615 Final Project

Video Game Sales

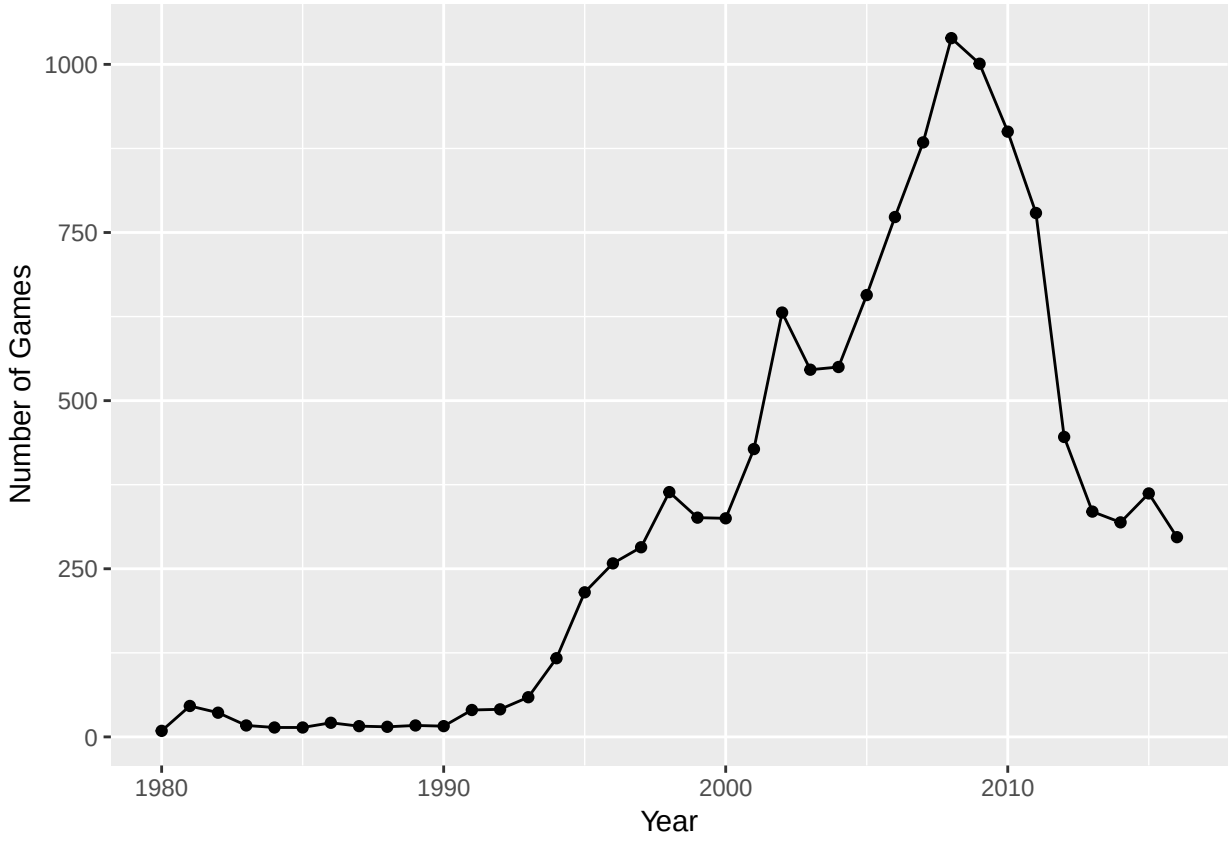
Xinyi Wang

12/9/2018

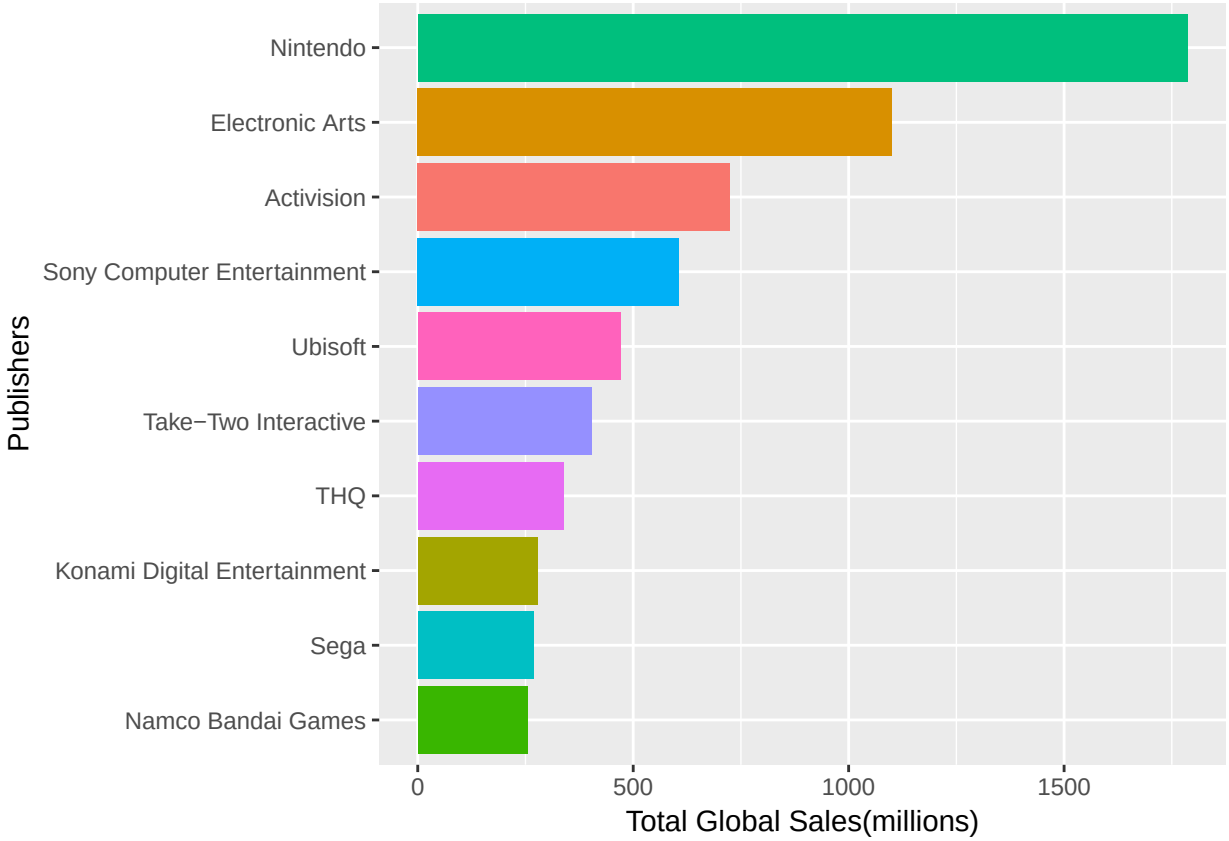
Read and Clean Data

EDA

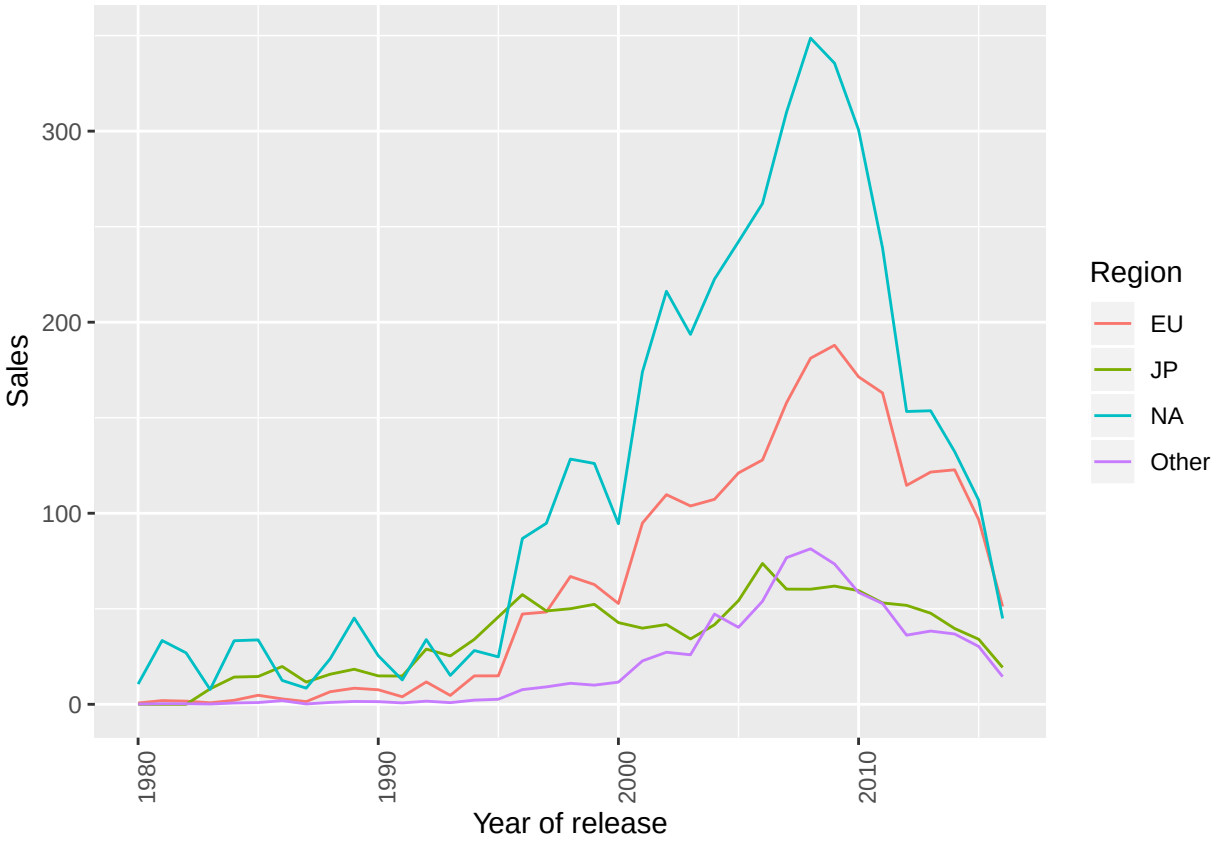
1.Games Released Each Year by All Publishers(1980-2016)



2.top 10 publishers with higher revenue from 1980-2016



3.Sales per region in timeline



4.global sales map by platform and year

```
##
## Attaching package: 'maps'

## The following object is masked from 'package:purrr':
##
##     map

##  [1] "French Southern and Antarctic Lands"
##  [2] "Antigua"
##  [3] "Barbuda"
##  [4] "Saint Barthelemy"
##  [5] "Ivory Coast"
##  [6] "Curacao"
##  [7] "Canary Islands"
##  [8] "UK"
##  [9] "Heard Island"
## [10] "Chagos Archipelago"
## [11] "Siachen Glacier"
## [12] "Nevis"
## [13] "Saint Kitts"
## [14] "Saint Martin"
## [15] "Bonaire"
```

```
## [16] "Sint Eustatius"
## [17] "Saba"
## [18] "Madeira Islands"
## [19] "Azores"
## [20] "Palestine"
## [21] "South Sandwich Islands"
## [22] "South Georgia"
## [23] "Ascension Island"
## [24] "Timor-Leste"
## [25] "Trinidad"
## [26] "Tobago"
## [27] "USA"
## [28] "Vatican"
## [29] "Grenadines"
## [30] "Saint Vincent"
## [31] "Virgin Islands"
```

```
## [1] "United Kingdom"
```

```
##
## Attaching package: 'plotly'
```

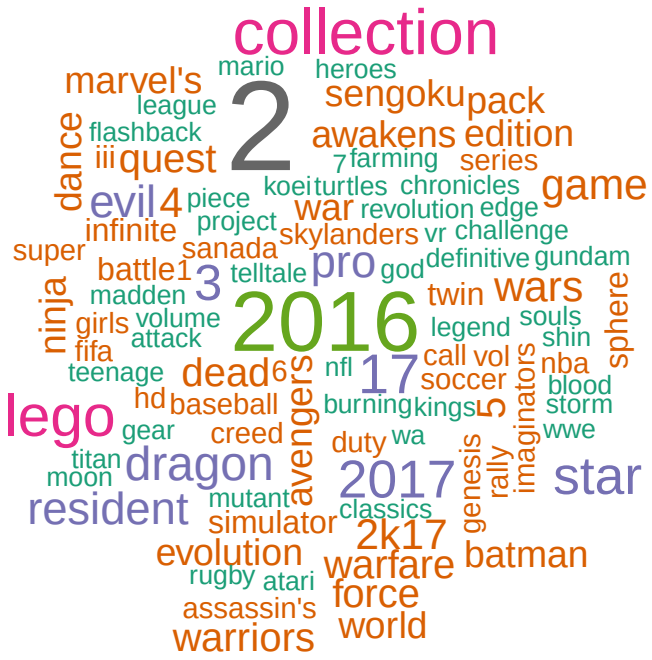
```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

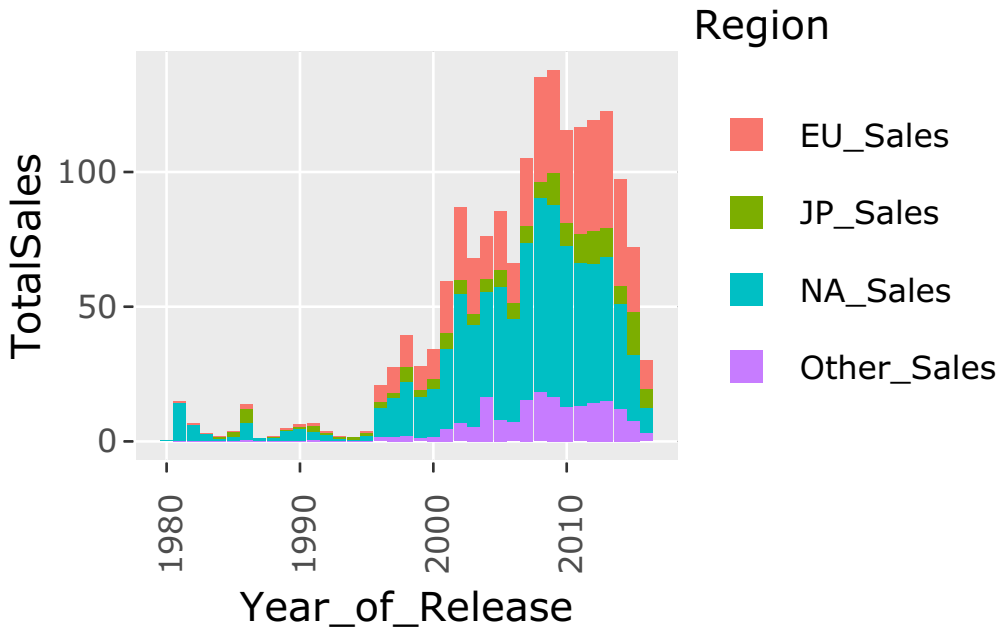
```
## The following object is masked from 'package:graphics':
##
##     layout
```

5.game name text analysis (word cloud)

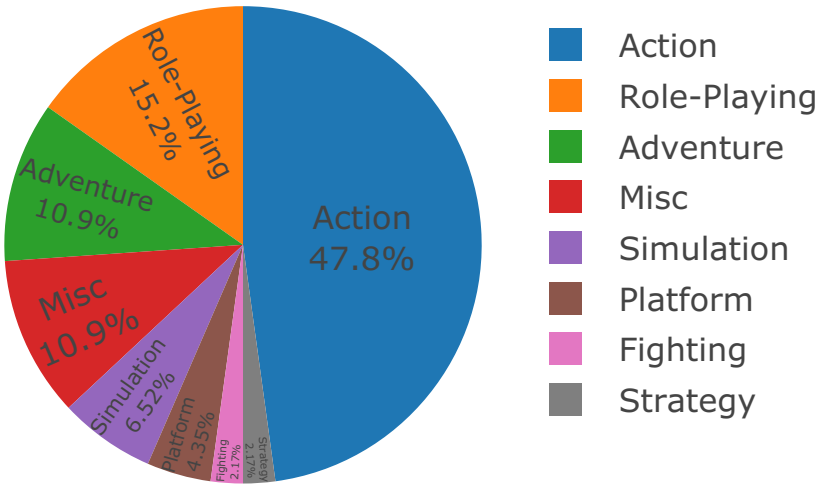
```
## Joining, by = "word"
## Loading required package: RColorBrewer
```



6.Sales by Genre

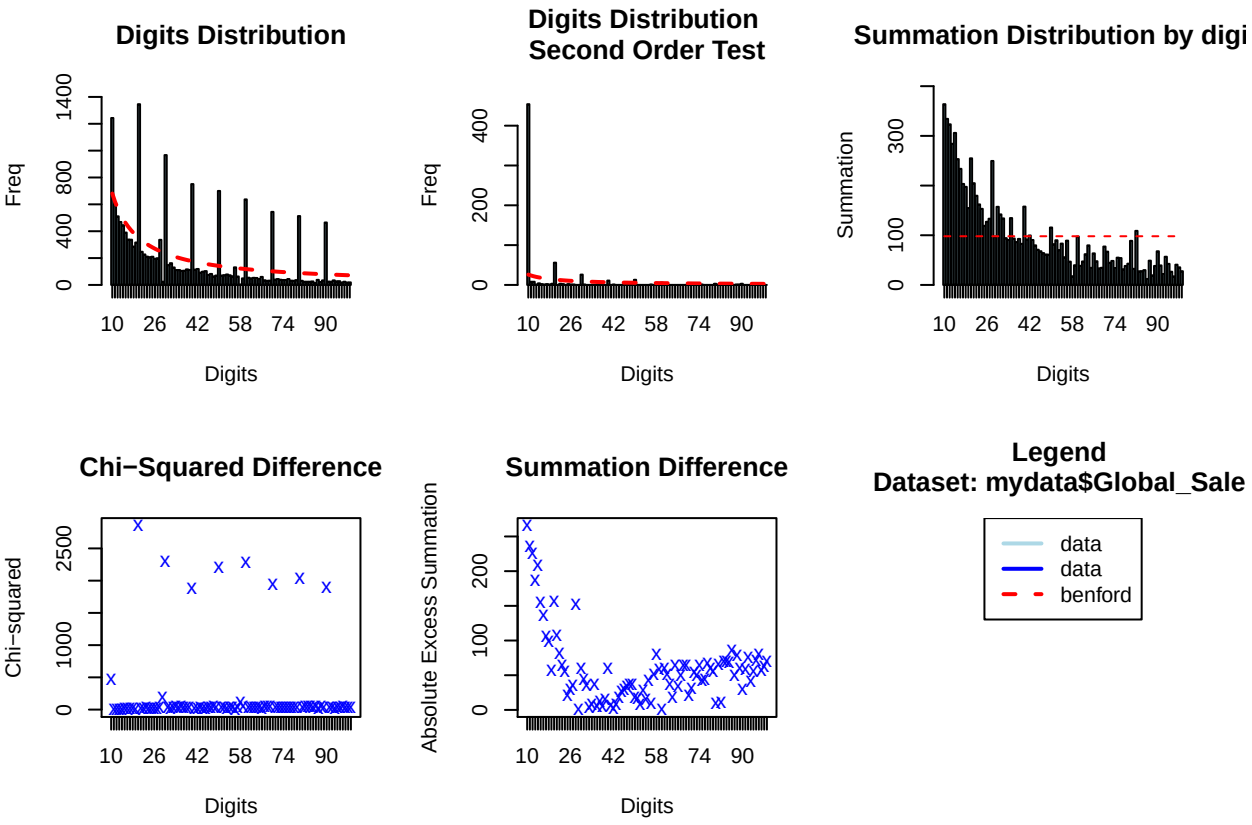


7.Pie chart of genres in certain year and platform



Benford Law

```
library(benford.analysis)
bfd <- benford(mydata$Global_Sales)
plot(bfd)
```




```
library(BenfordTests)
# Euclidean Distance Test for Benford's Law
```

```
edist.benftest(mydata$Global_Sales)
```

```
##
##  Euclidean Distance Test for Benford Distribution
##
## data:  mydata$Global_Sales
## d_star = 2.6538, p-value < 2.2e-16
```

```
# The p-value is smaller than 0.05 so that we reject the null hypothesis. Therefore, the goodness-of-fi
```

Top 10 Games

```
#Take 2016 as example
toptable <- mydata %>%
  select(Name,Global_Sales,Year_of_Release,Platform) %>%
  filter(Year_of_Release==2016) %>%
  arrange(desc(Global_Sales)) %>%
  distinct(Name,Global_Sales) %>%
  head(10)
```