Final Exam

Kelin Pan

August 22, 2018

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ----------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

Discussion

In this report, I examined the data of video games provided by my boss Doc B which shows the sales condition and users ranks from the year 1980 to year 2018. Results show that Palcom company has the highest mean of global sales although it only published one game. I also find that action games have the largest total sales in North America region, which means this kind of game is extremely popular among this area. What's more, I also used regular expression to find some key words about games names of particular game companies and used function and for loop the calculate the publishers with video game sales above the median value. Below I will illustrate my code and provide additional details about my analysis.

```
vg<-read.csv("C:/Users/panke/Downloads/WFU/R/final exam/videoGame.csv")
summary(vg)
```

```
##                            Name          Platform    Year_of_Release
##   Need for Speed: Most Wanted:  12    PS2   :2161    2008   :1427
##   FIFA 14                    :   9    DS    :2152    2009   :1426
##   LEGO Marvel Super Heroes   :   9    PS3   :1331    2010   :1255
##   Madden NFL 07              :   9    Wii   :1320    2007   :1197
##   Ratatouille                :   9    X360  :1262    2011   :1136
##   Angry Birds Star Wars      :   8    PSP   :1209    2006   :1006
##   (Other)                    :16663   (Other):7284   (Other):9272
##          Genre                          Publisher
##   Action      :3370    Electronic Arts            : 1356
##   Sports      :2348    Activision                 :  985
##   Misc        :1750    Namco Bandai Games         :  939
##   Role-Playing:1500    Ubisoft                    :  933
##   Shooter     :1323    Konami Digital Entertainment:  834
```

```
##   Adventure   :1303   THQ                    :  715
##   (Other)     :5125   (Other)                :10957
##     NA_Sales          EU_Sales          JP_Sales         Other_Sales
##   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.0000   Min.   : 0.00000
##   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.00000
##   Median : 0.0800   Median : 0.020   Median : 0.0000   Median : 0.01000
##   Mean   : 0.2633   Mean   : 0.145   Mean   : 0.0776   Mean   : 0.04733
##   3rd Qu.: 0.2400   3rd Qu.: 0.110   3rd Qu.: 0.0400   3rd Qu.: 0.03000
##   Max.   :41.3600   Max.   :28.960   Max.   :10.2200   Max.   :10.57000
##
##   Global_Sales      Critic_Score    Critic_Count      User_Score
##   Min.   : 0.0100   Min.   :13.00   Min.   :  3.00         :6704
##   1st Qu.: 0.0600   1st Qu.:60.00   1st Qu.: 12.00   tbd    :2425
##   Median : 0.1700   Median :71.00   Median : 21.00   7.8    : 324
##   Mean   : 0.5335   Mean   :68.97   Mean   : 26.36   8      : 290
##   3rd Qu.: 0.4700   3rd Qu.:79.00   3rd Qu.: 36.00   8.2    : 282
##   Max.   :82.5300   Max.   :98.00   Max.   :113.00   8.3    : 254
##                     NA's   :8582    NA's   :8582     (Other):6440
##     User_Count        Developer         Rating
##   Min.   :    4.0            :6623            :6769
##   1st Qu.:   10.0   Ubisoft  : 204   E       :3991
##   Median :   24.0   EA Sports: 172   T       :2961
##   Mean   :  162.2   EA Canada: 167   M       :1563
##   3rd Qu.:   81.0   Konami   : 162   E10+    :1420
##   Max.   :10665.0   Capcom   : 139   EC      :   8
##   NA's   :9129      (Other)  :9252   (Other) :   7
```
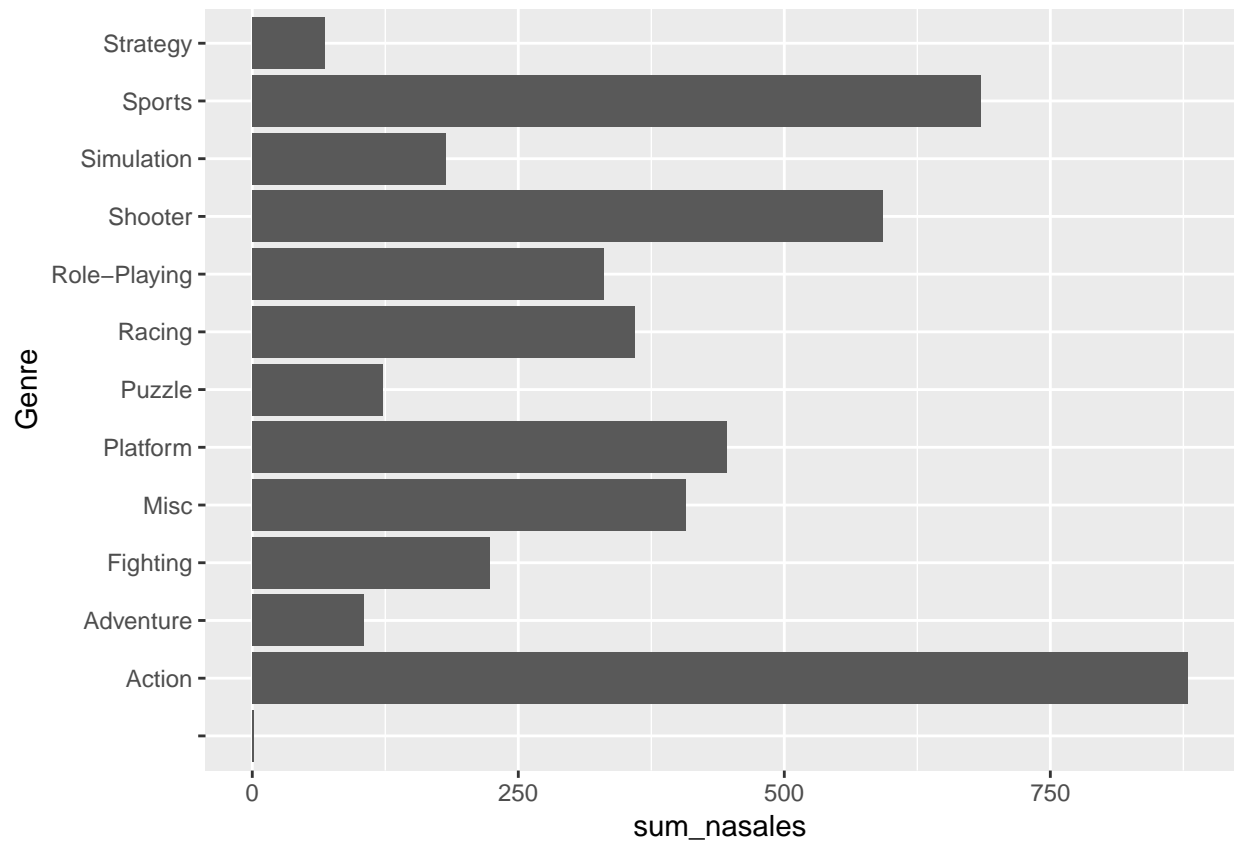
```r
#plot 1#
a<-vg%>%group_by(Publisher)%>%
  summarise(mean_global_sales=mean(Global_Sales,na.rm=T))%>%
  filter(mean_global_sales>=1)

ggplot(data=a)+
  geom_bar(aes(x=Publisher,y=mean_global_sales),stat="identity",position="dodge")+
  coord_flip()
```
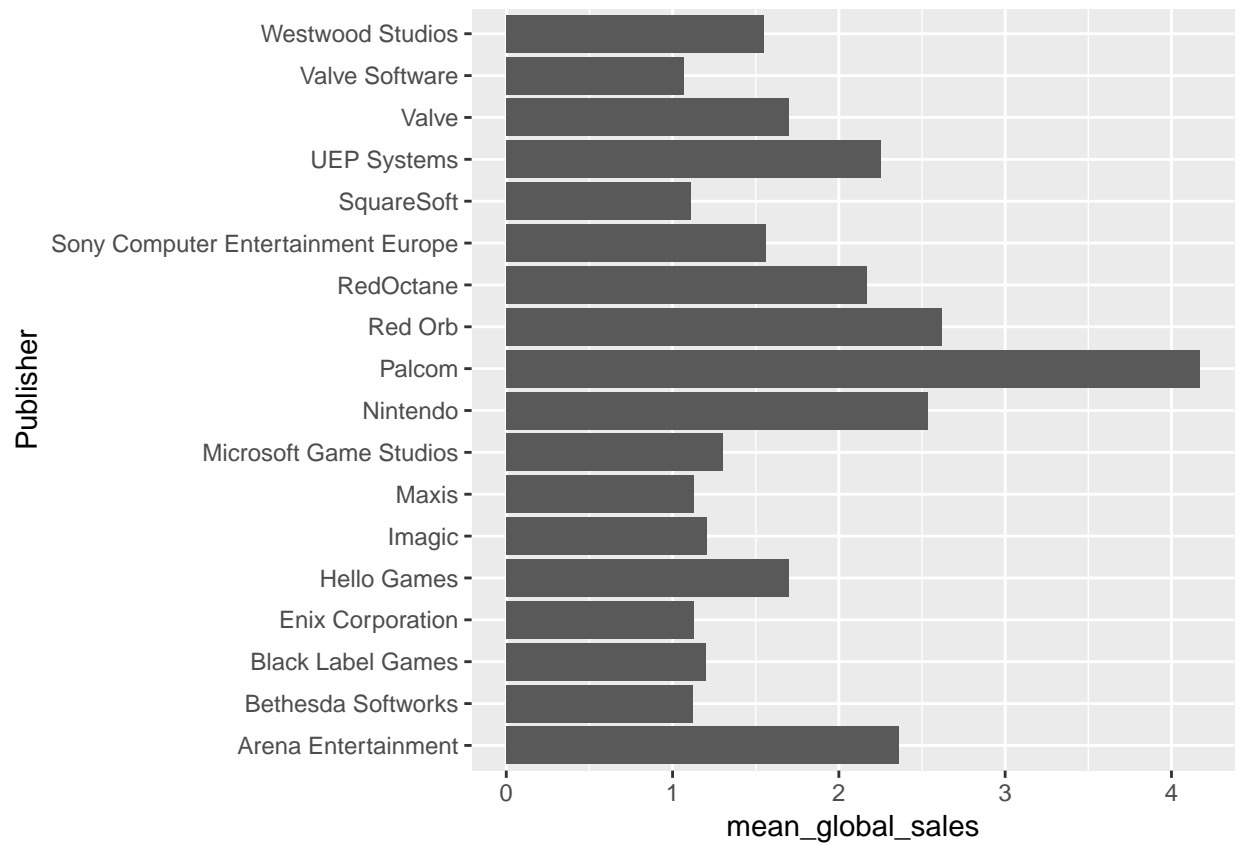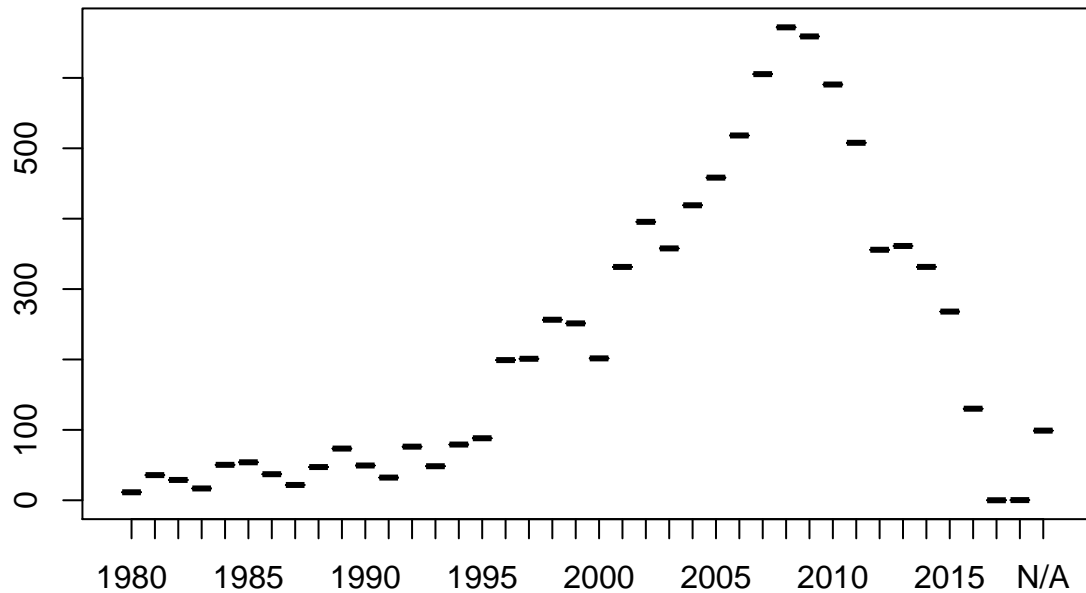
```r
#plot 2#
b<-vg%>%
  group_by(Genre)%>%
  summarise(sum_nasales=sum(NA_Sales,na.rm=T))
ggplot(b)+
  geom_bar(aes(x=Genre,y=sum_nasales),stat="identity")+
  coord_flip()
```

```
#plot 3#
c<-vg%>%
  group_by(Year_of_Release)%>%
  summarise(sum_global_sales=sum(Global_Sales,na.rm=T))

plot(x=c$Year_of_Release,y=c$sum_global_sales,main="Total Global Sales with Year")
```

Total Global Sales with Year


```
palcom <- vg[grepl('palcom', vg$Publisher, ignore.case = TRUE),]
palcom
```

```
##                           Name Platform Year_of_Release  Genre Publisher
## 280 Teenage Mutant Ninja Turtles     NES            1989 Action    Palcom
##     NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales Critic_Score
## 280     3.38     0.44     0.31        0.04         4.17           NA
##     Critic_Count User_Score User_Count Developer Rating
## 280           NA                    NA
```

```
g<-sum(str_detect(vg$Publisher, fixed("Red Orb",ignore_case=TRUE)))
g
```

```
## [1] 2
```

```
Red_orb <- vg[grepl('Red Orb', vg$Publisher, ignore.case = TRUE),]
Red_orb
```

```
##                           Name Platform Year_of_Release     Genre Publisher
## 522                       Myst      PC            1994 Adventure   Red Orb
## 638 Riven: The Sequel to Myst      PC            1997 Adventure   Red Orb
##     NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales Critic_Score
## 522     0.02     2.79        0         0.0         2.81           NA
## 638     1.52     0.82        0         0.1         2.43           83
##     Critic_Count User_Score User_Count   Developer Rating
## 522           NA                    NA
## 638           12                    NA Cyan Worlds      E
```

```
d<-count(vg,Publisher)
d<-arrange(d,desc(n))
d
```

```
## # A tibble: 582 x 2
##    Publisher                        n
##    <fct>                        <int>
##  1 Electronic Arts               1356
##  2 Activision                     985
##  3 Namco Bandai Games             939
##  4 Ubisoft                        933
##  5 Konami Digital Entertainment   834
##  6 THQ                            715
##  7 Nintendo                       706
##  8 Sony Computer Entertainment    687
##  9 Sega                           638
## 10 Take-Two Interactive           422
## # ... with 572 more rows
```

```
electronic_arts<-vg[grepl('electronic arts', vg$Publisher, ignore.case = TRUE),]
h<-sum(str_detect(vg$Name,fixed("fifa",ignore_case=T)))
h
```

```
## [1] 142
```

Visualization, Transformation, Wrangling I & II

In this step, I find that the company Palcom has the largest mean global sales and Valve Software has the smallest mean global sales. As for genres, action games are the best selling games in North America area and Strategy games seem not to be popular. Finally, I plotted the relationship between year and total global sales and I find that with the trend of time, the total global sales increased, and it peaked at 2008 and 2009. After that, the total sales dropped dramatically and reached the lowest in 2018 year which means the sales of games globally is not good. After that, I used stringr to find some key words in the data set. Because I find that the Palcom is the best seller, so I want to know what game it has. And I found that there was only a game called teenage mutant ninja turtles which is amazing. And I also find that there are 157 games containing "fifa".

```
median(vg$NA_Sales,na.rm=T)
```

```
## [1] 0.08
```

```
md<-function(x){
  if(x>0.08){
    return("above")
  }else if(x==0.08){
    return("equal")
  }else{
    return("below")
  }
}
```

```
md(vg$NA_Sales[5581])
```

```
## [1] "below"
```

```
n=1
output<-vector("integer",length=1)
```

```r
for(i in seq_along(vg$NA_Sales)){
  if(md(vg$NA_Sales[[i]])=="above"){
    output<-n
    n<-n+1
  }else{
    n<-n
  }
}
output
```

```
## [1] 7994
```

```r
vg2<-filter(vg,NA_Sales>0.08)
vg_number_publisher<-vg2%>%count(Publisher)%>%arrange(desc(n))

vg3<-vg%>%group_by(Year_of_Release,Genre,Publisher)%>%
  summarise(total_sales=sum(Global_Sales,na.rm=T))
vg3
```

```
## # A tibble: 6,077 x 4
## # Groups:   Year_of_Release, Genre [?]
##     Year_of_Release Genre    Publisher                   total_sales
##     <fct>           <fct>    <fct>                              <dbl>
##  1 1980            Action   Activision                          0.34
##  2 1980            Fighting Activision                          0.77
##  3 1980            Misc     Activision                          1.42
##  4 1980            Misc     Atari                               1.29
##  5 1980            Shooter  Atari                               7.07
##  6 1980            Sports   Activision                          0.49
##  7 1981            Action   20th Century Fox Video Games        1.13
##  8 1981            Action   Activision                          0.82
##  9 1981            Action   Atari                               2.69
## 10 1981            Action   Coleco                              0.99
## # ... with 6,067 more rows
```
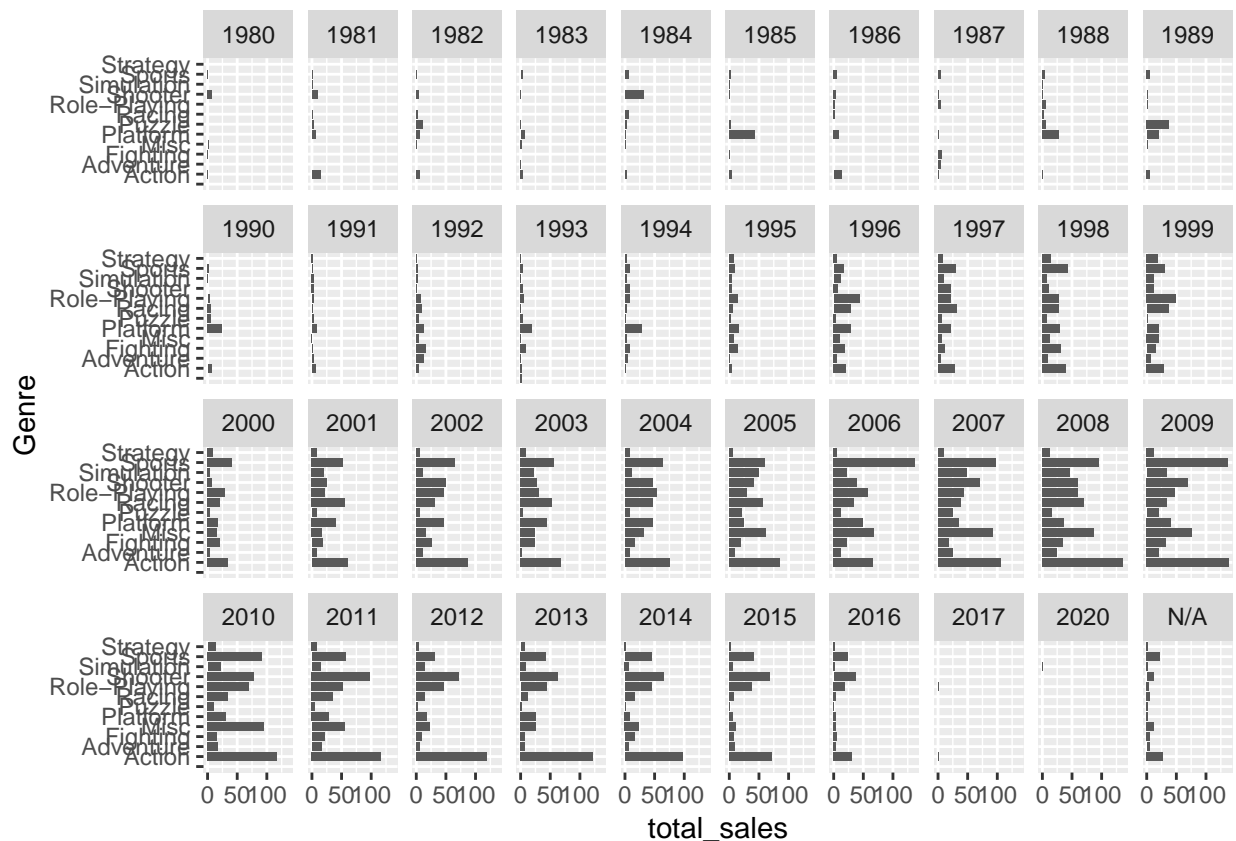
```r
ggplot(dat=vg3)+
  geom_bar(aes(x=Genre,y=total_sales),stat="identity")+
  facet_wrap(~Year_of_Release,nrow=4)+
  coord_flip()
```

```r
vg_nintendo2006<-filter(vg3,Year_of_Release==2006&Publisher=="Nintendo"&Genre=="Sports")
e<-sum(vg_nintendo2006$total_sales)
e
```

```
## [1] 84.38
```

```r
vg2009<-filter(vg3,Year_of_Release==2009&Publisher!="Nintendo"&Genre=="Sports")
f<-sum(vg2009$total_sales)
f
```

```
## [1] 81.13
```

Wrangling III, Functions, Vectors, Iterations

I used the function I created to calculate the number of publishers with video game sales above the median and I find that there are 7994 above. But in the data, there are some replicated publishers, so I count and find that there are 263 publishers with above median sales. I also find the variations in video games with genre and year. And I find that sports video games released in 2006 by Nintendo generated more global sales than their 2009 counterparts.