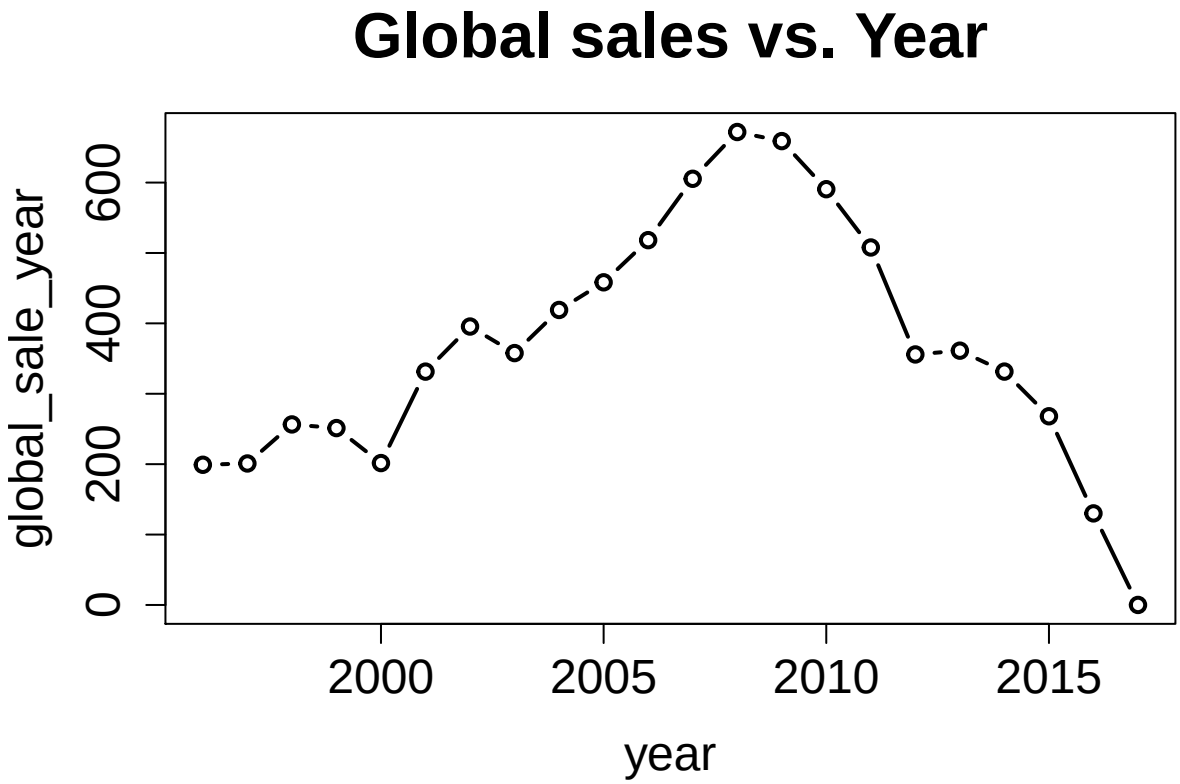
EDA

Yunguo Cai

4/7/2018

```
data = read.csv("/Users/cyunguo/Desktop/Video_Games_Sales_as_at_22_Dec_2016.csv")
```

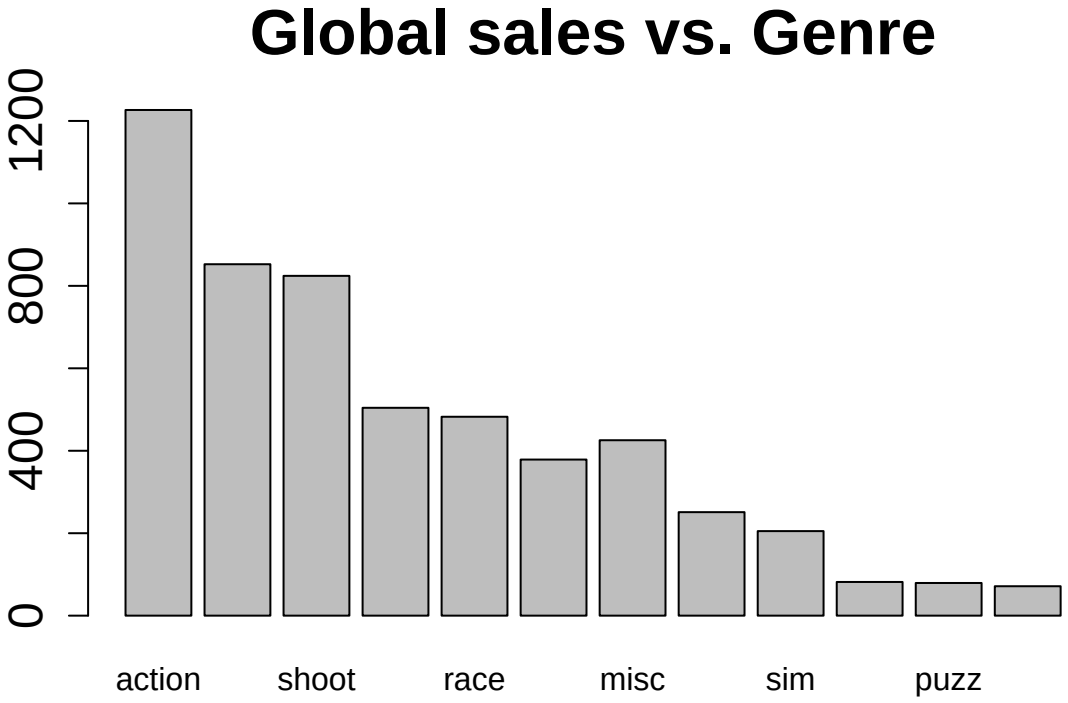
```
# year trend of games
all_1996 = data[data$Year_of_Release == "1996",]
all_1997 = data[data$Year_of_Release == "1997",]
all_1998 = data[data$Year_of_Release == "1998",]
all_1999 = data[data$Year_of_Release == "1999",]
all_2000 = data[data$Year_of_Release == "2000",]
all_2001 = data[data$Year_of_Release == "2001",]
all_2002 = data[data$Year_of_Release == "2002",]
all_2003 = data[data$Year_of_Release == "2003",]
all_2004 = data[data$Year_of_Release == "2004",]
all_2005 = data[data$Year_of_Release == "2005",]
all_2006 = data[data$Year_of_Release == "2006",]
all_2007 = data[data$Year_of_Release == "2007",]
all_2008 = data[data$Year_of_Release == "2008",]
all_2009 = data[data$Year_of_Release == "2009",]
all_2010 = data[data$Year_of_Release == "2010",]
all_2011 = data[data$Year_of_Release == "2011",]
all_2012 = data[data$Year_of_Release == "2012",]
all_2013 = data[data$Year_of_Release == "2013",]
all_2014 = data[data$Year_of_Release == "2014",]
all_2015 = data[data$Year_of_Release == "2015",]
all_2016 = data[data$Year_of_Release == "2016",]
all_2017 = data[data$Year_of_Release == "2017",]
year = c(1996:2017)
global_sale_year = c(sum(all_1996$Global_Sales), sum(all_1997$Global_Sales), sum(all_1998$Global_Sales)
plot(year, global_sale_year, type = 'b', lwd = 2,main="Global sales vs. Year",cex.axis = 1.5,cex.lab =
```



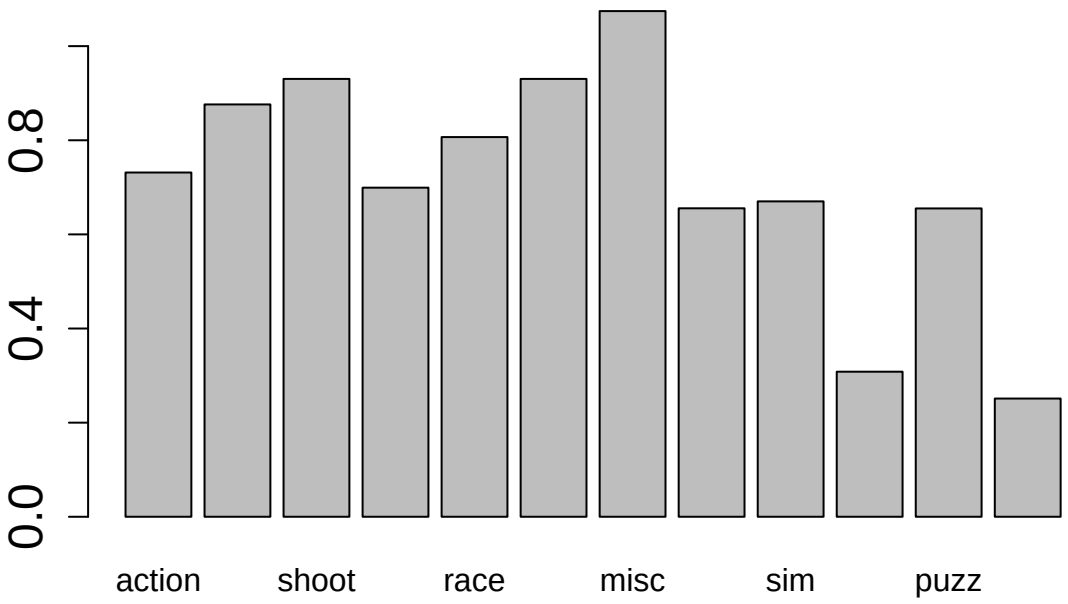

```
data = na.omit(data)
# All genre games global sale
action = data[data$Genre == "Action",]
actionsale = sum(action$Global_Sales)
#summary(adventure$Global_Sales)
actionmean = mean(action$Global_Sales)
adventure = data[data$Genre == "Adventure",]
adventuresale = sum(adventure$Global_Sales)
#summary(adventure$Global_Sales)
adventuremean = mean(adventure$Global_Sales)
fighting = data[data$Genre == "Fighting",]
fightingsale = sum(fighting$Global_Sales)
#summary(fighting$Global_Sales)
fightingmean = mean(fighting$Global_Sales)
misc = data[data$Genre == "Misc",]
miscsale = sum(misc$Global_Sales)
#summary(misc$Global_Sales)
miscmean = mean(misc$Global_Sales)
platform = data[data$Genre == "Platform",]
platformsale = sum(platform$Global_Sales)
#summary(platform$Global_Sales)
platformmean = mean(platform$Global_Sales)
puzzle = data[data$Genre == "Puzzle",]
puzzlesale = sum(puzzle$Global_Sales)
#summary(puzzle$Global_Sales)
puzzlemean = mean(puzzle$Global_Sales)
racing = data[data$Genre == "Racing",]
racingsale = sum(racing$Global_Sales)
#summary(racing$Global_Sales)
```

```
racingmean = mean(racing$Global_Sales)
rpg = data[data$Genre == "Role-Playing",]
rpgsale = sum(rpg$Global_Sales)
#summary(rpg$Global_Sales)
rpgmean = mean(rpg$Global_Sales)
shooter = data[data$Genre == "Shooter",]
shootersale = sum(shooter$Global_Sales)
#summary(shooter$Global_Sales)
shootermean = mean(shooter$Global_Sales)
simulation = data[data$Genre == "Simulation",]
simulationsale = sum(simulation$Global_Sales)
#summary(simulation$Global_Sales)
simulationmean = mean(simulation$Global_Sales)
sports = data[data$Genre == "Sports",]
sportsale = sum(sports$Global_Sales)
#summary(sports$Global_Sales)
sportmean = mean(sports$Global_Sales)
strategy = data[data$Genre == "Strategy",]
strategysale = sum(strategy$Global_Sales)
#summary(strategy$Global_Sales)
strategymean = mean(strategy$Global_Sales)
```

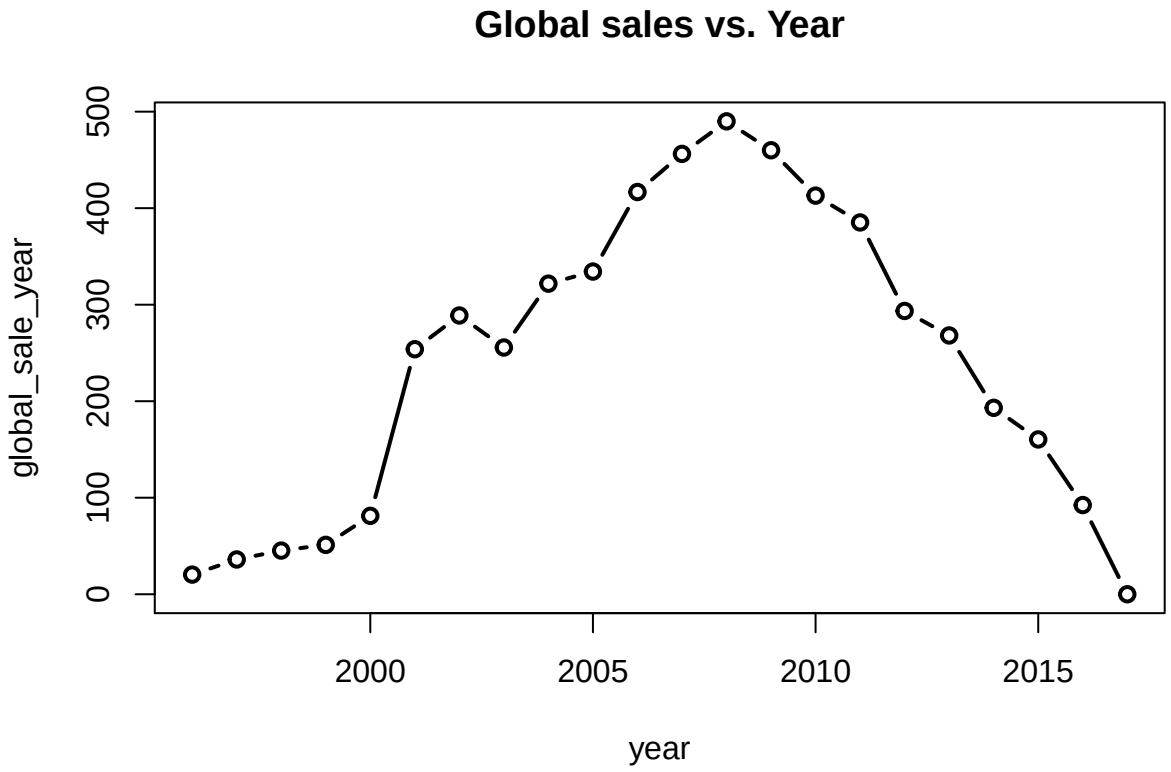
```
Global_Sales_ALL = c(actionsale, adventuresale, fightingsale, miscsale, platformsale, puzzlesale, racing

Global_Sales_MEAN = c(actionmean, adventuremean, fightingmean, miscmean, platformmean, puzzlemean, racin

barplot(height = c(Global_Sales_ALL[1], Global_Sales_ALL[11], Global_Sales_ALL[9], Global_Sales_ALL[8],
```




```
barplot(height = c(Global_Sales_MEAN[1], Global_Sales_MEAN[11], Global_Sales_MEAN[9], Global_Sales_MEAN
```




```r
# year trend of games
all_1996 = data[data$Year_of_Release == "1996",]
all_1997 = data[data$Year_of_Release == "1997",]
all_1998 = data[data$Year_of_Release == "1998",]
all_1999 = data[data$Year_of_Release == "1999",]
all_2000 = data[data$Year_of_Release == "2000",]
all_2001 = data[data$Year_of_Release == "2001",]
all_2002 = data[data$Year_of_Release == "2002",]
all_2003 = data[data$Year_of_Release == "2003",]
all_2004 = data[data$Year_of_Release == "2004",]
all_2005 = data[data$Year_of_Release == "2005",]
all_2006 = data[data$Year_of_Release == "2006",]
all_2007 = data[data$Year_of_Release == "2007",]
all_2008 = data[data$Year_of_Release == "2008",]
all_2009 = data[data$Year_of_Release == "2009",]
all_2010 = data[data$Year_of_Release == "2010",]
all_2011 = data[data$Year_of_Release == "2011",]
all_2012 = data[data$Year_of_Release == "2012",]
all_2013 = data[data$Year_of_Release == "2013",]
all_2014 = data[data$Year_of_Release == "2014",]
all_2015 = data[data$Year_of_Release == "2015",]
all_2016 = data[data$Year_of_Release == "2016",]
all_2017 = data[data$Year_of_Release == "2017",]
year = c(1996:2017)
global_sale_year = c(sum(all_1996$Global_Sales), sum(all_1997$Global_Sales), sum(all_1998$Global_Sales)
plot(year, global_sale_year, type = 'b', lwd = 2,main="Global sales vs. Year")
```


Global sales vs. Year
global_sale_year
0
100
200
300
400
500
2000
2005
2010
2015
year