Data project

Grant Mooslin

9/23/2024

I want to use R to learn about the relationship of video game genres and their sales to see what we can find based off the data I have. What I will begin with is reading and importing the data and putting it into the "vgsales" variable.

```r
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
```

```r
library(ggplot2)
vgsales <- read.csv("C:/math130/data/vgsales.csv")
```

First I want to specify the factors of the data by establishing the genres of the different games there are.

```r
factor(vgsales$Genre, levels = c("Shooter", "Racing", "Platform", "Puzzle", "Sports", "Misc", "Role-Play
```

```
## .
##      Shooter       Racing     Platform       Puzzle       Sports         Misc
##         1310         1249          886          582         2346         1739
## Role-Playing       Action     Fighting
##         1488         3316          848
```

So we are looking at the sample sizes and I noticed that the top 3 are action, sports, and misc. Now let's separate the data by genre.

```r
shooter <- filter(vgsales, Genre =="Shooter")
action <- filter(vgsales, Genre =="Action")
roleplaying <- filter(vgsales, Genre =="Role-Playing")
puzzle <- filter(vgsales, Genre =="Puzzle")
platform <- filter(vgsales, Genre =="Platform")
racing <- filter(vgsales, Genre =="Racing")
misc <- filter(vgsales, Genre =="Misc")
sports <- filter(vgsales, Genre =="Sports")
fighting <- filter(vgsales, Genre =="Fighting")
```

Now we have all the games separated by genre we can now start extracting the means of the sales by genre. Since the data for the video game sales is in millions we will multiply the result by 1 million to get the full number of average sales

```
genre_mean_sales <- vgsales %>%
  filter(Genre %in% c("Shooter", "Racing", "Platform", "Puzzle", "Sports", "Misc", "Role-Playing", "Act:
  summarise(mean_global_sales = mean(Global_Sales, na.rm = TRUE)* 1e6)

print(genre_mean_sales)
```

```
## # A tibble: 9 x 2
##   Genre        mean_global_sales
##   <chr>                    <dbl>
## 1 Action                 528100.
## 2 Fighting               529375
## 3 Misc                   465762.
## 4 Platform               938341.
## 5 Puzzle                 420876.
## 6 Racing                 586101.
## 7 Role-Playing           623233.
## 8 Shooter                791885.
## 9 Sports                 567319.
```

After looking at the data we can see the platform games have had the biggest sales however lets put this on a bar chart to help visualize how big of a difference it is.

```
genre_mean_sales <- vgsales %>%
  filter(Genre %in% c("Shooter", "Racing", "Platform", "Puzzle", "Sports", "Misc", "Role-Playing", "Act:
  group_by(Genre) %>%
  summarise(mean_global_sales = mean(Global_Sales, na.rm = TRUE) * 1e6)

# Create a bar chart
ggplot(genre_mean_sales, aes(x = reorder(Genre, -mean_global_sales), y = mean_global_sales)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Mean Global Sales by Genre",
       x = "Genre",
       y = "Mean Global Sales") +
  theme_minimal() +
  coord_flip()
```

Mean Global Sales by Genre
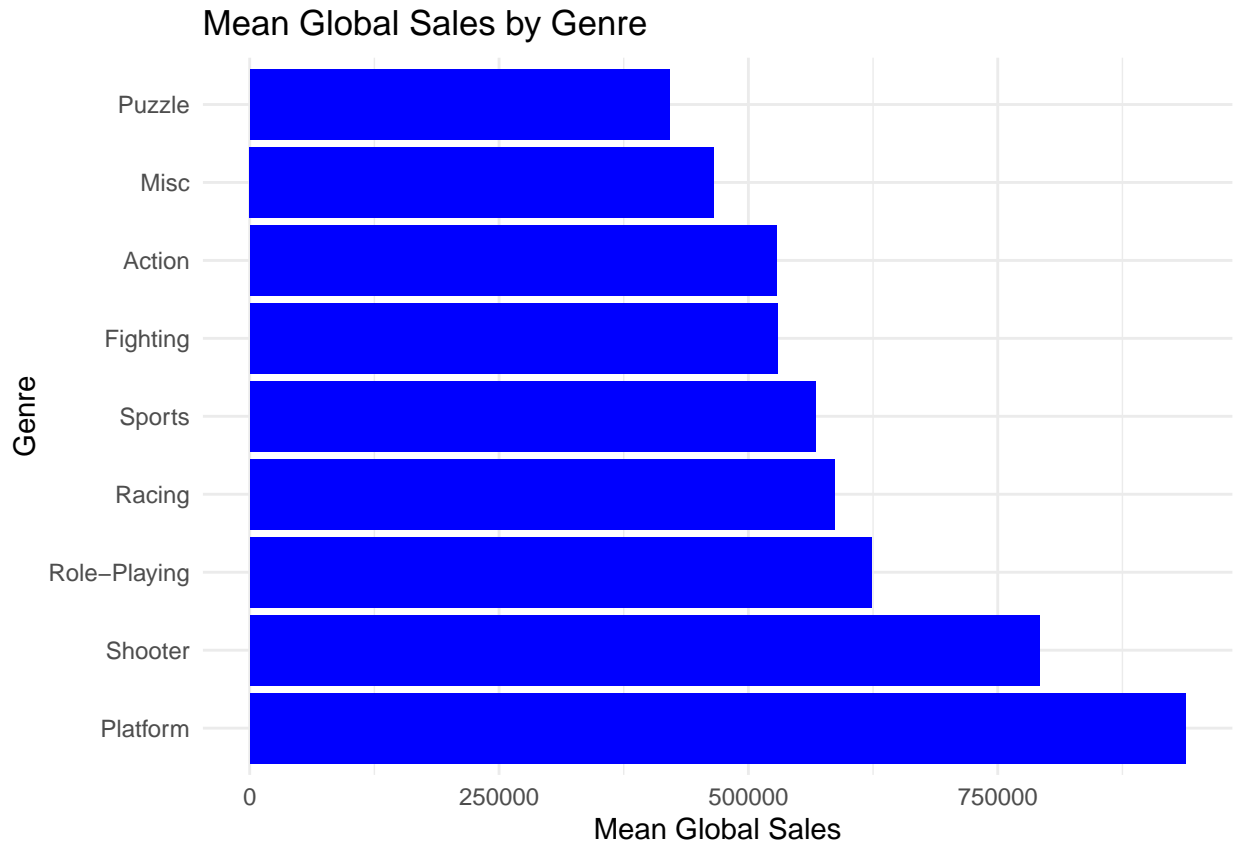
Now we can see that through this bar char that platforms are big but also shooters as well have very high global sales within this data set. Lets take a closer look at both the shooter and platform genres and see how they look on box plot to help visualize these genres.

```
ggplot(vgsales, aes(x = Genre, y = Global_Sales)) +
  geom_boxplot(fill = "blue") +
  labs(title = "Box Plot of Global Sales by Genre",
       x = "Genre",
       y = "Global Sales") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))
```

Box Plot of Global Sales by Genre
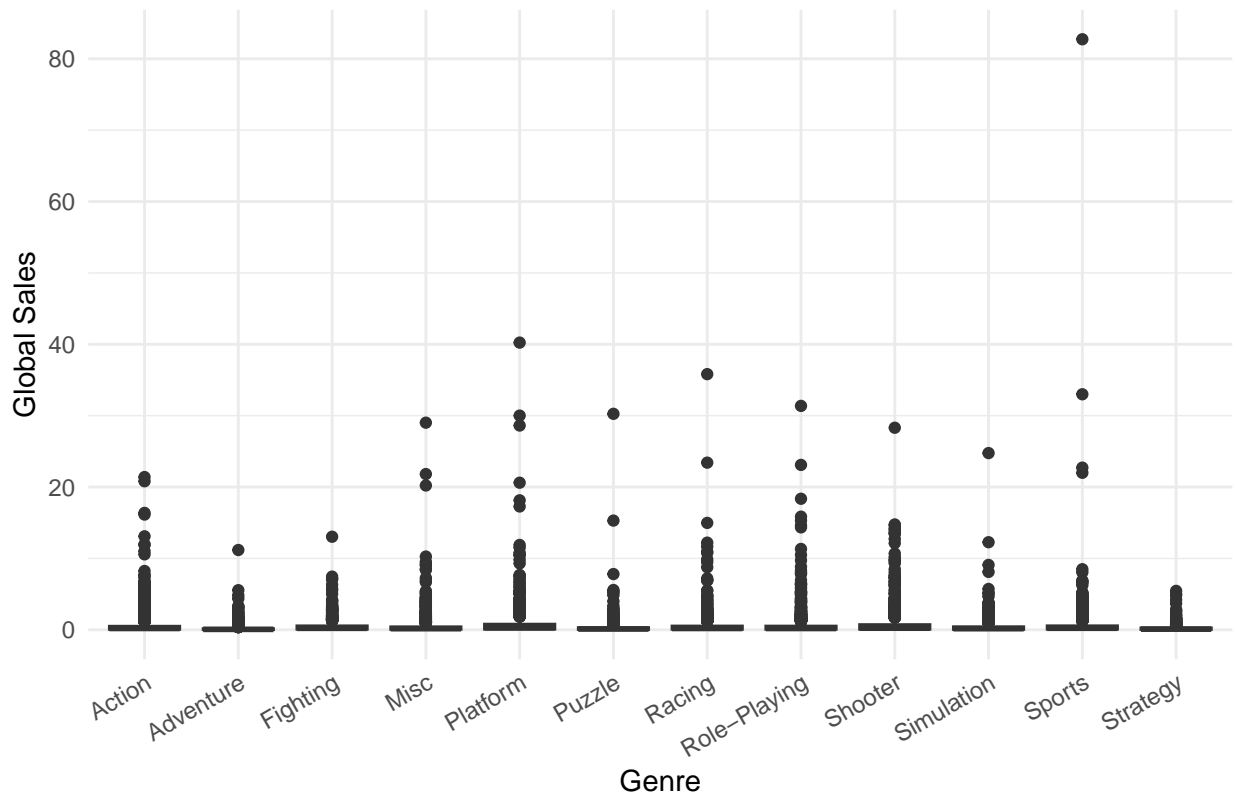
This box plot reveals that within the global sales for shooters and platforms there seems to be one game on the sports genre that has significantly more sales than any other game in this graph with over 80 million copies sold!

Let's explore this further and see what game is above 80 million copies sold

```
vgsales %>%
  filter(Genre == "Sports", Global_Sales > 80)
```

```
##   Rank       Name Platform Year  Genre Publisher NA_Sales EU_Sales JP_Sales
## 1    1 Wii Sports      Wii 2006 Sports  Nintendo    41.49    29.02     3.77
##   Other_Sales Global_Sales
## 1        8.46        82.74
```

Apparently Wii sports was the one game that sold better than all the other games by a huge margin being more than double then that of Super Mario for the Nintendo NES. So by using the R code we were able to find what games people like to buy on average as well as discovering what game sold the best even though it was in a genre that wasn't as popular as the platformer genre.