

## BUS4066 - Group 11 - Video Game Sales

---

### Introduction

This dataset contains a list of video games with sales greater than 100,000 copies.

Fields include:

Rank - Ranking of overall sales

Name - The games name

Platform - Platform of the games release (i.e. PC,PS4, etc.)

Year - Year of the game's release

Genre - Genre of the game

Publisher - Publisher of the game

NA\_Sales - Sales in North America (in millions)

EU\_Sales - Sales in Europe (in millions)

JP\_Sales - Sales in Japan (in millions)

Other\_Sales - Sales in the rest of the world (in millions)

Global\_Sales - Total worldwide sales.

Group repository - [https://github.com/conie-gbc/BUS4066-Analytics\\_\\_Grp11](https://github.com/conie-gbc/BUS4066-Analytics__Grp11)

---

Load data - vgsales.csv from <https://www.kaggle.com/gregorut/videogamesales>

```
VideoGamedata <- readr::read_csv(file ="vgsales.csv")
```

---

## Structure of dataset

```
str(VideoGamedata)
```

```
## spec_tbl_df[,11] [16,598 x 11] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Rank      : num [1:16598] 1 2 3 4 5 6 7 8 9 10 ...
## $ Name      : chr [1:16598] "Wii Sports" "Super Mario Bros." "Mario Kart Wii" "Wii Sports Resort"
## $ Platform  : chr [1:16598] "Wii" "NES" "Wii" "Wii" ...
## $ Year      : chr [1:16598] "2006" "1985" "2008" "2009" ...
## $ Genre     : chr [1:16598] "Sports" "Platform" "Racing" "Sports" ...
## $ Publisher : chr [1:16598] "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
## $ NA_Sales  : num [1:16598] 41.5 29.1 15.8 15.8 11.3 ...
## $ EU_Sales  : num [1:16598] 29.02 3.58 12.88 11.01 8.89 ...
## $ JP_Sales  : num [1:16598] 3.77 6.81 3.79 3.28 10.22 ...
## $ Other_Sales : num [1:16598] 8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...
## $ Global_Sales: num [1:16598] 82.7 40.2 35.8 33 31.4 ...
## - attr(*, "spec")=
## .. cols(
## ..   Rank = col_double(),
## ..   Name = col_character(),
## ..   Platform = col_character(),
## ..   Year = col_character(),
## ..   Genre = col_character(),
## ..   Publisher = col_character(),
## ..   NA_Sales = col_double(),
## ..   EU_Sales = col_double(),
## ..   JP_Sales = col_double(),
## ..   Other_Sales = col_double(),
## ..   Global_Sales = col_double()
## .. )
```

---

## List the variables in dataset

```
names(VideoGamedata)
```

```
## [1] "Rank"      "Name"      "Platform"  "Year"      "Genre"
## [6] "Publisher" "NA_Sales"  "EU_Sales"  "JP_Sales"  "Other_Sales"
## [11] "Global_Sales"
```

---

## Top 15 rows of dataset

```
head (VideoGamedata, 15)
```

```
## # A tibble: 15 x 11
##   Rank Name           Platform Year  Genre Publisher NA_Sales EU_Sales JP_Sales
##   <dbl> <chr>         <chr>   <chr> <chr>   <chr>      <dbl>   <dbl>   <dbl>
## 1     1 1 Wii Sports    Wii     2006 Sports Nintendo  41.5    29.0    3.77
## 2     2 2 Super Mario~ NES      1985 Platf~ Nintendo  29.1     3.58    6.81
## 3     3 3 Mario Kart ~ Wii     2008 Racing Nintendo  15.8    12.9    3.79
## 4     4 4 Wii Sports ~ Wii     2009 Sports Nintendo  15.8    11.0    3.28
## 5     5 5 Pokemon Red~ GB       1996 Role~ Nintendo  11.3     8.89   10.2
## 6     6 6 Tetris        GB       1989 Puzzle Nintendo  23.2     2.26    4.22
## 7     7 7 New Super M~ DS       2006 Platf~ Nintendo  11.4     9.23    6.5
## 8     8 8 Wii Play      Wii     2006 Misc  Nintendo  14.0     9.2    2.93
## 9     9 9 New Super M~ Wii     2009 Platf~ Nintendo  14.6     7.06    4.7
## 10    10 10 Duck Hunt    NES      1984 Shoot~ Nintendo  26.9     0.63    0.28
## 11    11 11 Nintendogs    DS       2005 Simul~ Nintendo   9.07     11     1.93
## 12    12 12 Mario Kart ~ DS       2005 Racing Nintendo   9.81     7.57    4.13
## 13    13 13 Pokemon Gol~ GB       1999 Role~ Nintendo    9      6.18    7.2
## 14    14 14 Wii Fit       Wii     2007 Sports Nintendo   8.94     8.03    3.6
## 15    15 15 Wii Fit Plus Wii     2009 Sports Nintendo   9.09     8.59    2.53
## # ... with 2 more variables: Other_Sales <dbl>, Global_Sales <dbl>
```

---

## Remove missing values in dataset

```
str(VideoGamedata)
complete.cases(VideoGamedata)
x <- VideoGamedata[complete.cases(VideoGamedata), ]
str(x)
```

---

## User defined function using any of the variables from the data set

```
Total <- function(x,y)
{x+y}
Total (VideoGamedata$Other_Sales,VideoGamedata$Global_Sales)
```

---

Data manipulation techniques and filter rows based on any logical criteria that exist in dataset

```
VideoGamedata %>% filter(Global_Sales > 30.00)
```

```
## # A tibble: 7 x 11
##   Rank Name      Platform Year  Genre  Publisher NA_Sales EU_Sales JP_Sales
##   <dbl> <chr>      <chr>   <chr> <chr>   <chr>      <dbl>   <dbl>   <dbl>
## 1     1  Wii Sports   Wii     2006  Sports Nintendo    41.5    29.0    3.77
## 2     2 Super Mario~ NES     1985  Platfo~ Nintendo    29.1     3.58    6.81
## 3     3 Mario Kart ~ Wii     2008  Racing Nintendo    15.8    12.9    3.79
## 4     4 Wii Sports ~ Wii     2009  Sports Nintendo    15.8    11.0    3.28
## 5     5 Pokemon Red~ GB      1996  Role-P~ Nintendo    11.3     8.89   10.2
## 6     6 Tetris      GB      1989  Puzzle Nintendo    23.2     2.26    4.22
## 7     7 New Super M~ DS      2006  Platfo~ Nintendo    11.4     9.23    6.5
## # ... with 2 more variables: Other_Sales <dbl>, Global_Sales <dbl>
```

Dependent & independent variables and use reshaping techniques and create a new data frame by joining those variables from your dataset

```
##We want to predict North America Sales from Global Sales
```

```
fit <-lm(NA_Sales ~ Global_Sales, data=VideoGamedata)
```

```
##Values of coefficients(0s) are -0.0009505 and 0.4942274,
```

```
##-> Prediction equation for model is as below:
```

```
##NA_Sales = -0.0009505 + 0.4942274*Global_Sales
```

```
summary(fit)
```

```
##
```

```
## Call:
```

```
## lm(formula = NA_Sales ~ Global_Sales, data = VideoGamedata)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -5.2423 -0.0287  0.0014  0.0415 12.9394
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.0009505  0.0022689  -0.419   0.675
```

```
## Global_Sales  0.4942274  0.0013791 358.380 <2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.2763 on 16596 degrees of freedom
```

```
## Multiple R-squared:  0.8856, Adjusted R-squared:  0.8856
```

```
## F-statistic: 1.284e+05 on 1 and 16596 DF, p-value: < 2.2e-16
```

```
#Actual values of North America Sales of top 15 are as below
```

```
head(VideoGamedata$NA_Sales,15)
```

```
## [1] 41.49 29.08 15.85 15.75 11.27 23.20 11.38 14.03 14.59 26.93  9.07  9.81
```

```
## [13]  9.00  8.94  9.09
```

```
#Predicted values of North America Sales of top 15 are as below
```

```
head(fitted(fit),15)
```

```
##           1           2           3           4           5           6           7           8
```

```
## 40.89142 19.88676 17.70227 16.30855 15.50296 14.95437 14.83081 14.34153
```

```
##           9           10          11          12          13          14          15
```

```
## 14.14384 13.99063 12.23612 11.57386 11.41570 11.22790 10.87205
```

```
##Use reshaping techniques cbind()
```

```
##to add the predicted North America Sales as last column
```

```
predictedNA_Sales <-c(fitted(fit))
```

```
NewVideoGamedata <-cbind(VideoGamedata, predictedNA_Sales)
```

```
head(NewVideoGamedata,15)
```

```
##      Rank      Name Platform Year      Genre Publisher
## 1      1      Wii Sports    Wii 2006    Sports  Nintendo
## 2      2  Super Mario Bros.   NES 1985 Platform  Nintendo
## 3      3    Mario Kart Wii    Wii 2008    Racing  Nintendo
## 4      4  Wii Sports Resort    Wii 2009    Sports  Nintendo
```

```
## 5      5      Pokemon Red/Pokemon Blue      GB 1996 Role-Playing Nintendo
## 6      6      Tetris      GB 1989      Puzzle Nintendo
## 7      7      New Super Mario Bros.      DS 2006      Platform Nintendo
## 8      8      Wii Play      Wii 2006      Misc Nintendo
## 9      9      New Super Mario Bros. Wii      Wii 2009      Platform Nintendo
## 10     10     Duck Hunt      NES 1984      Shooter Nintendo
## 11     11     Nintendogs      DS 2005      Simulation Nintendo
## 12     12     Mario Kart DS      DS 2005      Racing Nintendo
## 13     13     Pokemon Gold/Pokemon Silver      GB 1999 Role-Playing Nintendo
## 14     14     Wii Fit      Wii 2007      Sports Nintendo
## 15     15     Wii Fit Plus      Wii 2009      Sports Nintendo
##      NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales predictedNA_Sales
## 1      41.49    29.02    3.77      8.46      82.74      40.89142
## 2      29.08     3.58    6.81     0.77     40.24     19.88676
## 3      15.85    12.88    3.79     3.31     35.82     17.70227
## 4      15.75    11.01    3.28     2.96     33.00     16.30855
## 5      11.27     8.89   10.22     1.00     31.37     15.50296
## 6      23.20     2.26    4.22     0.58     30.26     14.95437
## 7      11.38     9.23    6.50     2.90     30.01     14.83081
## 8      14.03     9.20    2.93     2.85     29.02     14.34153
## 9      14.59     7.06    4.70     2.26     28.62     14.14384
## 10     26.93     0.63    0.28     0.47     28.31     13.99063
## 11      9.07    11.00    1.93     2.75     24.76     12.23612
## 12      9.81     7.57    4.13     1.92     23.42     11.57386
## 13      9.00     6.18    7.20     0.71     23.10     11.41570
## 14      8.94     8.03    3.60     2.15     22.72     11.22790
## 15      9.09     8.59    2.53     1.79     22.00     10.87205
```

```
##list the variables in NewVideoGamedata
```

```
names(NewVideoGamedata)
```

```
## [1] "Rank"      "Name"      "Platform"
## [4] "Year"      "Genre"     "Publisher"
## [7] "NA_Sales"  "EU_Sales"  "JP_Sales"
## [10] "Other_Sales" "Global_Sales" "predictedNA_Sales"
```

```
##list the structure of NewVideoGamedata
```

```
str(NewVideoGamedata)
```

```
## 'data.frame': 16598 obs. of 12 variables:
## $ Rank      : num  1 2 3 4 5 6 7 8 9 10 ...
## $ Name      : chr   "Wii Sports" "Super Mario Bros." "Mario Kart Wii" "Wii Sports Resort" ...
## $ Platform   : chr   "Wii" "NES" "Wii" "Wii" ...
## $ Year       : chr   "2006" "1985" "2008" "2009" ...
## $ Genre      : chr   "Sports" "Platform" "Racing" "Sports" ...
## $ Publisher  : chr   "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
## $ NA_Sales   : num  41.5 29.1 15.8 15.8 11.3 ...
## $ EU_Sales   : num  29.02 3.58 12.88 11.01 8.89 ...
## $ JP_Sales   : num  3.77 6.81 3.79 3.28 10.22 ...
## $ Other_Sales : num  8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...
## $ Global_Sales : num  82.7 40.2 35.8 33 31.4 ...
## $ predictedNA_Sales: num  40.9 19.9 17.7 16.3 15.5 ...
```

---

## Remove missing values in dataset

```
##To check missing values in dataset  
VideoGameDataNew=is.na(VideoGamedata)  
##To remove Missing values from Dataset VideoGameDataNew  
na.omit(VideoGameDataNew)  
##result hidden due to number of rows
```

---

## Identify and remove duplicated data in dataset

```
VideoGameDataDuplicate =VideoGamedata  
##VideoGameDataDuplicate[!duplicated(VideoGameDataDuplicate$Sepal.Width),]  
VideoGameDataDuplicate %>% distinct()  
##result hidden due to number of rows
```

## Reorder multiple rows in descending order

```
VideoGamedata %>% arrange(desc(Genre, Publisher))
```

```
## # A tibble: 16,598 x 11
##   Rank Name      Platform Year  Genre Publisher  NA_Sales EU_Sales JP_Sales
##   <dbl> <chr>      <chr>   <chr> <chr>   <chr>      <dbl>   <dbl>   <dbl>
## 1 166 Pokemon S~ N64      1999 Strat~ Nintendo    3.18    1.24    0.94
## 2 205 Warzone 2~ PS       1999 Strat~ Eidos Inte~  2.79    1.89    0
## 3 218 StarCraft~ PC       2010 Strat~ Activision  2.56    1.68    0
## 4 268 Warcraft ~ PC       1995 Strat~ Activision  1.7     2.27    0
## 5 336 Pokémon T~ GB       1998 Strat~ Nintendo    1.49    0.73    1.38
## 6 512 Command &~ PC       1996 Strat~ Virgin Int~  1.37    1.34    0
## 7 548 Pokémon S~ N64      2000 Strat~ Nintendo    1.02    0.36    1.13
## 8 579 Halo Wars X360     2009 Strat~ Microsoft ~  1.53    0.82    0.04
## 9 649 Theme Hos~ PC       1997 Strat~ Electronic~  2.3     0.1     0
## 10 807 Sim Theme~ PC       1998 Strat~ Electronic~  2.04    0.04    0
## # ... with 16,588 more rows, and 2 more variables: Other_Sales <dbl>,
## # Global_Sales <dbl>
```



## Rename column names in dataset

```
VideoGamedata %>%
```

```
  rename(
    NorthAmerica_Sales = NA_Sales,
    Europe_Sales = EU_Sales,
    Japan_Sales = JP_Sales,
    Other_Sales = Other_Sales,
    Global_Sales = Global_Sales
  )
```

```
## # A tibble: 16,598 x 11
```

```
##   Rank Name      Platform Year  Genre Publisher NorthAmerica_Sa~ Europe_Sales
##   <dbl> <chr>    <chr>    <chr> <chr> <chr>          <dbl>          <dbl>
## 1     1 1 Wii Spor~ Wii      2006 Sports Nintendo          41.5          29.0
## 2     2 2 Super Ma~ NES      1985 Platf~ Nintendo          29.1           3.58
## 3     3 3 Mario Ka~ Wii      2008 Racing Nintendo          15.8          12.9
## 4     4 4 Wii Spor~ Wii      2009 Sports Nintendo          15.8          11.0
## 5     5 5 Pokemon ~ GB      1996 Role~~ Nintendo          11.3           8.89
## 6     6 6 Tetris   GB      1989 Puzzle Nintendo          23.2           2.26
## 7     7 7 New Supe~ DS      2006 Platf~ Nintendo          11.4           9.23
## 8     8 8 Wii Play  Wii      2006 Misc  Nintendo          14.0           9.2
## 9     9 9 New Supe~ Wii      2009 Platf~ Nintendo          14.6           7.06
## 10    10 10 Duck Hunt NES      1984 Shoot~ Nintendo          26.9           0.63
## # ... with 16,588 more rows, and 3 more variables: Japan_Sales <dbl>,
## #   Other_Sales <dbl>, Global_Sales <dbl>
```

## Add new variables in data frame by using a mathematical function

```
VideoGamedata$NAAndEUSales = VideoGamedata$EU_Sales+VideoGamedata$NA_Sales
head (VideoGamedata, 15)
```

```
## # A tibble: 15 x 12
##   Rank Name      Platform Year  Genre Publisher NA_Sales EU_Sales JP_Sales
##   <dbl> <chr>      <chr>   <chr> <chr>   <chr>      <dbl>   <dbl>   <dbl>
## 1     1 1 Wii Sports   Wii    2006 Sports Nintendo   41.5    29.0    3.77
## 2     2 2 Super Mario~ NES    1985 Platf~ Nintendo   29.1     3.58    6.81
## 3     3 3 Mario Kart ~ Wii    2008 Racing Nintendo   15.8    12.9    3.79
## 4     4 4 Wii Sports ~ Wii    2009 Sports Nintendo   15.8    11.0    3.28
## 5     5 5 Pokemon Red~ GB     1996 Role~~ Nintendo   11.3     8.89   10.2
## 6     6 6 Tetris      GB     1989 Puzzle Nintendo   23.2     2.26    4.22
## 7     7 7 New Super M~ DS     2006 Platf~ Nintendo   11.4     9.23    6.5
## 8     8 8 Wii Play    Wii    2006 Misc  Nintendo   14.0     9.2    2.93
## 9     9 9 New Super M~ Wii    2009 Platf~ Nintendo   14.6     7.06    4.7
## 10    10 10 Duck Hunt   NES    1984 Shoot~ Nintendo   26.9     0.63    0.28
## 11    11 11 Nintendogs   DS     2005 Simul~ Nintendo    9.07    11     1.93
## 12    12 12 Mario Kart ~ DS     2005 Racing Nintendo    9.81    7.57    4.13
## 13    13 13 Pokemon Gol~ GB     1999 Role~~ Nintendo     9     6.18    7.2
## 14    14 14 Wii Fit     Wii    2007 Sports Nintendo    8.94    8.03    3.6
## 15    15 15 Wii Fit Plus Wii    2009 Sports Nintendo    9.09    8.59    2.53
## # ... with 3 more variables: Other_Sales <dbl>, Global_Sales <dbl>,
## #   NAAndEUSales <dbl>
```

## Create a training set using random number generator engine

```
set.seed(1234)
trainingdata = VideoGamedata %>% sample_frac(0.8, replace = FALSE)
head(trainingdata, 15)
```

```
## # A tibble: 15 x 12
##   Rank Name      Platform Year  Genre Publisher NA_Sales EU_Sales JP_Sales
##   <dbl> <chr>      <chr>   <chr> <chr>   <chr>      <dbl>   <dbl>   <dbl>
## 1  7453 Mortal Kom~ PS      2000  Fight~ Midway Ga~    0.12    0.08    0
## 2  8017 Reel Fishi~ PS      2000  Sports Victor In~    0.1     0.07    0
## 3  7163 Dark Souls~ XOne    2015  Role~~ Namco Ban~    0.13    0.07    0
## 4  8087 Battle Com~ SNES    1991  Strat~ Banpresto    0       0       0.18
## 5  9197 NFL Blitz ~ GC      2002  Sports Midway Ga~    0.11    0.03    0
## 6   623 Tomb Raide~ PS      1998  Action Eidos Int~    1.15    1.14    0.06
## 7 15243 Ryu-koku   PS2     2006  Adven~ KID          0       0       0.02
## 8 10886 Fishing Ma~ Wii     2009  Sports Hudson En~    0.09    0       0
## 9   935 GoldenEye ~ Wii     2010  Action Activision    0.85    0.71    0.13
## 10 12689 Catwoman  XB      2004  Action Electroni~    0.04    0.01    0
## 11  2949 Test Drive~ X360    N/A   Racing Atari       0.3     0.32    0
## 12 12239 Hasbro Fam~ X360    2011  Misc   Electroni~    0.06    0       0
## 13 15222 Legend: Ha~ PC      2007  Role~~ DTP Enter~    0       0.02    0
## 14 10425 Contact   DS      2006  Role~~ Rising St~    0.07    0       0.03
## 15  2623 Sid Meier'~ PS3     2008  Strat~ Take-Two ~    0.49    0.19    0.01
## # ... with 3 more variables: Other_Sales <dbl>, Global_Sales <dbl>,
## #   NAAndEUSales <dbl>
```

## Print Summary Statistics of dataset

```
summary(VideoGamedata)
```

```
##      Rank      Name      Platform      Year
## Min.   :    1  Length:16598  Length:16598  Length:16598
## 1st Qu.: 4151  Class :character  Class :character  Class :character
## Median : 8300  Mode  :character  Mode  :character  Mode  :character
## Mean   : 8301
## 3rd Qu.:12450
## Max.   :16600
##      Genre      Publisher      NA_Sales      EU_Sales
## Length:16598  Length:16598  Min.   : 0.0000  Min.   : 0.0000
## Class :character  Class :character  1st Qu.: 0.0000  1st Qu.: 0.0000
## Mode  :character  Mode  :character  Median : 0.0800  Median : 0.0200
##                                     Mean   : 0.2647  Mean   : 0.1467
##                                     3rd Qu.: 0.2400  3rd Qu.: 0.1100
##                                     Max.   :41.4900  Max.   :29.0200
##      JP_Sales      Other_Sales      Global_Sales      NAAndEUSales
## Min.   : 0.00000  Min.   : 0.00000  Min.   : 0.0100  Min.   : 0.0000
## 1st Qu.: 0.00000  1st Qu.: 0.00000  1st Qu.: 0.0600  1st Qu.: 0.0200
## Median : 0.00000  Median : 0.01000  Median : 0.1700  Median : 0.1200
## Mean   : 0.07778  Mean   : 0.04806  Mean   : 0.5374  Mean   : 0.4113
## 3rd Qu.: 0.04000  3rd Qu.: 0.04000  3rd Qu.: 0.4700  3rd Qu.: 0.3700
## Max.   :10.22000  Max.   :10.57000  Max.   :82.7400  Max.   :70.5100
```

---

Use numerical variables from the dataset and perform the statistical functions - mean

```
VideoGamedata %>%
  summarise(
    count = n(),
    mean_NorthAmericaSales = mean(VideoGamedata$NA_Sales, na.rm = TRUE),
    mean_EuropeSales = mean(VideoGamedata$EU_Sales, na.rm = TRUE),
    mean_JapanSales = mean(VideoGamedata$JP_Sales, na.rm = TRUE),
    mean_OtherSales = mean(VideoGamedata$Other_Sales, na.rm = TRUE),
    mean_GlobalSales = mean(VideoGamedata$Global_Sales, na.rm = TRUE)
  )

## # A tibble: 1 x 6
##   count mean_NorthAmericaSales mean_EuropeSales mean_JapanSales mean_OtherSales
##   <int>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 16598          0.265          0.147          0.0778         0.0481
## # ... with 1 more variable: mean_GlobalSales <dbl>
```

---

Use numerical variables from the dataset and perform the statistical functions - median

```
VideoGamedata %>%
  summarise(
    count = n(),
    median_NorthAmericaSales = median(VideoGamedata$NA_Sales, na.rm = TRUE),
    median_EuropeSales = median(VideoGamedata$EU_Sales, na.rm = TRUE),
    median_JapanSales = median(VideoGamedata$JP_Sales, na.rm = TRUE),
    median_OtherSales = median(VideoGamedata$Other_Sales, na.rm = TRUE),
    median_GlobalSales = median(VideoGamedata$Global_Sales, na.rm = TRUE)
  )

## # A tibble: 1 x 6
##   count median_NorthAmerica~ median_EuropeSal~ median_JapanSal~ median_OtherSal~
##   <int>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 16598          0.08          0.02          0          0.01
## # ... with 1 more variable: median_GlobalSales <dbl>
```

Use numerical variables from the dataset and perform the statistical functions - range

```
VideoGamedata %>%
  summarise(
    count = n(),
    mode_NorthAmericaSales = range(VideoGamedata$NA_Sales, na.rm = TRUE),
    mode_EuropeSales = range(VideoGamedata$EU_Sales, na.rm = TRUE),
    mode_JapanSales = range(VideoGamedata$JP_Sales, na.rm = TRUE),
    mode_OtherSales = range(VideoGamedata$Other_Sales, na.rm = TRUE),
    mode_GlobalSales = range(VideoGamedata$Global_Sales, na.rm = TRUE)
  )
```

```
## # A tibble: 2 x 6
##   count mode_NorthAmericaSales mode_EuropeSales mode_JapanSales mode_OtherSales
##   <int>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 16598            0            0            0            0
## 2 16598          41.5          29.0          10.2          10.6
## # ... with 1 more variable: mode_GlobalSales <dbl>
```

---

Find the correlation between any 2 variables by applying least square linear regression model - correlation

```
X = VideoGamedata[, "Global_Sales"]
Y = VideoGamedata[, "NA_Sales"]
CORRELATION = cor(Y,X,method = "pearson")
CORRELATION
```

```
##           Global_Sales
## NA_Sales    0.9410474
```

```
X = VideoGamedata[, "Global_Sales"]
Y = VideoGamedata[, "EU_Sales"]
CORRELATION = cor(Y,X,method = "pearson")
CORRELATION
```

```
##           Global_Sales
## EU_Sales    0.9028358
```

```
X = VideoGamedata[, "Global_Sales"]
Y = VideoGamedata[, "JP_Sales"]
CORRELATION = cor(Y,X,method = "pearson")
CORRELATION
```

```
##           Global_Sales
## JP_Sales    0.6118155
```

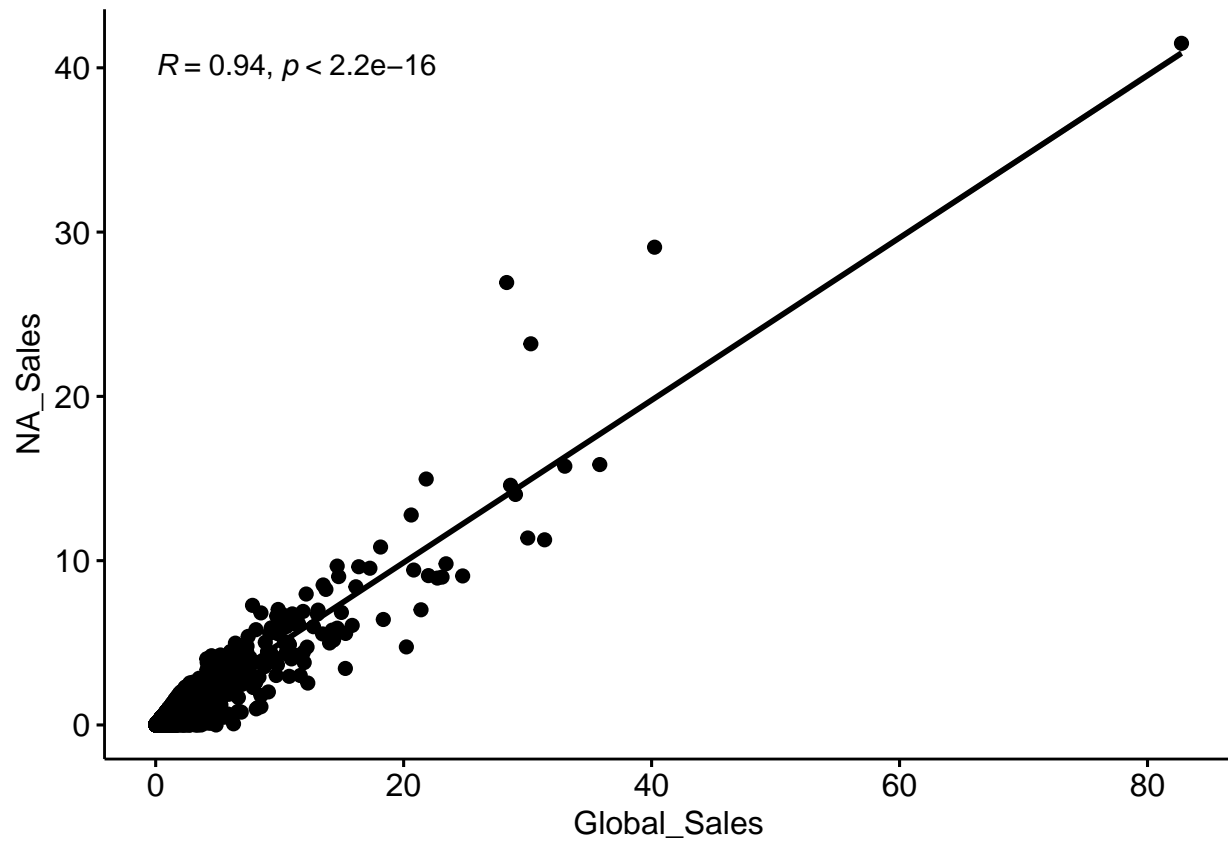
```
X = VideoGamedata[, "Global_Sales"]
Y = VideoGamedata[, "Other_Sales"]
CORRELATION = cor(Y,X,method = "pearson")
CORRELATION
```

```
##           Global_Sales
## Other_Sales  0.7483308
```



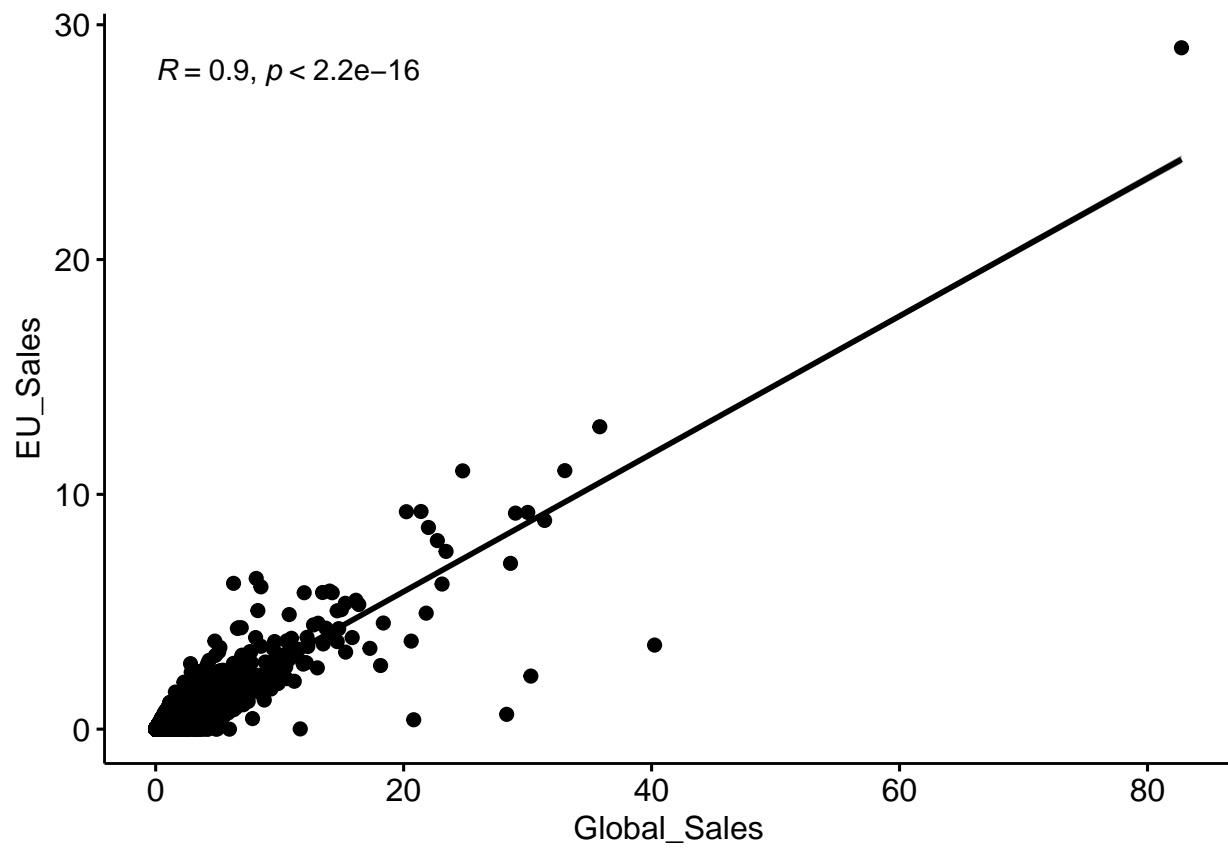
Plot a scatter plot for any 2 variables in your dataset - graph in correlation - North America / Global Sales

```
ggscatter(VideoGamedata, x = "Global_Sales", y = "NA_Sales", add = "reg.line", conf.int = TRUE, cor.coef = TRUE)
```



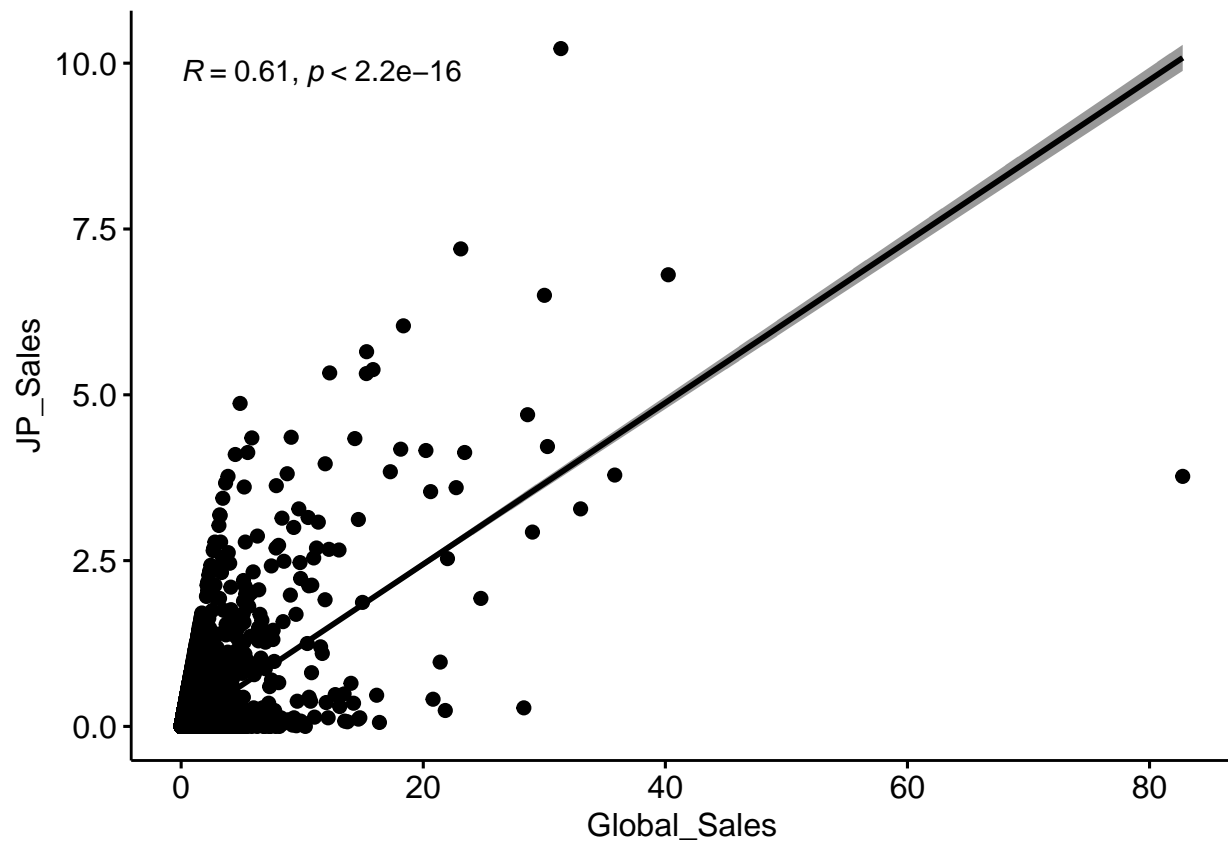
## - Europe / Global Sales

```
ggscatter(VideoGamedata, x = "Global_Sales", y = "EU_Sales", add = "reg.line", conf.int = TRUE, cor.coef = TRUE)
```



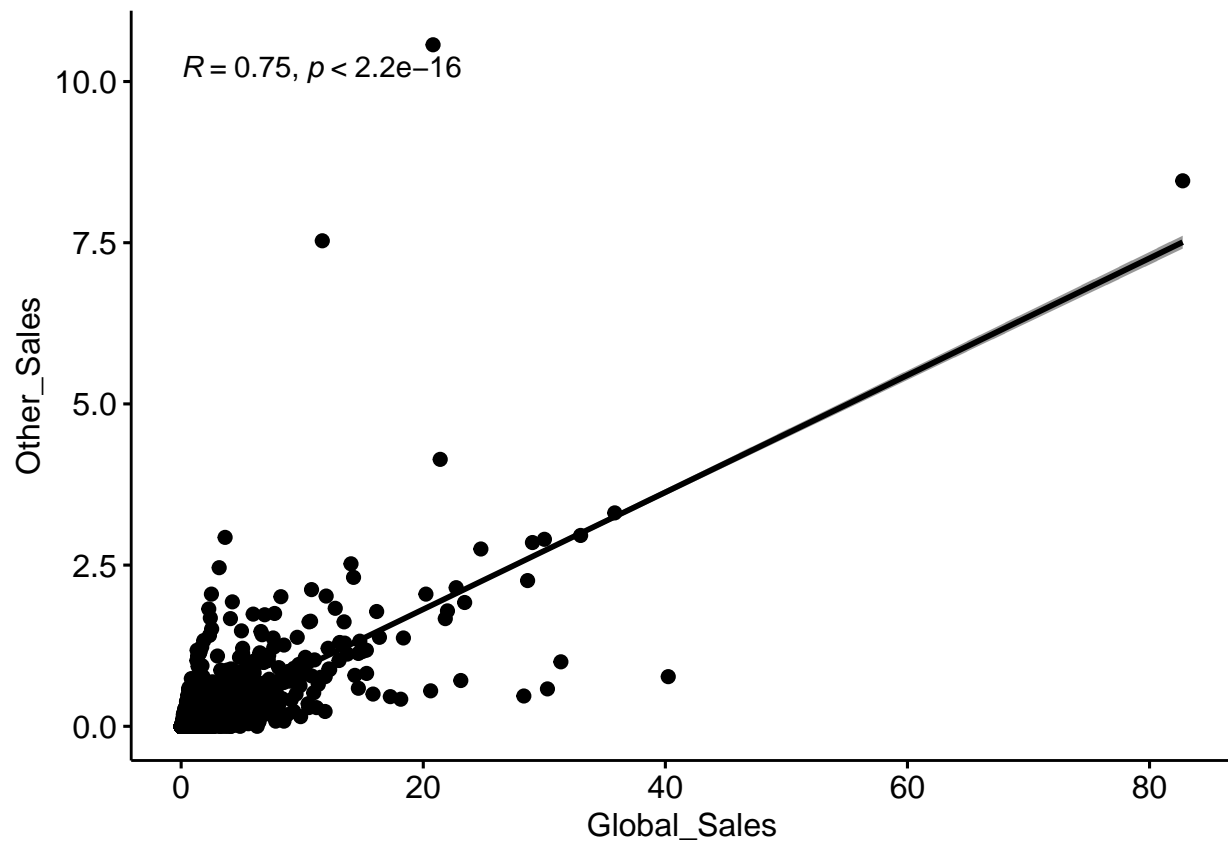
### - Japan / Global Sales

```
ggscatter(VideoGamedata, x = "Global_Sales", y = "JP_Sales", add = "reg.line", conf.int = TRUE, cor.coef = TRUE)
```



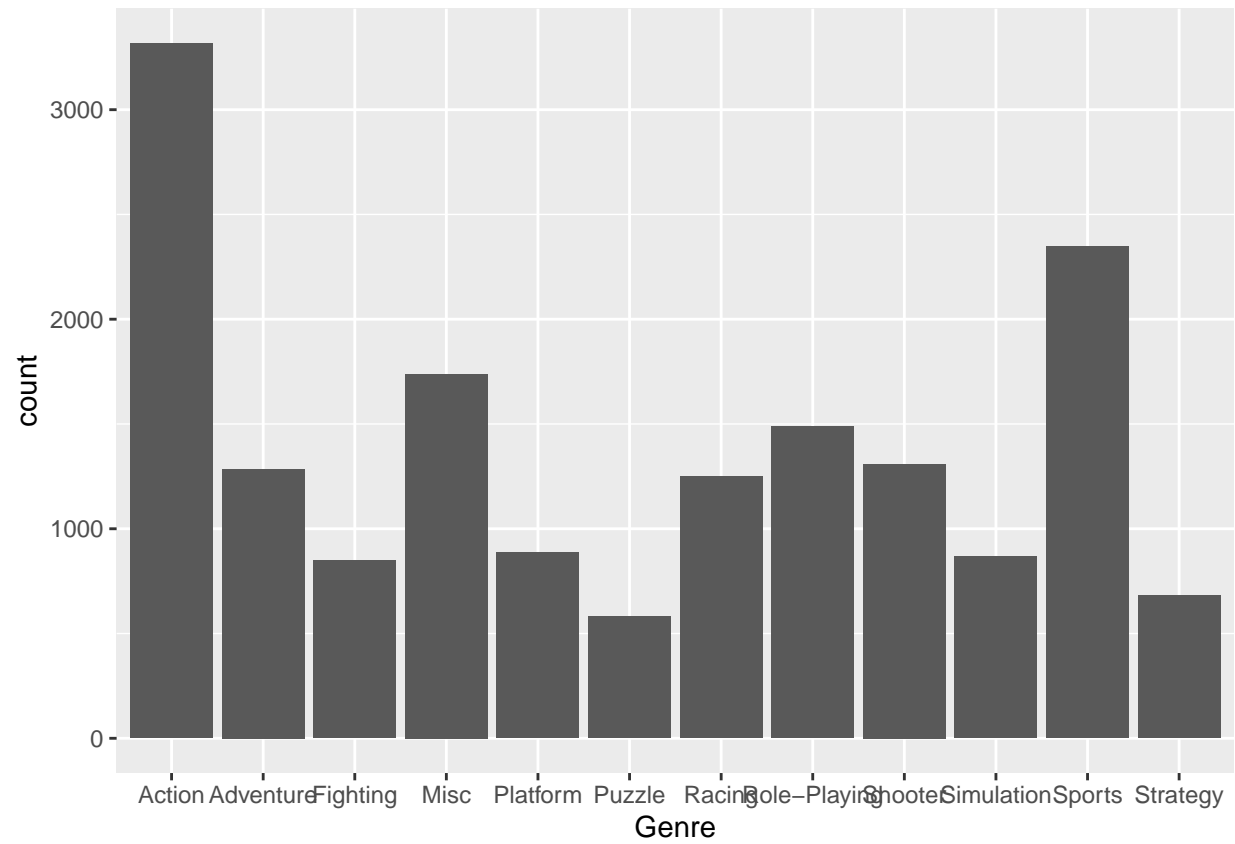
### - Other Countries / Global Sales

```
ggscatter(VideoGamedata, x = "Global_Sales", y = "Other_Sales", add = "reg.line", conf.int = TRUE, cor.co
```



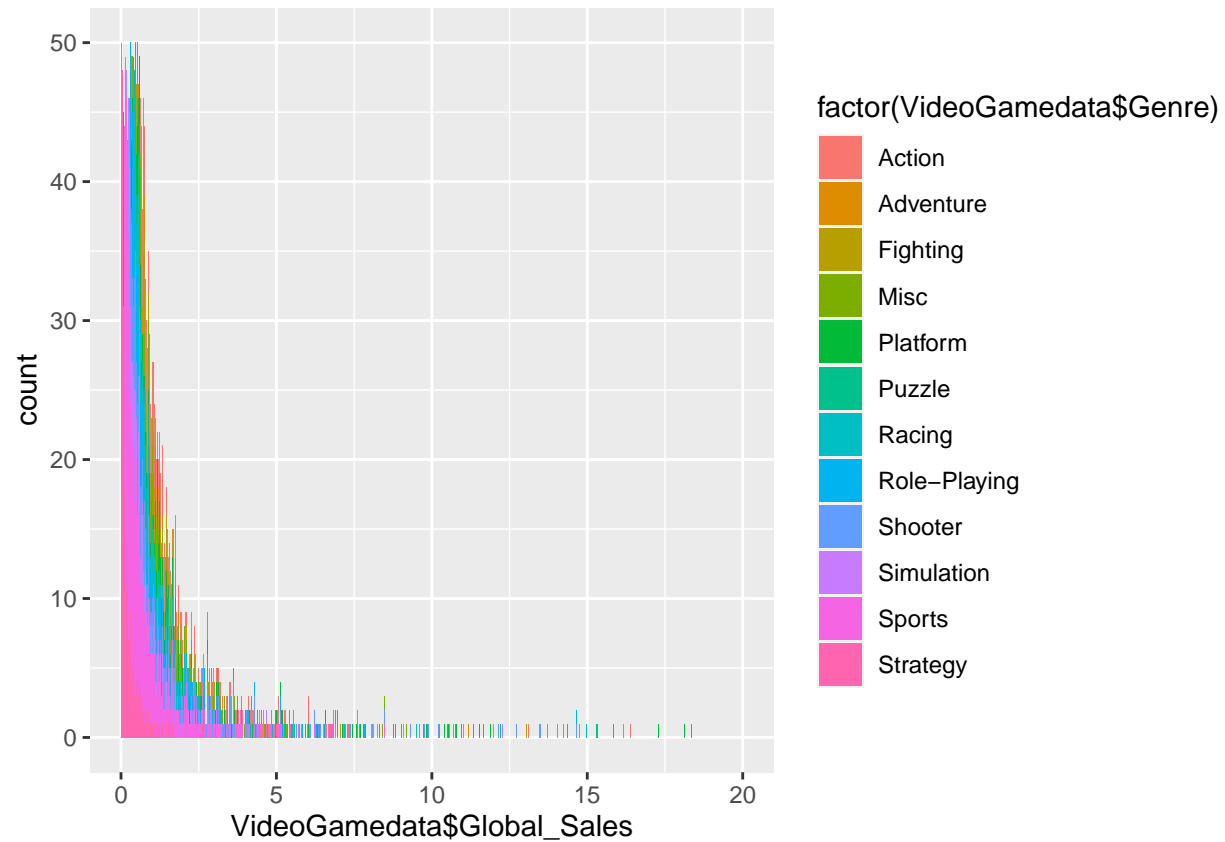
## graph - bar plot Global\_Sales

```
ggplot(data = VideoGamedata, aes(x = Genre)) + geom_bar()
```



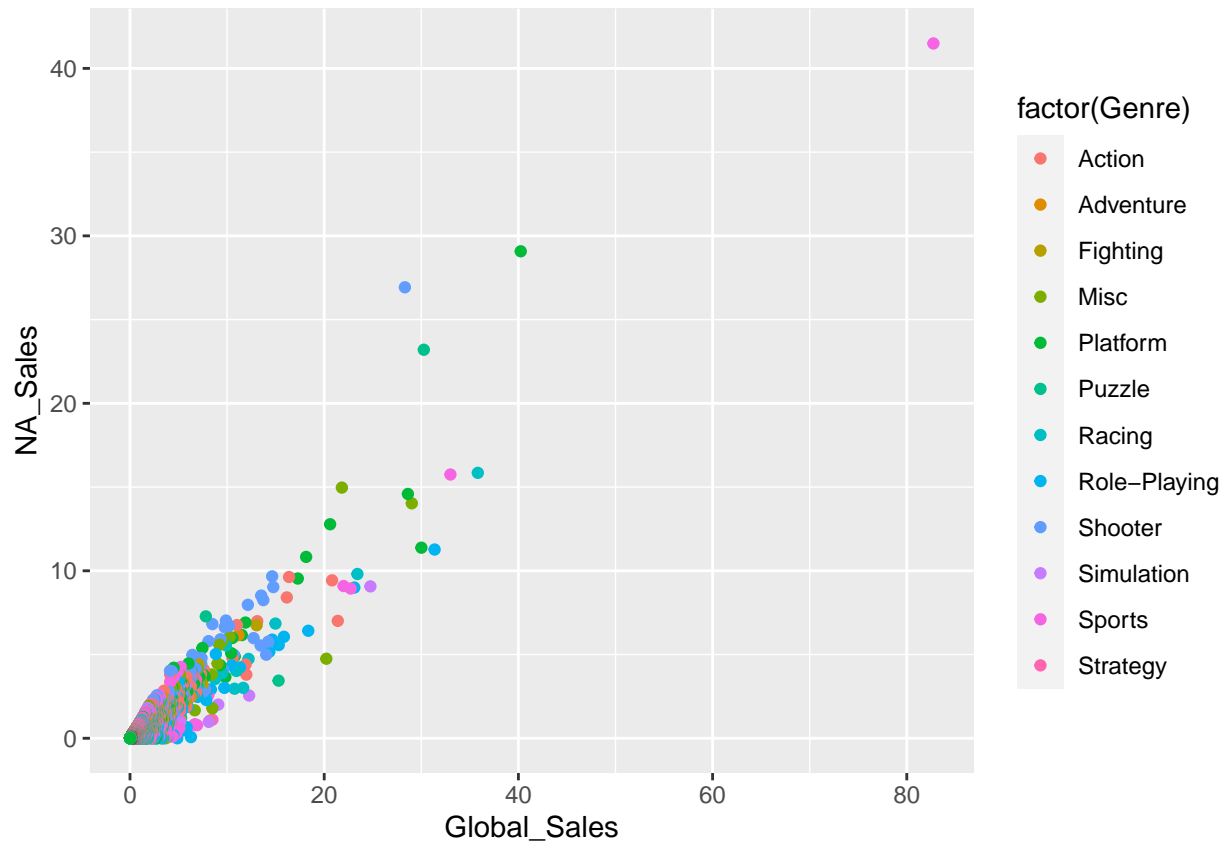
Plot a bar plot for any 2 variables in your dataset - barplot Global\_Sale with Genre

```
ggplot(data = VideoGamedata, aes(x = VideoGamedata$Global_Sales, fill = factor(VideoGamedata$Genre))) + g
```



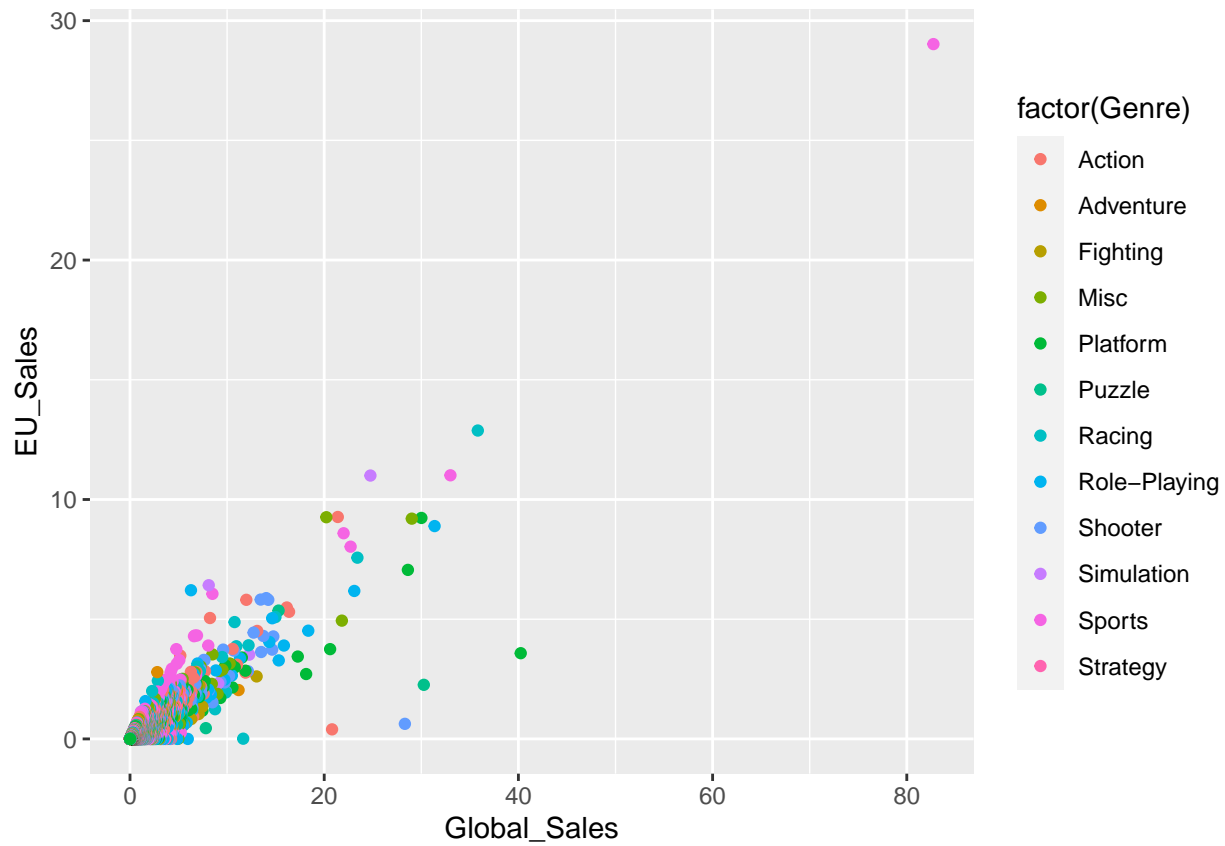
## Scatter plot - North America / Global Sales

```
ggplot(data = VideoGamedata,aes(x = Global_Sales,y = NA_Sales,col = factor(Genre)))+ geom_point()
```



## - Europe / Global Sales

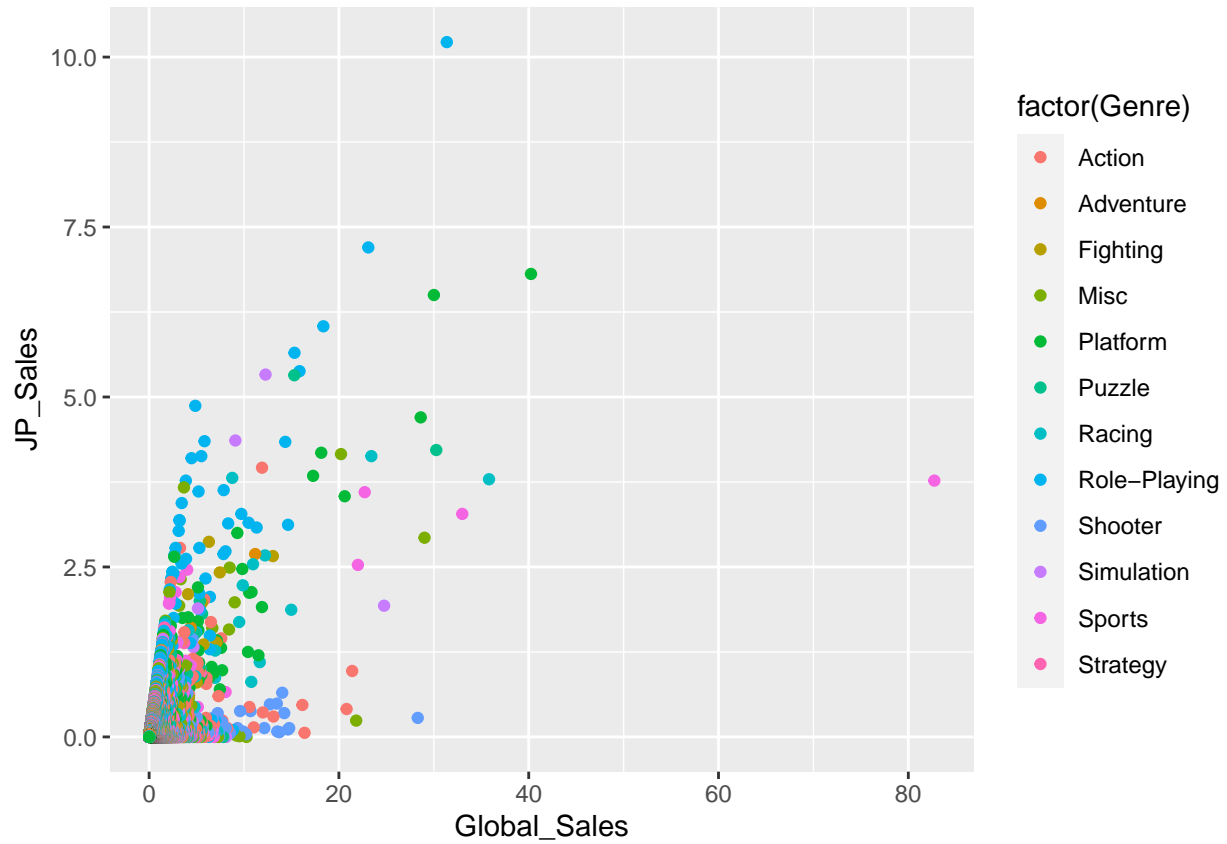
```
ggplot(data = VideoGamedata, aes(x = Global_Sales, y = EU_Sales, col = factor(Genre))) + geom_point()
```





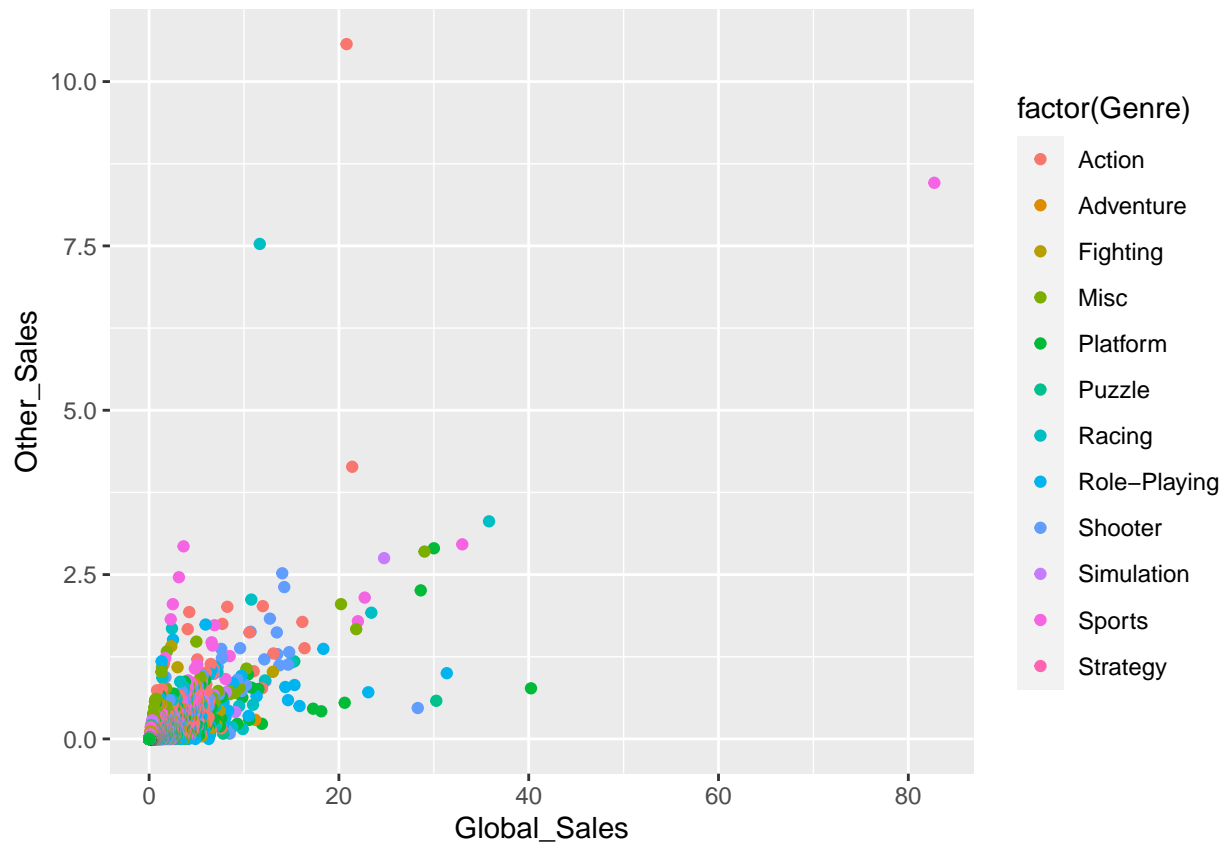
## - Japan / Global Sales

```
ggplot(data = VideoGamedata, aes(x = Global_Sales, y = JP_Sales, col = factor(Genre))) + geom_point()
```



## - Other Countries / Global Sales

```
ggplot(data = VideoGamedata, aes(x = Global_Sales, y = Other_Sales, col = factor(Genre))) + geom_point()
```



---

## Conclusion

Wii Sports game has the most sales overall, and specifically for North America and Europe.

Pokemon Red/Pokemon Blue has the most sales in Japan and Grand Theft Auto: San Andreas in Other countries.

PS2 Platform is leading overall video game sales.

Between the year of 2008 and 2010 video game sales are the highest while it dramatically dropped between 2017 and 2020.

Action genre has the most recorded sales around the world.