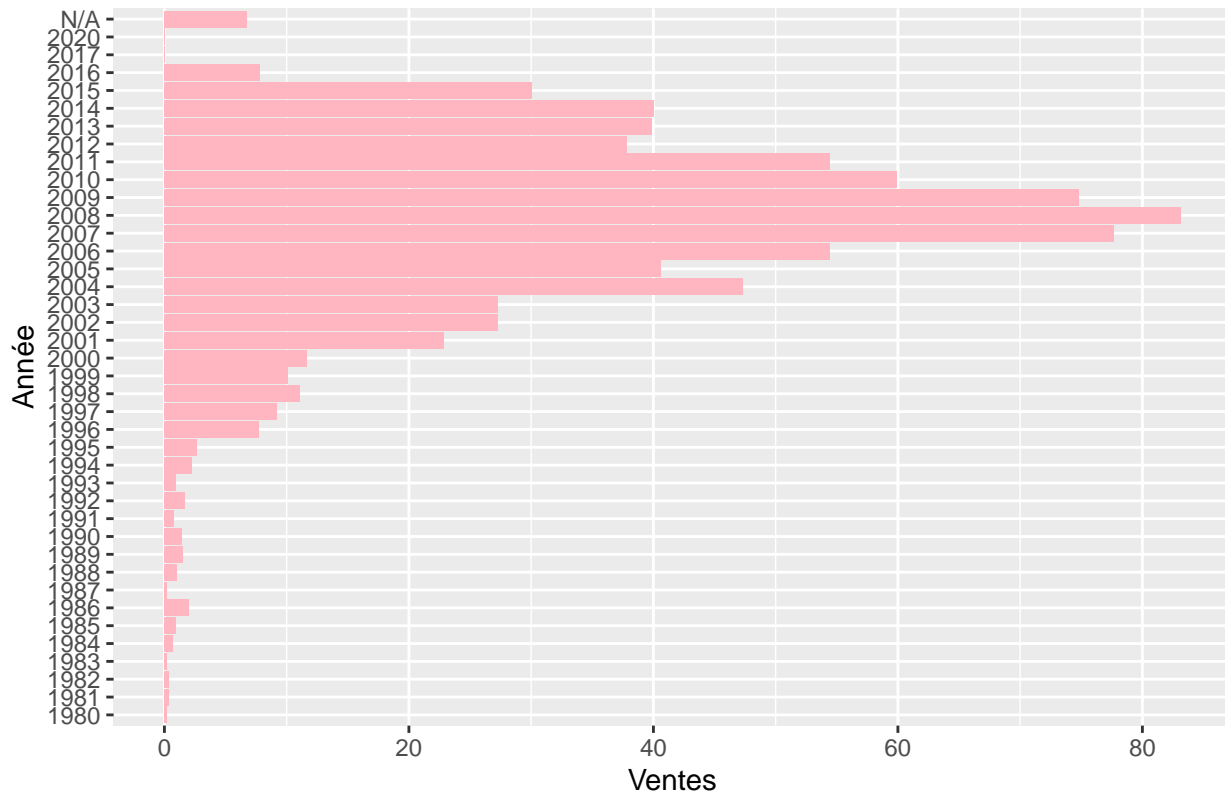


Ce deuxième graphique représente les jeux vidéo qui ont le plus réussi. Pour simplifier, nous avons inclus uniquement les jeux ayant totalisé plus de 5 millions de ventes. On observe au Japon que les jeux-vidéos les plus vendus sont totalement différents des genres observés en Europe et en Amérique du Nord. Le genre le plus répandu est "MMORPG"/Aventure. Il s'agit d'un genre de jeux vidéo associant le jeu de rôle et le jeu en ligne massivement multijoueur, permettant à un grand nombre de personnes d'interagir simultanément.

Other_Sales -> Corresponds aux nombres de ventes dans les autres régions du monde. C'est une variable quantitative continue.

```
ggplot(vgsales, aes(y = Year, x = Other_Sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Comparaison du nombre de ventes dans les autres régions du monde par année",
        y = "Année",
        x = "Ventes")
```

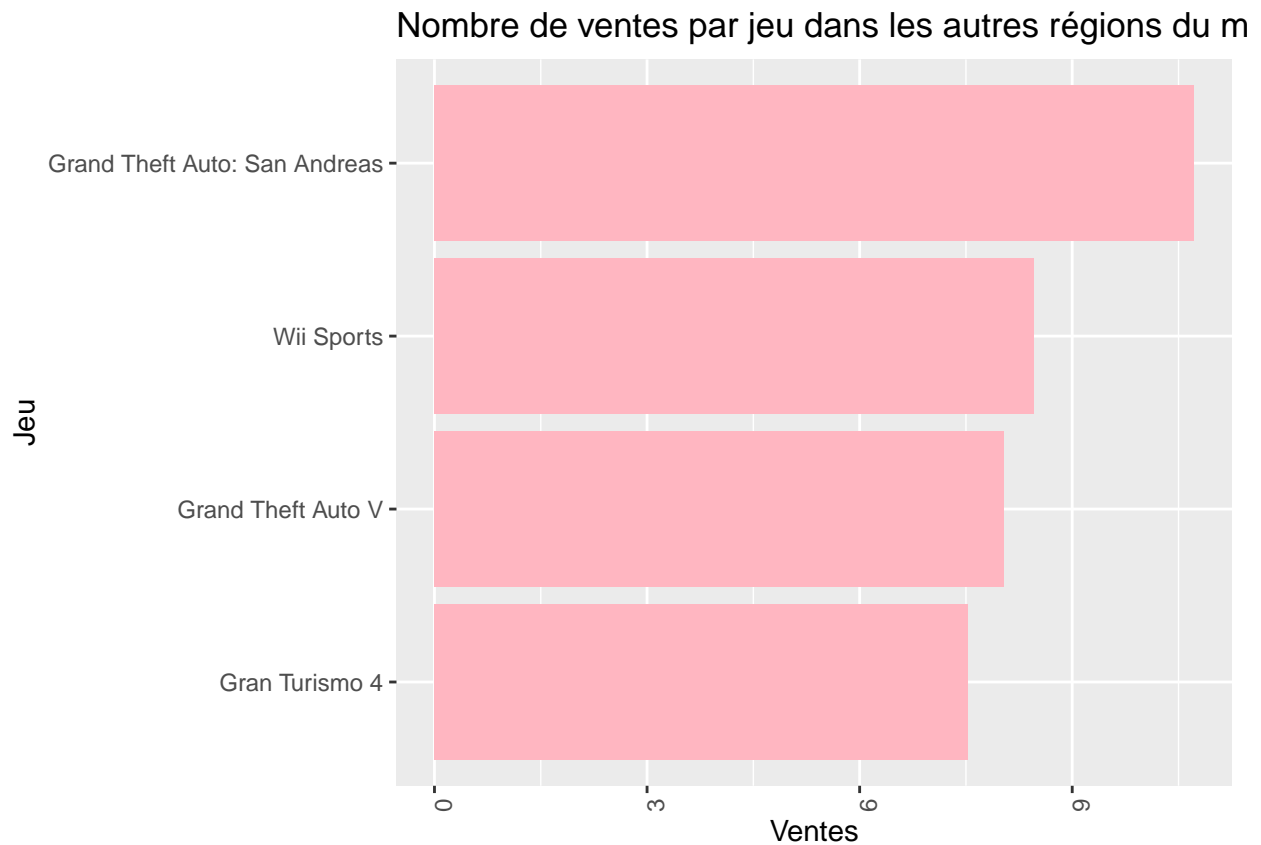
Comparaison du nombre de ventes dans les autres régions du monde par



Le graphique présenté ci-dessus illustre les ventes de jeux vidéo par année, sur les autres régions du monde. Une forte croissance des ventes est observée à partir de 2001 et aux alentours de l'année 2005.

```
# Filtrer les jeux avec plus de 20 ventes
filtered_data3 <- vgsales %>%
  group_by(Name) %>%
  summarise(total_sales = sum(Other_Sales)) %>%
  filter(total_sales > 5)

# Création du diagramme adapté
ggplot(filtered_data3, aes(x = reorder(Name, total_sales), y = total_sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Nombre de ventes par jeu dans les autres régions du monde (>5 millions ventes)",
       x = "Jeu",
       y = "Ventes") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```

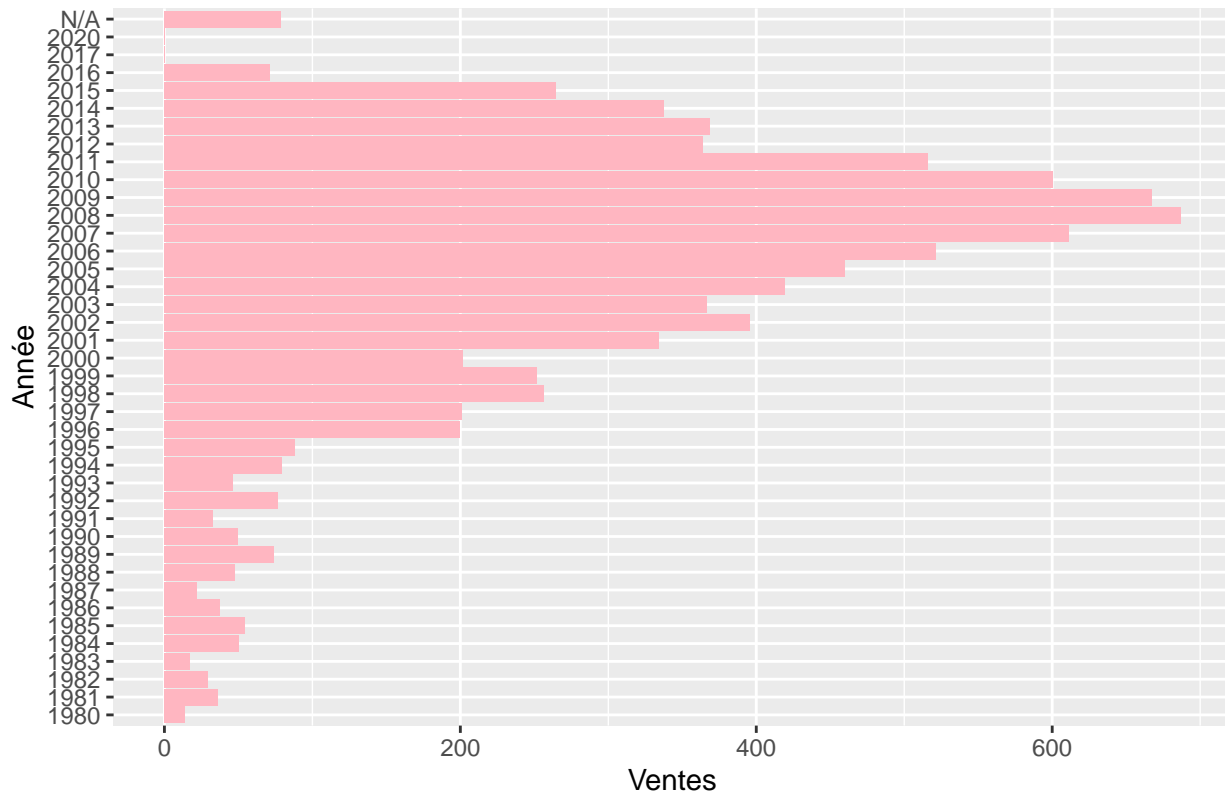


Ce deuxième graphique représente les jeux vidéo qui on le plus réussis. Pour simplifier, nous avons inclus uniquement les jeux ayant totalisé plus de 5 millions de ventes. Les jeux énoncés dans ce graphique sont globalement similaire a ceux vendus en Amérique du Nord.

Global_Sales -> Corresponds aux nombres de ventes dans le monde entier. C'est une variable quantitative continu.

```
ggplot(vgsales, aes(y = Year, x = Global_Sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Comparaison du nombre de ventes dans le entier monde par année",
        y = "Année",
        x = "Ventes")
```

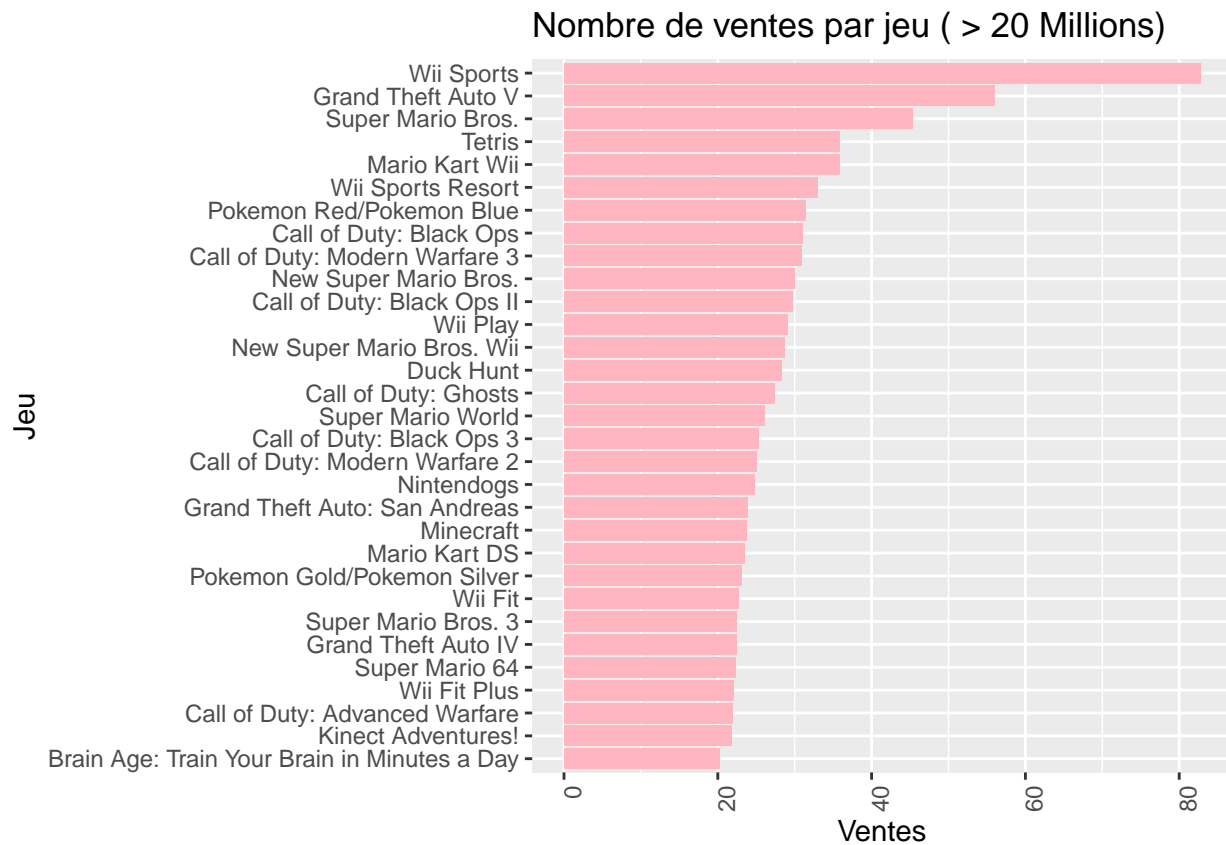
Comparaison du nombre de ventes dans le entier monde par année



Le graphique présenté ci-dessus illustre les ventes de jeux vidéo par année, sur le monde entier. Une forte croissance des ventes est observée à partir de 1996 et aux alentours de l'année 2005.

```
# Filtrer les jeux avec plus de 20 ventes
filtered_data2 <- vgsales %>%
  group_by(Name) %>%
  summarise(total_sales = sum(Global_Sales)) %>%
  filter(total_sales > 20)

# Création du diagramme adapté
ggplot(filtered_data2, aes(x = reorder(Name, total_sales), y = total_sales)) +
  geom_bar(stat = "identity", fill = "LightPink") +
  labs(title = "Nombre de ventes par jeu ( > 20 Millions)",
       x = "Jeu",
       y = "Ventes") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```



Ce deuxième graphique représente les jeux vidéo qui on le plus réussis. Pour simplifier, nous avons inclus uniquement les jeux ayant totalisé plus de 20 millions de ventes. Les jeux les plus vendus mondialement sont plus present sur les zones géographiques tels que l'Amérique du Nord ainsi que l'Europe. **Le top 20 :**

- Wii Sports
- Grand Theft Auto V
- Super Mario Bros.
- Tetris
- Mario Kart Wii
- Wii Sports Resort
- Pokemon Red/Pokemon Blue
- Call of Duty: Black Ops
- Call of Duty: Modern Warfare 3
- New Super Mario Bros.
- Call of Duty: Black Ops II
- Wii Play
- New Super Mario Bros. Wii
- Duck Hunt
- Call of Duty: Ghosts
- Super Mario World

- Call of Duty: Black Ops 3
- Call of Duty: Modern Warfare 2
- Nintendogs
- Grand Theft Auto: San Andreas
- Minecraft

Analyse des données

En examinant les données, nous avons constaté que l'Amérique du Nord est le territoire géographique avec les ventes de jeux vidéo les plus élevées, suivie de l'Europe et du Japon. Cette tendance est cohérente avec la taille de la population de chaque région, les territoires avec une population plus importante ayant tendance à vendre plus de jeux vidéo.

En ce qui concerne les jeux les plus vendus, nous avons identifié des titres emblématiques tels que "Wii Sports", "Grand Theft Auto V" et "Super Mario Bros.". Ces jeux ont marqué l'industrie du jeu vidéo par leurs ventes massives et leur impact culturel significatif.

En examinant les données par région géographique, nous avons observé des différences dans les préférences de genre et de plateforme. Par exemple, les jeux de sport sont plus populaires en Europe, tandis que le Japon montre une préférence marquée pour les jeux de rôle massivement multijoueurs (MMORPG).

Conclusion

La mode des jeux vidéo a pris naissance principalement au Japon mais a clairement explosé en Amérique du Nord qui domine les ventes, avec une expansion rapide vers d'autres régions du monde.

Les jeux vidéo les plus marquants ont souvent une portée mondiale, mais il existe des variations dans les préférences de genre et de plateforme selon les régions géographiques.

Ces variations témoignent de la diversité culturelle et des tendances de consommation propres à chaque région. En somme, les jeux vidéo ont réussi à devenir un phénomène mondial en s'adaptant aux goûts et aux préférences des joueurs à travers le monde.