Fostanes Final Project

Christian Fostanes

6/15/2021

Big Question - Does Video Game Genre Affect Sales in Different Regions?

```
# Reading the data and convert some columns.
vgs.dat <- read.csv("vgsales.csv")
vgs.dat$Platform <- as.factor(vgs.dat$Platform)
vgs.dat$Genre <- as.factor(vgs.dat$Genre)
vgs.dat$Publisher <- as.factor(vgs.dat$Publisher)
vgs.dat$Year <- as.factor(vgs.dat$Year)

# Checks if there are any missing values
sum(is.na(vgs.dat))
```

```
## [1] 0
```

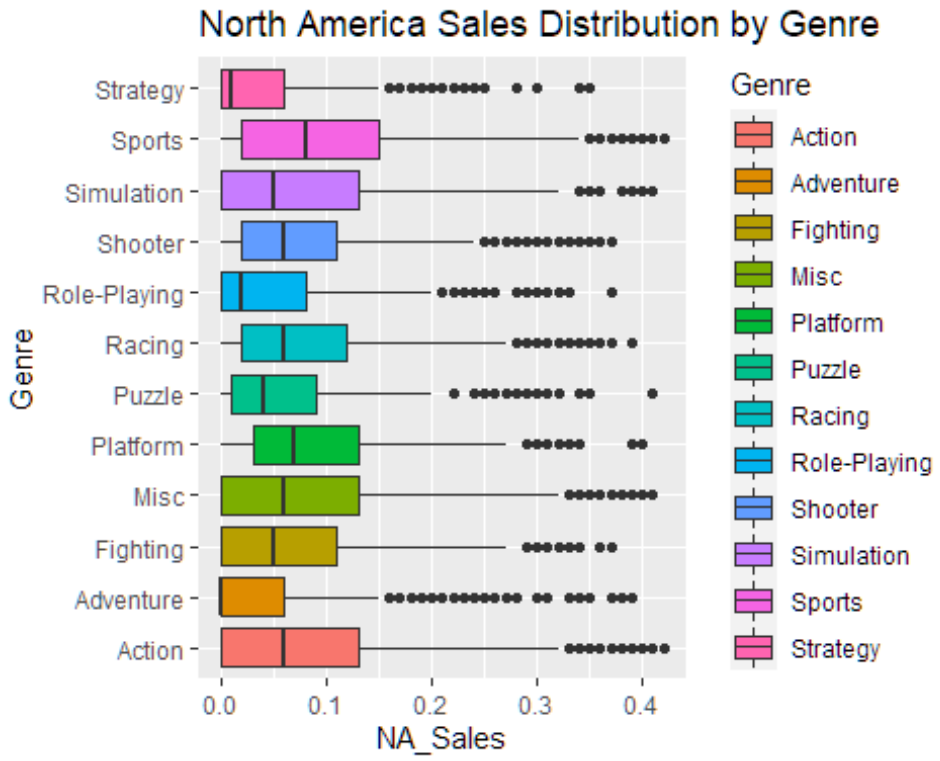
```
# Remove outliers
# NA
outliers <- boxplot(vgs.dat$NA_Sales, plot = FALSE)$out
vgs.dat <- vgs.dat[-c(which(vgs.dat$NA_Sales %in% outliers)),]
# EU
outliers <- boxplot(vgs.dat$EU_Sales, plot = FALSE)$out
vgs.dat <- vgs.dat[-c(which(vgs.dat$EU_Sales %in% outliers)),]
# JP
outliers <- boxplot(vgs.dat$JP_Sales, plot = FALSE)$out
vgs.dat <- vgs.dat[-c(which(vgs.dat$JP_Sales %in% outliers)),]
# Other Regions
outliers <- boxplot(vgs.dat$Other_Sales, plot = FALSE)$out
vgs.dat <- vgs.dat[-c(which(vgs.dat$Other_Sales %in% outliers)),]
# Global
outliers <- boxplot(vgs.dat$Global_Sales, plot = FALSE)$out
vgs.dat <- vgs.dat[-c(which(vgs.dat$Global_Sales %in% outliers)),]

summary(vgs.dat)
```

```
##       Rank           Name              Platform         Year
##  Min.   : 4416   Length:10240       DS     :1572   2009   : 970
##  1st Qu.: 8130   Class :character   PS2    :1205   2008   : 919
##  Median :11389   Mode  :character   Wii    : 872   2010   : 848
##  Mean   :11052                      PSP    : 798   2011   : 759
##  3rd Qu.:14039                      PC     : 783   2007   : 755
##  Max.   :16600                      X360   : 750   2006   : 700
##                                     (Other):4260   (Other):5289
##        Genre                            Publisher       NA_Sales
##  Action   :2046   Ubisoft                   : 617   Min.   :0.00000
```

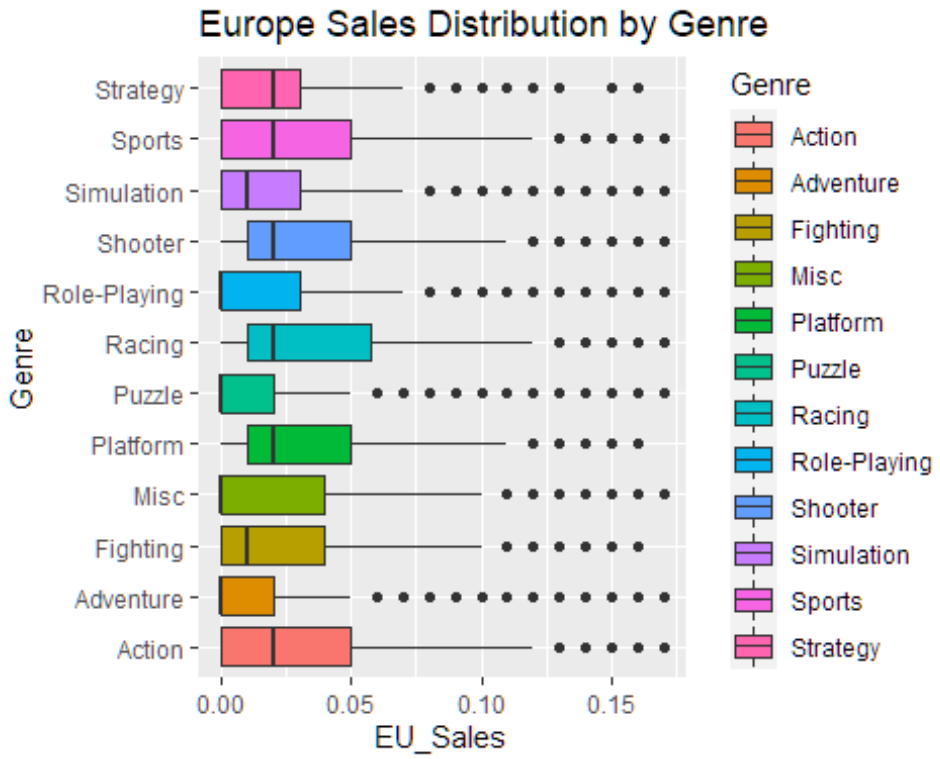
```
##  Sports   :1335   Activision                    : 588   1st Qu.:0.00000
##  Misc     :1152   Electronic Arts               : 587   Median :0.05000
##  Adventure:1038   Namco Bandai Games            : 462   Mean   :0.07515
##  Racing   : 822   THQ                           : 459   3rd Qu.:0.12000
##  Shooter  : 753   Konami Digital Entertainment: 455   Max.   :0.42000
##  (Other)  :3094   (Other)                       :7072
##     EU_Sales          JP_Sales         Other_Sales        Global_Sales
##  Min.   :0.0000   Min.   :0.000000   Min.   :0.000000   Min.   :0.0100
##  1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0400
##  Median :0.0100   Median :0.000000   Median :0.010000   Median :0.0800
##  Mean   :0.0283   Mean   :0.007873   Mean   :0.008749   Mean   :0.1205
##  3rd Qu.:0.0400   3rd Qu.:0.000000   3rd Qu.:0.010000   3rd Qu.:0.1800
##  Max.   :0.1700   Max.   :0.070000   Max.   :0.050000   Max.   :0.4400
##
```

Plots

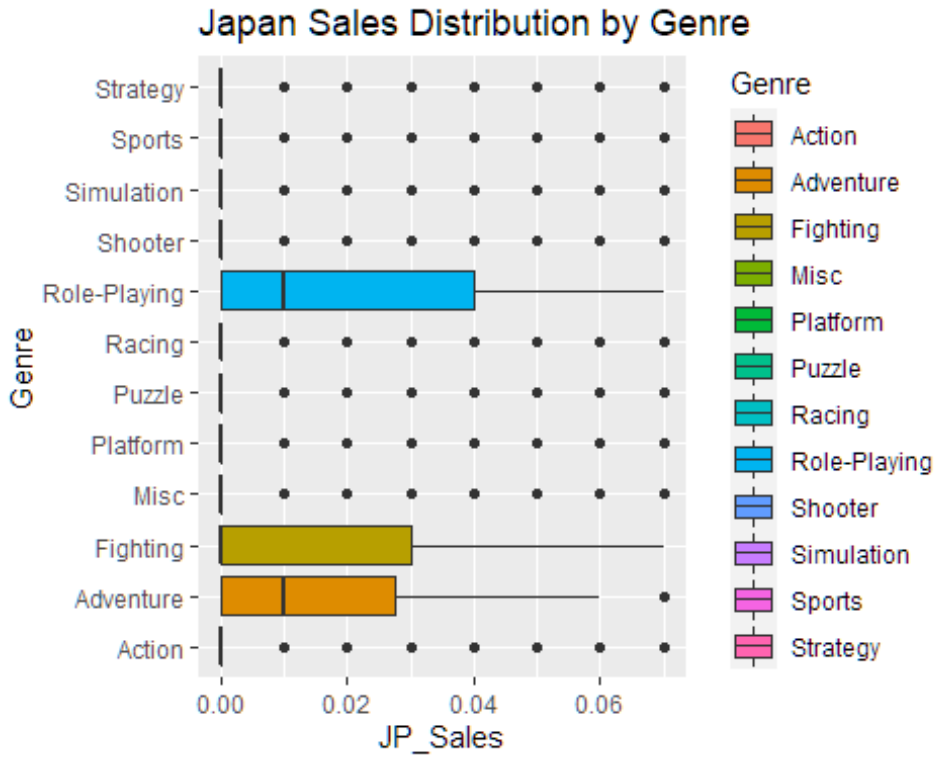
```
# Box Plots for Different Regions
ggplot(vgs.dat, aes(NA_Sales, Genre, fill = Genre)) + geom_boxplot() + ggtitl
e("North America Sales Distribution by Genre")
```



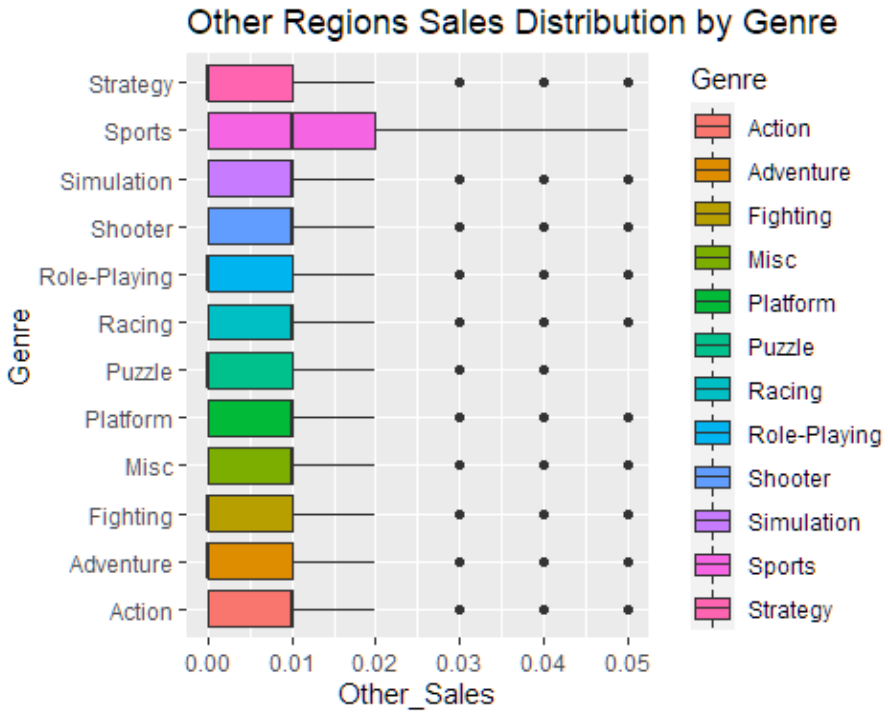

```
ggplot(vgs.dat, aes(EU_Sales, Genre, fill = Genre)) + geom_boxplot() + ggtitl
e("Europe Sales Distribution by Genre")
```



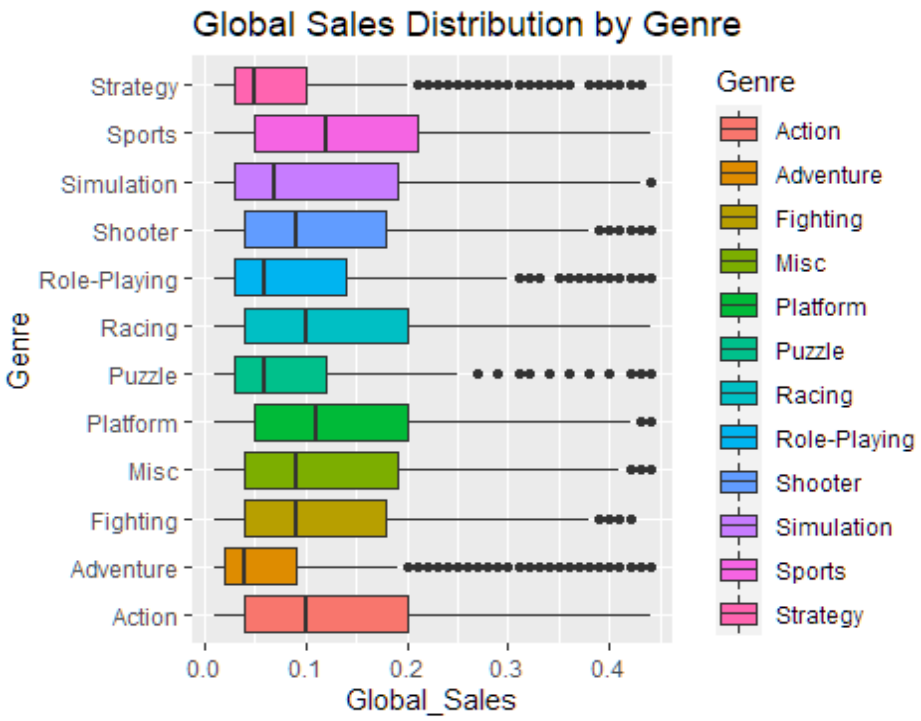

```
ggplot(vgs.dat, aes(JP_Sales, Genre, fill = Genre)) + geom_boxplot() + ggtitl
e("Japan Sales Distribution by Genre")
```



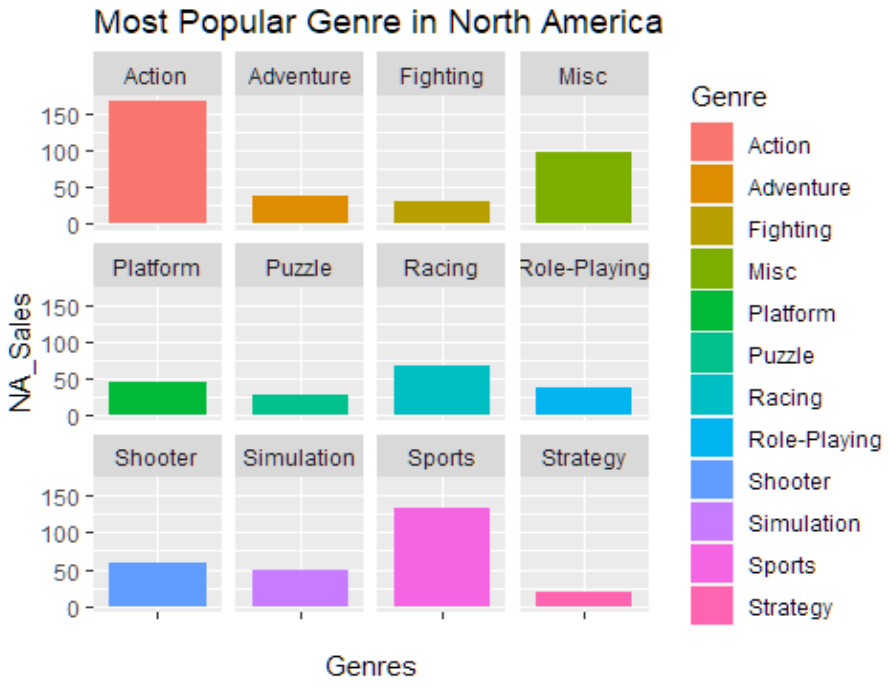
```
ggplot(vgs.dat, aes(Other_Sales, Genre, fill = Genre)) + geom_boxplot() + ggt
itle("Other Regions Sales Distribution by Genre")
```



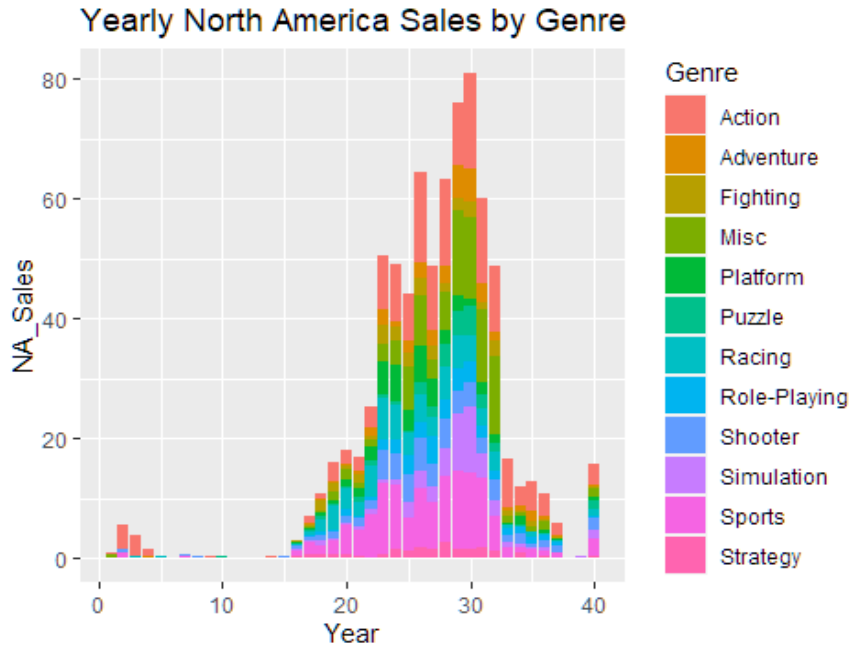

```
ggplot(vgs.dat, aes(Global_Sales, Genre, fill = Genre)) + geom_boxplot() + gg
title("Global Sales Distribution by Genre")
```



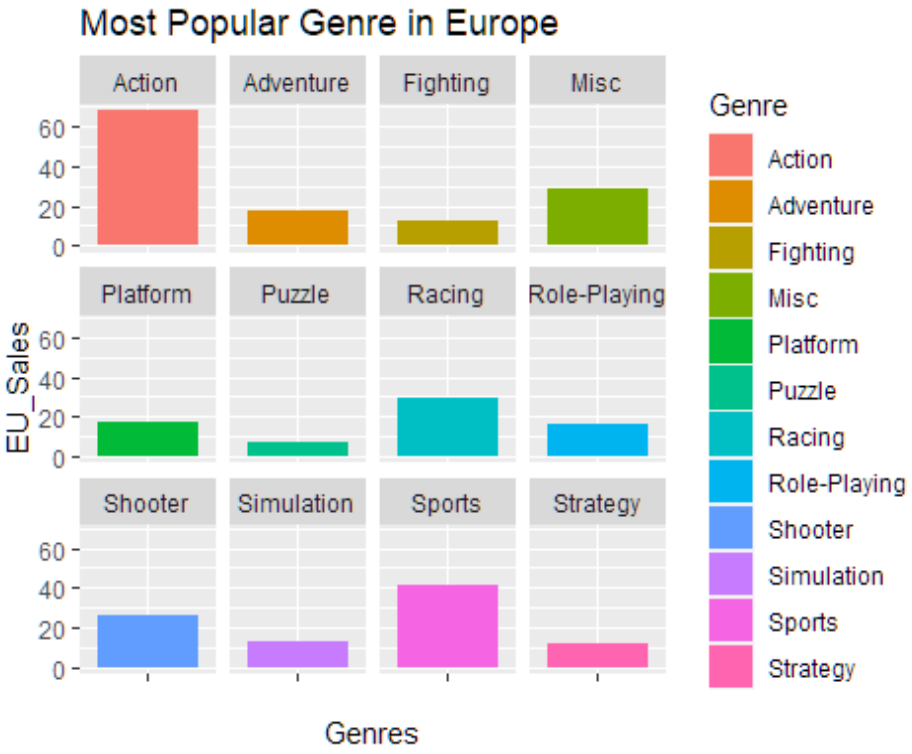
```
vgs.dat$Year <- as.numeric(vgs.dat$Year)

# NA Plots
ggplot(vgs.dat, aes("", NA_Sales, fill = Genre)) + geom_bar(stat = "identity"
) + ggtitle("North America Sales by Genre") + facet_wrap(~Genre) + xlab("Genr
es") + ggtitle("Most Popular Genre in North America")
```



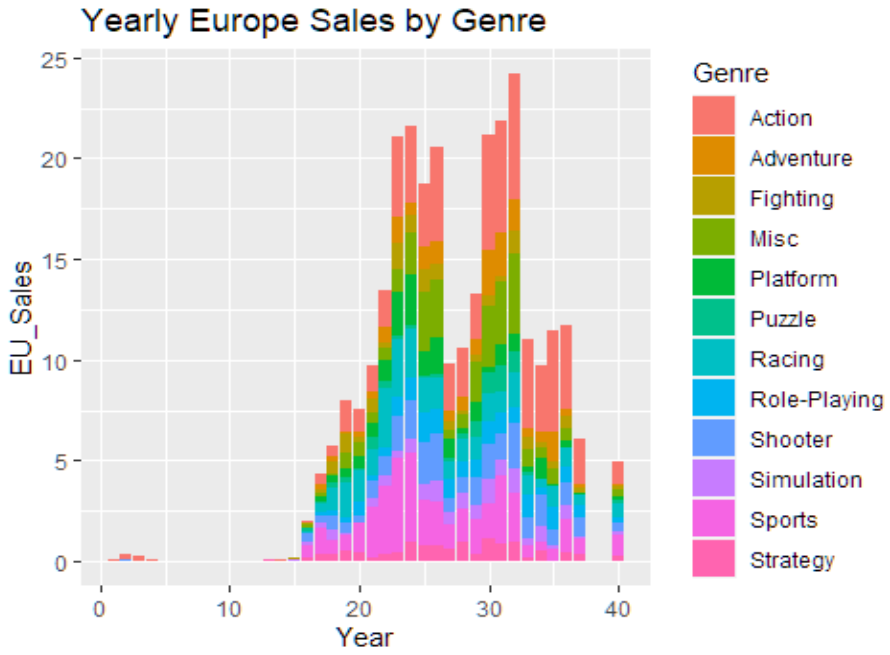

```
ggplot(vgs.dat, aes(Year, NA_Sales, fill = Genre)) + geom_bar(stat = "identit
y") + ggtitle("Yearly North America Sales by Genre")
```



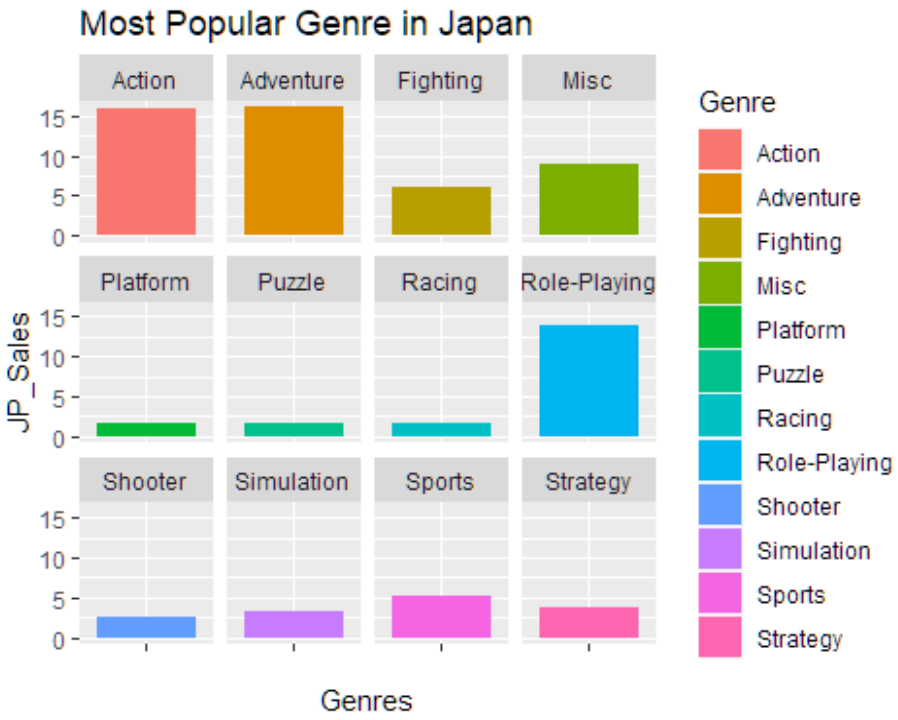
```
# EU Plots
ggplot(vgs.dat, aes("", EU_Sales, fill = Genre)) + geom_bar(stat = "identity"
) + ggtitle("Europe Sales by Genre") + facet_wrap(~Genre) + xlab("Genres") +
ggtitle("Most Popular Genre in Europe")
```



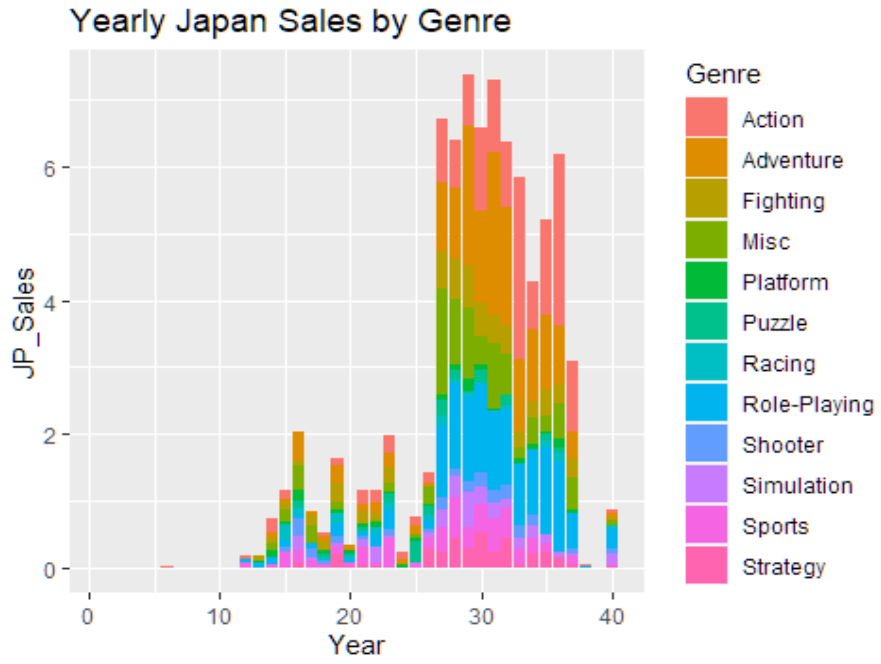

```
ggplot(vgs.dat, aes(Year, EU_Sales, fill = Genre)) + geom_bar(stat = "identit
y") + ggtitle("Yearly Europe Sales by Genre")
```



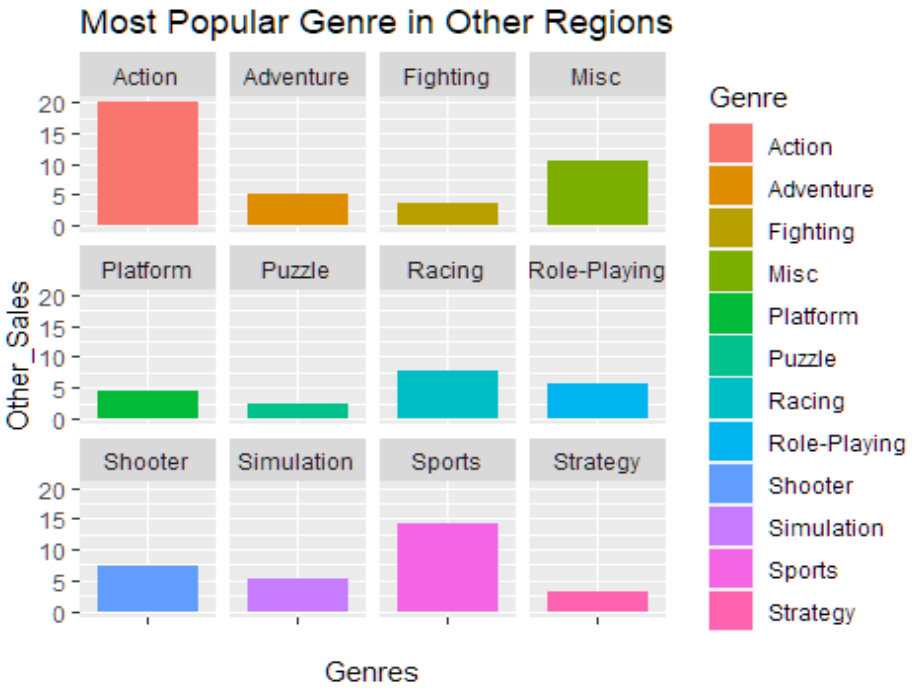
```
# JP Plots
ggplot(vgs.dat, aes("", JP_Sales, fill = Genre)) + geom_bar(stat = "identity"
) + ggtitle("Japan Sales by Genre") + facet_wrap(~Genre) + xlab("Genres") + g
gtitle("Most Popular Genre in Japan")
```



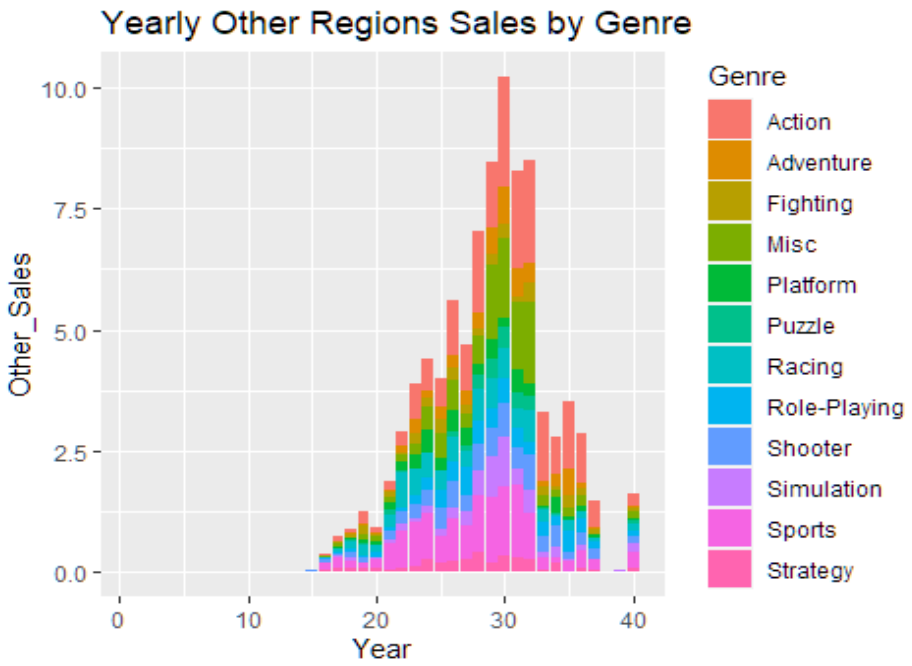

```
ggplot(vgs.dat, aes(Year, JP_Sales, fill = Genre)) + geom_bar(stat = "identit
y") + ggtitle("Yearly Japan Sales by Genre")
```



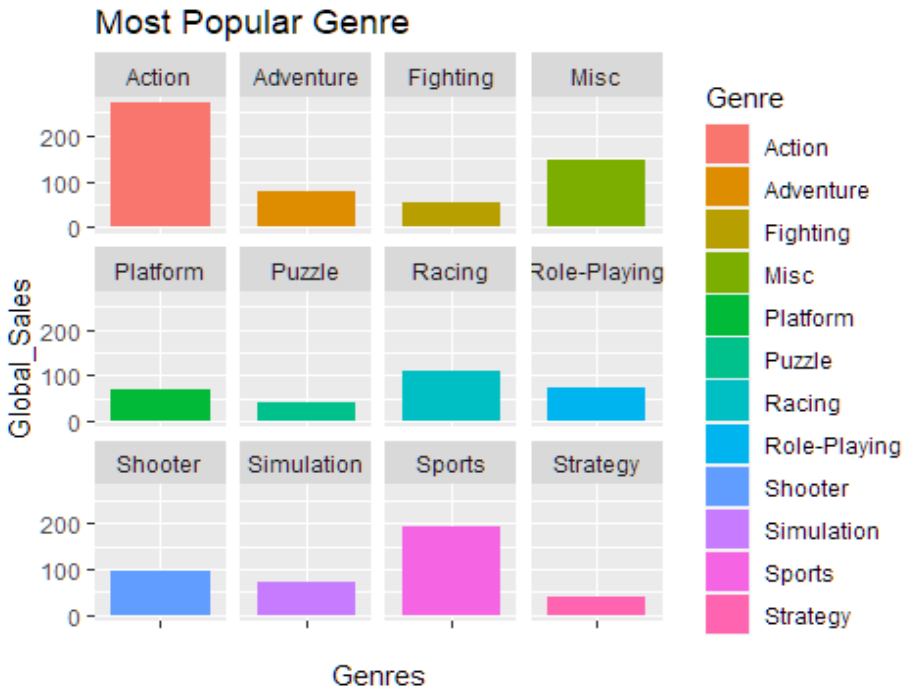
```
# Other Regions Plots
ggplot(vgs.dat, aes("", Other_Sales, fill = Genre)) + geom_bar(stat = "identi
ty") + ggtitle("Other Regions Sales by Genre") + facet_wrap(~Genre) + xlab("G
enres") + ggtitle("Most Popular Genre in Other Regions")
```



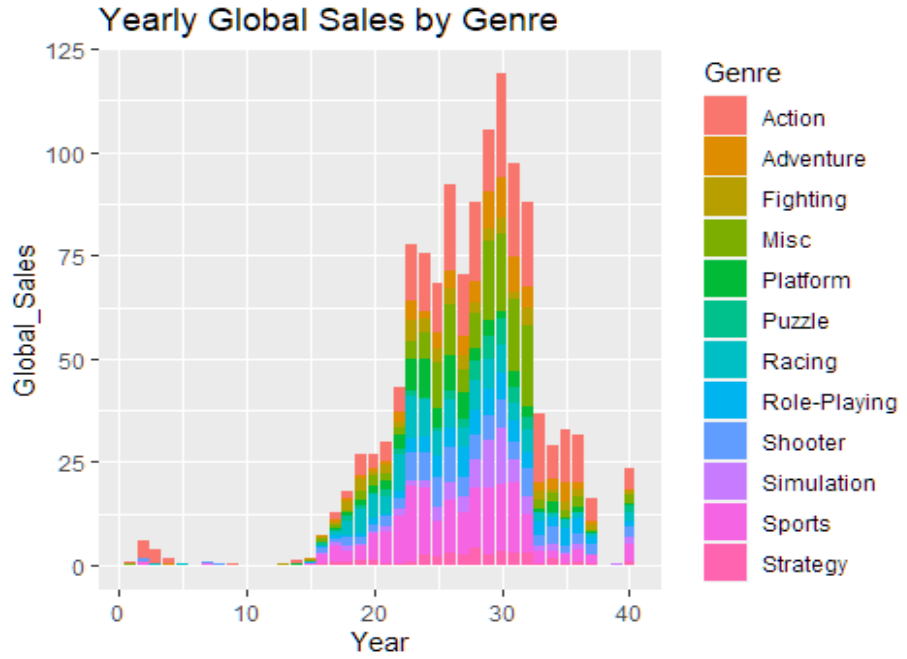

```
ggplot(vgs.dat, aes(Year, Other_Sales, fill = Genre)) + geom_bar(stat = "iden
tity") + ggtitle("Yearly Other Regions Sales by Genre")
```



```
# Global Plots
ggplot(vgs.dat, aes("", Global_Sales, fill = Genre)) + geom_bar(stat = "ident
ity") + ggtitle("Global Sales by Genre") + facet_wrap(~Genre) + xlab("Genres"
) + ggtitle("Most Popular Genre")
```




```
ggplot(vgs.dat, aes(Year, Global_Sales, fill = Genre)) + geom_bar(stat = "ide
ntity") + ggtitle("Yearly Global Sales by Genre")
```



```
vgs.dat <- glm (as.factor(Genre) ~ NA_Sales + EU_Sales + JP_Sales + Other_Sal
es + Global_Sales, data = vgs.dat, family = binomial)
summary(vgs.dat)

##
## Call:
## glm(formula = as.factor(Genre) ~ NA_Sales + EU_Sales + JP_Sales +
##     Other_Sales + Global_Sales, family = binomial, data = vgs.dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9138   0.6186   0.6371   0.6691   0.8964
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.57960    0.04224  37.392   <2e-16 ***
## NA_Sales      4.02413    4.89588   0.822    0.411
## EU_Sales      0.14846    4.99221   0.030    0.976
## JP_Sales      1.86187    5.01300   0.371    0.710
## Other_Sales   8.33467    5.44215   1.532    0.126
## Global_Sales -4.79363    4.87184  -0.984    0.325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 10243  on 10239  degrees of freedom
## Residual deviance: 10192  on 10234  degrees of freedom
## AIC: 10204
##
## Number of Fisher Scoring iterations: 4
```

Report

The data set I chose depicts different video game titles released from 1980 to 2020 and their sales in different regions. I chose this data set because I enjoy playing video games as a hobby and I was curious if certain video game genres are more popular in certain regions than others. This data set is perfect fit for answering that question.

Before plotting the data, I first checked if the data set has any missing values. Next, I checked for outliers and removed those from the data set.

A bar graph was the appropriate type of plot because I wanted to know which genres are the most popular in different regions. My premilinary conclusion was genre does have an effect on sales in different regions.

Based on the data plots, here are my findings:

- Most popular genre in North America and Europe is Action, while the least popular is Strategy

- Most popular genre in Japan is Role-Playing, least popular is Shooter
- As for other regions, it has the same results as North America and Europe. The most popular genre is Action and the least popular is Strategy.
- Overall, the most popular video game genre is Action, while the least popular is Strategy
- Sales peaked during the period of 2005-2010

Based on my findings, I can say that there is a connection between video game genre and sales in different regions. Western regions such as North America and Europe prefer the Action video game genre and Japanese gamers prefer the Role-Playing genre. Overall, the most popular video game genre in the world is action, while the least popular is Strategy. While it's not definitive, I believe cultures play a part when it comes to genre popularity. Shooter games are the least popular genre in Japan. Shooter games are often violent and Japanese people are not a fan of violence. If you look at popular games that cam from Japan, they often aren't violent and are family friendly. Western media in general is very action packed. Most popular movies and shows from Western regions fall under the Action genre, so it's no surprise that most popular video game genre is also Action. The peak of overall sales during the 2005-2010 period is due to the number of video game titles releases during that period.