

Video Game Sales Analysis

Alex Flores

September 27, 2018

Contents

| | |
|--|----------|
| Introduction | 1 |
| Loading and checking the data | 1 |
| Question: How much did sales increase on average each year between 2000 and 2008? | 3 |
| Question: Are all of the countries in the dataset following the same trend as the global sales data? | 4 |
| How much did exactly did each country increase in sales on average, in each year between 2000 and 2008? | 5 |
| Question: Which genres in North America caused such a rapid increase in sales between the years 2000 and 2008 for North America? | 6 |
| Question: What percent of North American video games sold in between 2000 and 2008, were the result of Sports and Actions games? | 7 |
| Summary | 8 |

Introduction

Before we begin, if you would like to see the data for yourself, please visit the following link: (<https://data.world/julienf/video-games-global-sales-in-volume-1983-2017>).

Video games are a integral part of the US economy. In 2015 alone, the video game industry alone contributed a total of **\$11.7 BILLION** to the US gross domestic product accoring to the ESA(Entertainment Software Association). In addition, there are **2,568** game company locations in the US employing about 65,678 people. In total though, about **220 THOUSAND** people are directly or indirectly employed by the US game industry in 2015. With all this said, it is important to consider how video game sales have changed over time throughout the world. It is important to consider certain questions such as, are video game sales increasing or are they stagnating? It is important for economists to consider this because video games play an integral part of the US economy. With that being said, lets jump into our data so we can solve this question.

Loading and checking the data

The first step is to load the packages we are going to be using. Then we change the working directory to where my Video Game files are located. After that we read the file into R via the read.csv function. Once we have done this, we can actually work with our sales data.

Note: All sales units are in millions

```
library("knitr")
library("dplyr")
library("tidyverse")
setwd("C:\\Users\\flore\\Desktop\\Job Search Documents\\R Working Directory")
sales <- read.csv("VideoGameSalesGlobal.csv")
```

```
head(sales,10)
```

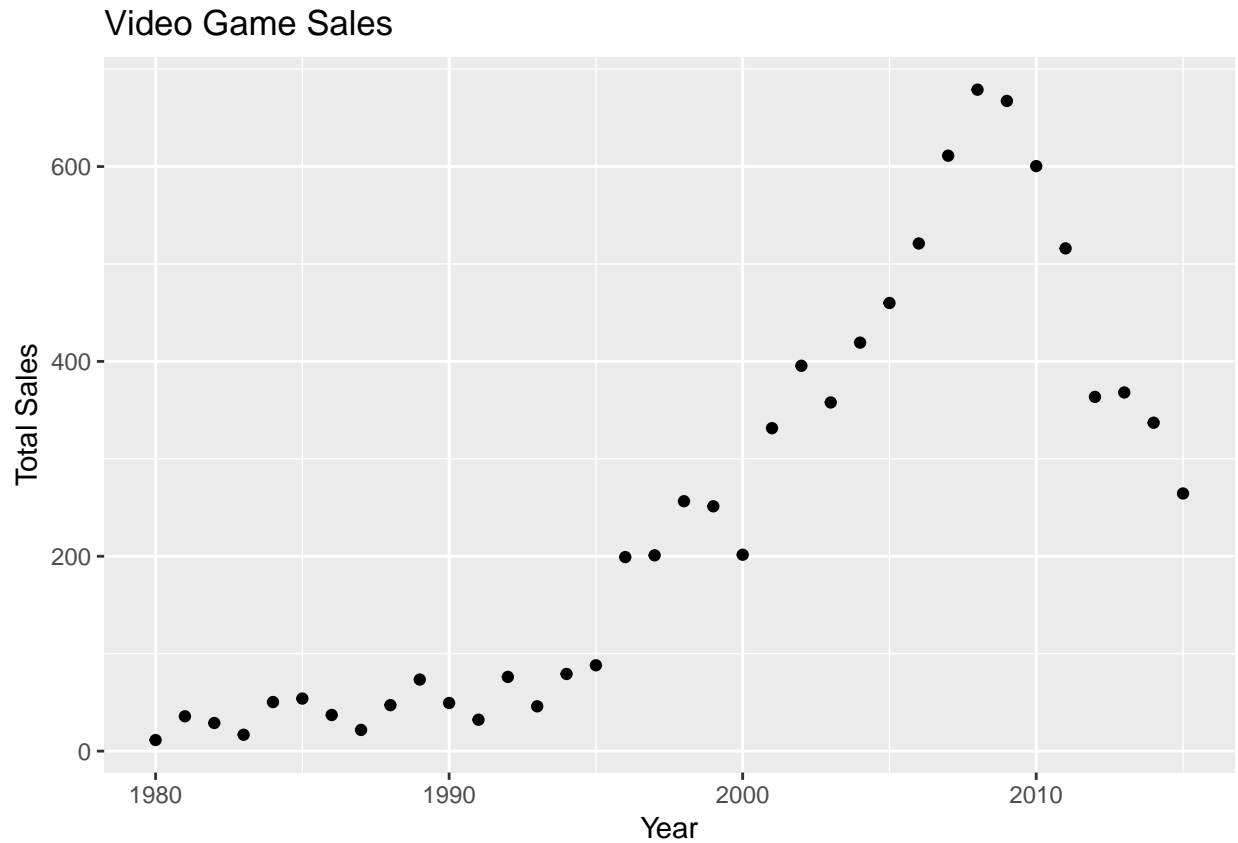
| Rank | Name | Platform | Year | Genre | Publisher | NA_Sales | EU_Sales | JP_Sales |
|------|---------------------------|----------|------|--------------|-----------|----------|----------|----------|
| 1 | Wii Sports | Wii | 2006 | Sports | Nintendo | 41.49 | 29.02 | 3.77 |
| 2 | Super Mario Bros. | NES | 1985 | Platform | Nintendo | 29.08 | 3.58 | 6.81 |
| 3 | Mario Kart Wii | Wii | 2008 | Racing | Nintendo | 15.85 | 12.88 | 3.79 |
| 4 | Wii Sports Resort | Wii | 2009 | Sports | Nintendo | 15.75 | 11.01 | 3.28 |
| 5 | Pokemon Red/Pokemon Blue | GB | 1996 | Role-Playing | Nintendo | 11.27 | 8.89 | 10.22 |
| 6 | Tetris | GB | 1989 | Puzzle | Nintendo | 23.20 | 2.26 | 4.22 |
| 7 | New Super Mario Bros. | DS | 2006 | Platform | Nintendo | 11.38 | 9.23 | 6.50 |
| 8 | Wii Play | Wii | 2006 | Misc | Nintendo | 14.03 | 9.20 | 2.93 |
| 9 | New Super Mario Bros. Wii | Wii | 2009 | Platform | Nintendo | 14.59 | 7.06 | 4.70 |
| 10 | Duck Hunt | NES | 1984 | Shooter | Nintendo | 26.93 | 0.63 | 0.28 |

The next step is to change the Year variable into a character and then into a integer so we can use the filter function on it. The reason we can't convert "Years" directly to an integer is because the "Year" data is stored internally as factors which already correspond to certain integers. We want to filter out the data that occurs before the Year 2016 because we want complete data and also we don't want predictions. We also filter out all the incomplete, "N/A" years.

```
sales$Year <- as.character(sales$Year)
sales$Year <- as.integer(sales$Year)
salesWithYear <- sales %>%
  filter(Year<2016, Year!="N/A")
```

The next thing we do is sum the total sales by years and graph the results. As we can see from the results. There is a **large spike** in sales between the years 2000 and 2008.

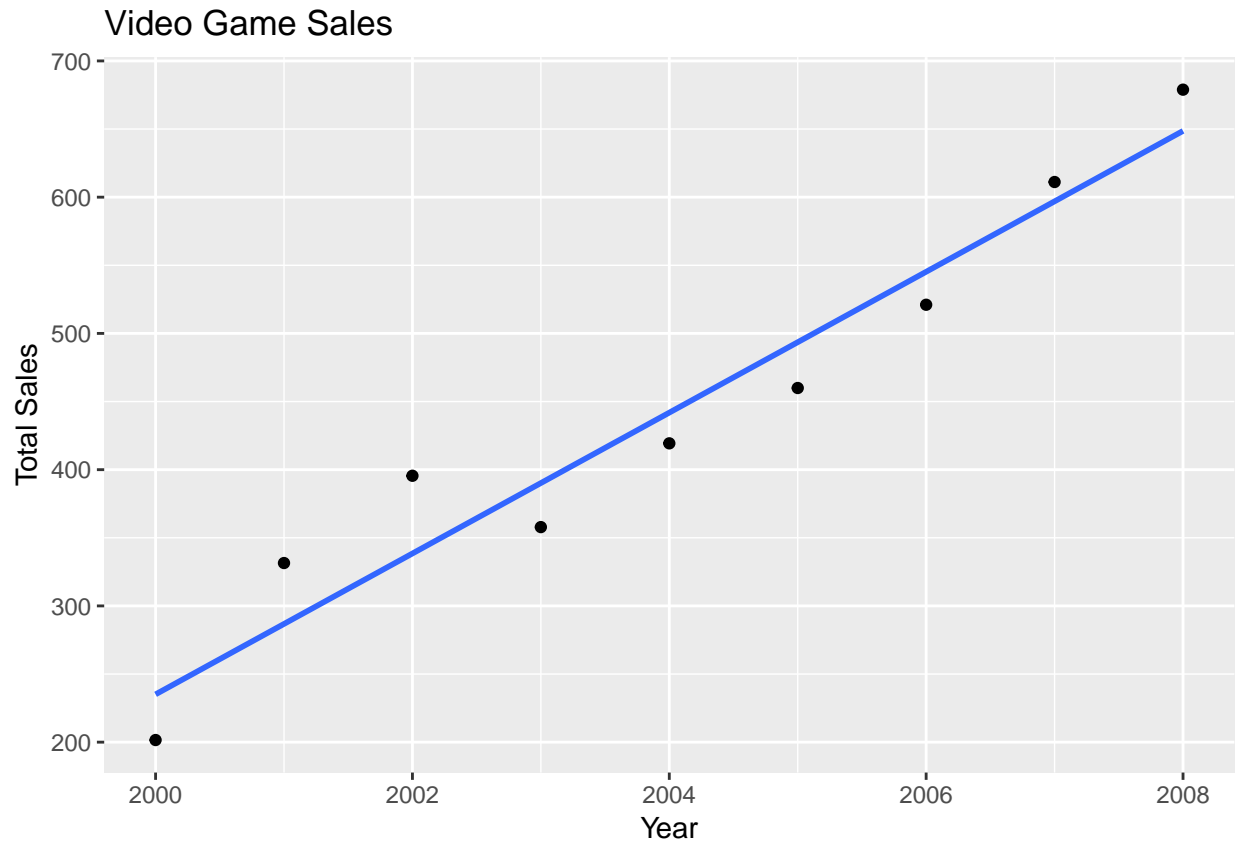
```
salesGroupedByYear <- salesWithYear %>%
  group_by(Year) %>%
  summarise(Total_Sales = sum(Global_Sales))
ggplot(salesGroupedByYear,
  aes(x=Year,y=Total_Sales)) +
  geom_point() +
  ggtitle("Video Game Sales") +
  labs(y="Total Sales",x="Year")
```



Question: How much did sales increase on average each year between 2000 and 2008?

To answer this question we can run a linear regression between the years between 2000 and 2008. First, I must filter the dataset between the years 2000 and 2008 then I can run a linear regression on the data. Then I summarize my linear model in R using the `coeff` function. From the linear regression, we can see that sales increased on average of roughly **51.70 Million Units** each year between 2000 and 2008.

```
outlierSales <- salesGroupedByYear %>%
  filter(Year > 1999, Year < 2009)
ggplot(outlierSales, aes(x=Year, y=Total_Sales)) +
  geom_point() +
  ggtitle("Video Game Sales") +
  geom_smooth(method = "lm", se = FALSE) +
  labs(y="Total Sales", x="Year")
```



```
LinearModel <- lm(Total_Sales ~ Year, outlierSales)
coef(LinearModel)
```

```
##      (Intercept)          Year
## -103147.24022      51.69117
```

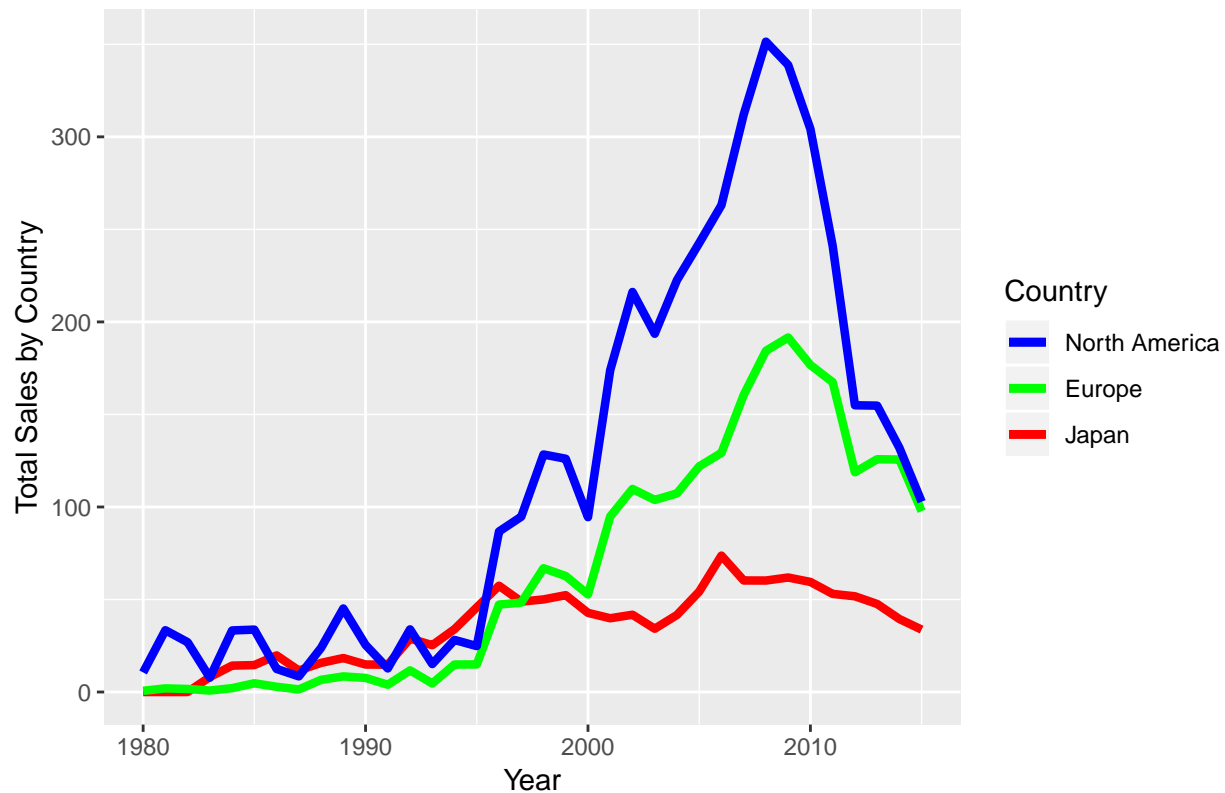
Question: Are all of the countries in the dataset following the same trend as the global sales data?

We can see that video game sales do increase around the world greatly between the years 2000 and 2008 but we can't tell exactly how much each specific country grows within this time period. This made me curious as to how much exactly each country is growing in sales over time. This leads to my next question...

```
salesGroupedByCountry <- salesWithYear %>%
  group_by(Year) %>%
  summarise(Total_NA=sum(NA_Sales),
            Total_EU=sum(EU_Sales),
            Total_JP=sum(JP_Sales))

ggplot(salesGroupedByCountry, aes(Year)) +
  geom_line(aes(y=Total_JP, color="red"), size=1.5) +
  geom_line(aes(y=Total_EU, color="green"), size=1.5) +
  geom_line(aes(y=Total_NA, color="blue"), size=1.5) +
  ggtitle("Video Game Sales by Country") +
  labs(y="Total Sales by Country", x="Year") +
  guides(color=guide_legend("Country")) +
  scale_color_manual(labels = c("North America", "Europe", "Japan"),
                    values = c("blue", "green", "red"))
```

Video Game Sales by Country

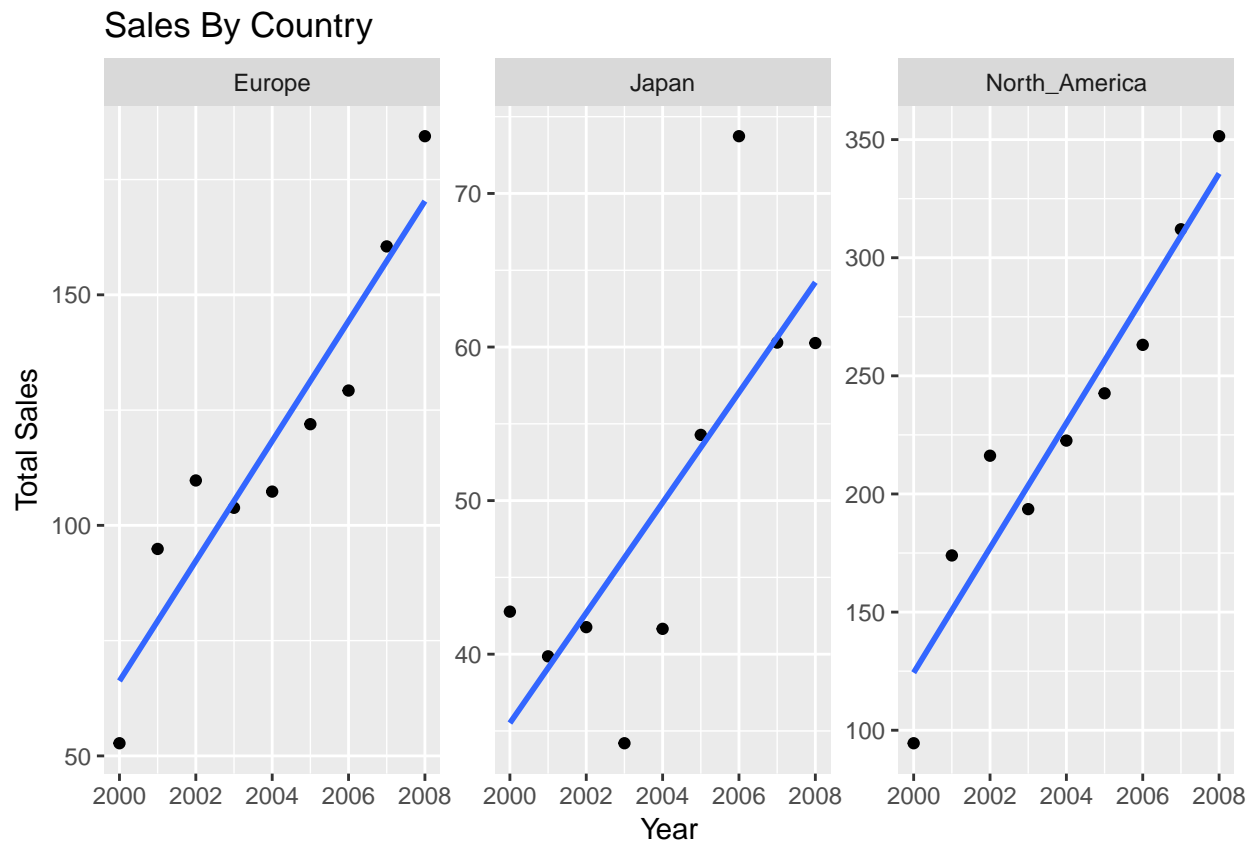


How much did exactly did each country increase in sales on average, in each year between 2000 and 2008?

We can tell exactly how much sales increased on average, by each year, respective to each country. If we take a look at the beta coefficient in our linear regression, we can see that Japan only increased sales by an average of **3.59 million** sales each year between 2000-2008. In contrast, Europe increased sales by an average of **13.01 million** sales beating out Japan. However, in these years North America was able to increase sales the most on average with an average increase of **26.41 million** sales each year between 2000-2008.

```
salesSummary <- sales[c(4,7:9)]
salesSummary <- salesSummary %>%
  arrange(Year)
salesSummary <- salesSummary %>%
  group_by(Year) %>%
  filter(Year>1999,Year<2009) %>%
  summarise(North_America=sum(NA_Sales),
            Europe=sum(EU_Sales),
            Japan=sum(JP_Sales))
NorthAmerica <- lm(North_America ~ Year, salesSummary)
Europe <- lm(Europe ~ Year, salesSummary)
Japan <- lm(Japan ~ Year, salesSummary)
coef(NorthAmerica)
coef(Europe)
coef(Japan)
```

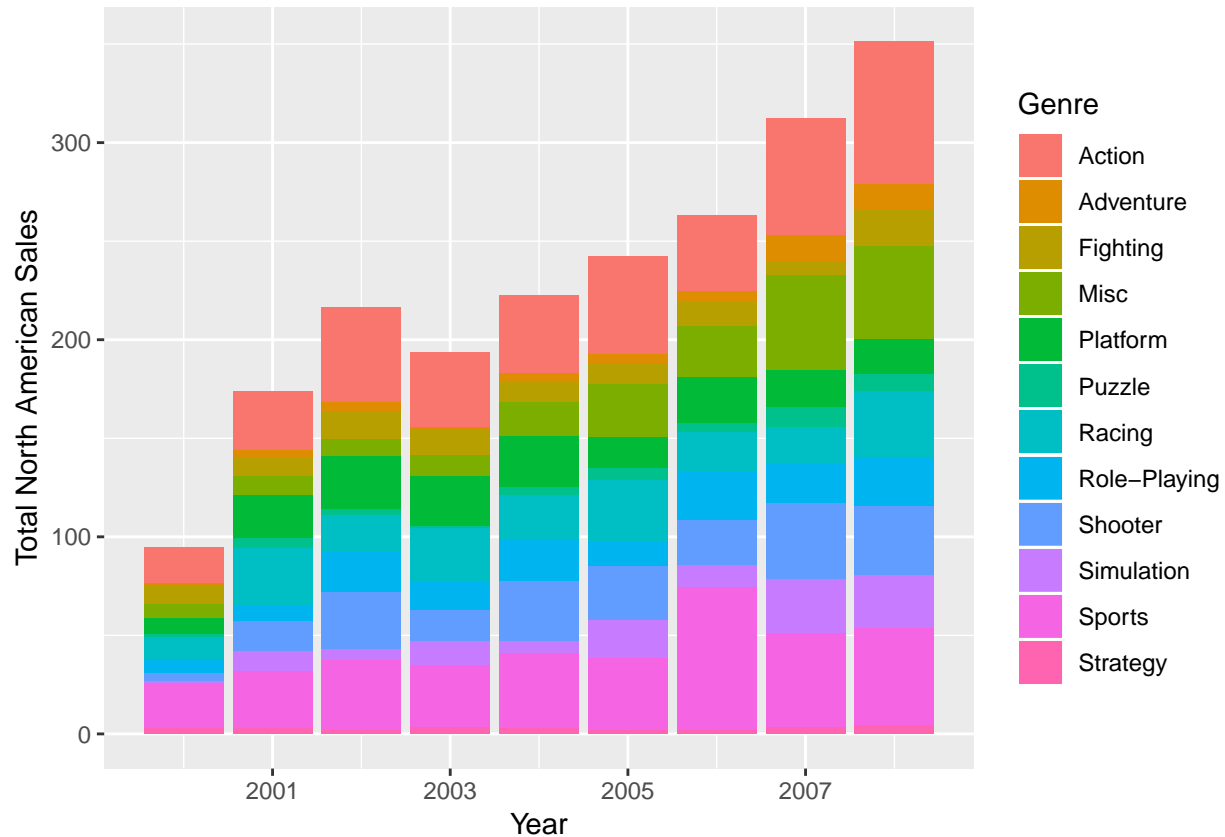
```
salesSummary <- gather(salesSummary, Country, Sales, ~Year)
ggplot(salesSummary, aes(x=Year, y=Sales)) +
  geom_point() +
  geom_smooth(method = "lm", se=FALSE) + facet_wrap(~Country, scales='free_y') +
  labs(title="Sales By Country", y="Total Sales")
```



Question: Which genres in North America caused such a rapid increase in sales between the years 2000 and 2008 for North America?

The first step, as usual is to create a data frame that we can work with using dplyr's group by and summarise verbs. We filter this new data table with three variables for the years between 2000-2008 and graph that data we create. From this graph, we can see that no one genre completely dominates the sales over time however, we can roughly see that Action and Sports games sell the most. This leads to my next question...

```
salesByGenre <- salesWithYear %>%
  group_by(Year, Genre) %>%
  summarise(Total_NA_Sales=sum(NA_Sales))
salesByGenre <- salesByGenre %>%
  filter(Year>1999, Year<2009)
ggplot(salesByGenre, aes(x=Year, y = Total_NA_Sales, fill = Genre)) +
  geom_bar(stat="identity") +
  labs(y="Total North American Sales", x="Year")
```



Question: What percent of North American video games sold in between 2000 and 2008, were the result of Sports and Actions games?

To do this we must filter our dataset first, then we can sum the total amount of North American Action and Sports games between 2000 and 2008. Afterwards, we can divide by the total amount of sales to find out that almost 1/5 of all Sales in this time period came from Action and Sports games.

```
GlobalSales <- salesWithYear %>%
  filter(Year>1999,Year<2009) %>%
  summarise((sum(Global_Sales)))
salesWithYear$Genre <- as.character(salesWithYear$Genre)
NA_Sales_Action_Sports <- salesWithYear %>%
  filter(Year>1999,Year<2009) %>%
  filter(Genre %in% c("Action","Sports")) %>%
  summarise(sum(NA_Sales))
NumberOfSales<-(as.integer(NA_Sales_Action_Sports))/as.integer(GlobalSales)
print(paste0("Action and Sport Games as a Percentage of Total Sales between 2000-2008: ",
  round(NumberOfSales*100,digits = 2),
  "%"))
```

```
## [1] "Action and Sport Games as a Percentage of Total Sales between 2000-2008: 18.94%"
```

Summary

- **Video game sales peaked around the year 2008.**
 - This peak is not the same for Japan which peaked slightly earlier.
- **Sales increased an average of 51.70 Million Units each year between 2000 and 2008.**
 - This also varied widely between countries. For the US, it was **26.41 million** and for Europe it was **13.01 million** however for Japan it was only **3.59 million**.
- **No single Genre dominated the video games sales industry however, Sports and Action games sold the most.**
 - Between 2000-2008, Sports and Action games accounted for almost **1/5** of all video game sales.