1990-2010年游戏销售数据的数据探索

```
Sys.setlocale("LC_ALL","Chinese")
```

```
## [1] "LC_COLLATE=Chinese (Simplified)_China.936;LC_CTYPE=Chinese (Simplified)_China.936;LC_MONETARY=Chinese (Simplified)_C
hina.936;LC_NUMERIC=C;LC_TIME=Chinese (Simplified)_China.936"
```

```
data = read.csv("D:/R/PROJECTS/GAME SALES/Datasets6073/vgsales.csv",
                stringsAsFactors = FALSE, header=TRUE)

# 查看缺失值:
sum(is.na(data))
```

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

```
# 把 Year 转换为 INT
data[,'Year']=as.integer(data[,'Year'])
```

```
## Warning: NAs introduced by coercion
```

```
sum(is.na(data[,'Year']))
```

```
## [1] 273
```

```
# 发现271个 NA

#
data1 = data[complete.cases(data),]
sum(is.na(data1))
```

```
## [1] 0
```

```
 sort(data1[!duplicated(data1$Year),'Year'])
```

```
##  [1] 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994
## [16] 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
## [31] 2010 2011 2012 2013 2014 2015 2016 2017 2020
```

```
# 查看是什么游戏
data1[data1$Year==2020,]
```

	R…	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales
	<int>	<chr>	<chr>	<int>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
5958	5959	Imagine: Makeup Artist	DS	2020	Simulation	Ubisoft	0.27	0	0

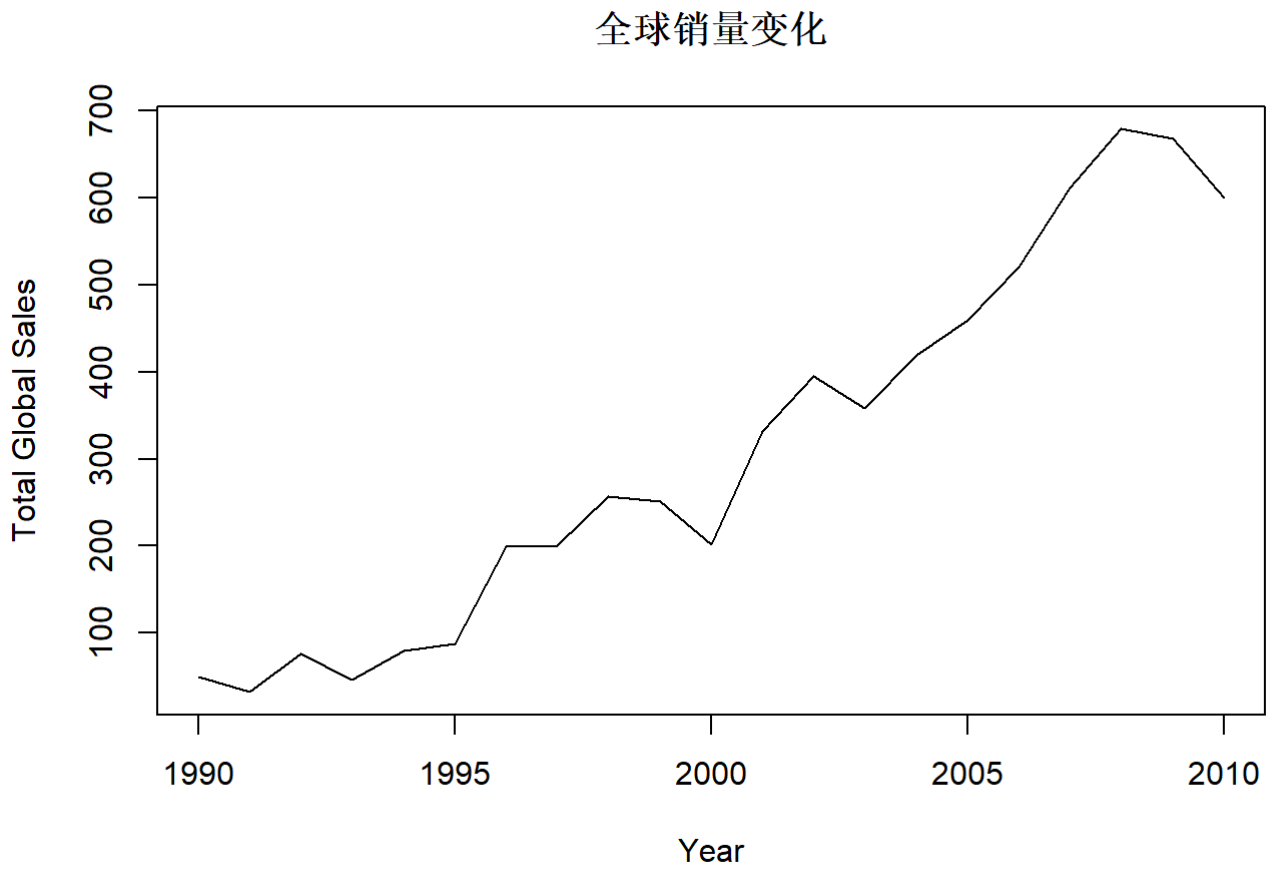
1 row | 1-10 of 12 columns

```
#经过查询，Imagine: Makeup Artist是2009年发售，对数据做出更新
data1[data1$Name=='Imagine: Makeup Artist','Year'] = 2009

sum(is.na(data1))
```

```
## [1] 0
```

```
#缺失值处理完毕

#对1990-2010年的数据进行分析
data1 = data1[data1$Year %in% c(seq(from = 1990, to = 2010, by = 1)),]
```

```
#全部游戏厂商不同年份总销售额对比情况

plot(sort(data1[!duplicated(data1$Year),'Year']),tapply(data1$Global_Sales,data1$Year,sum),type = 'l',xlab='Year', ylab='Tot
al Global Sales', main = '全球销量变化')
```



结论： 2008年之前销售额逐年升高，从2000年开始进入迅猛增长阶段，并在2008年达到顶峰。 但在2008年之后销售逐年下降。 猜测：从1980年开始到2008年，各大游戏市场趋近饱和且头部被大厂占领，小厂没有和大厂进行争夺的优势 各类游戏也没有出现太多新颖的玩法和机制，玩家的兴趣逐渐丧失。加上2008年的金融危机也对游戏产业造成了冲击，销量进一步下降。

```
#全球销量前十的游戏
head(data1[order(-data1$Global_Sales),],n=10)
```

	Ra…	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales
	<int>	<chr>	<chr>	<dbl>	<chr>	<chr>	<dbl>	<dbl>
1	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02
3	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88
4	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01
5	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89
7	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23
8	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20
9	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06
11	11	Nintendogs	DS	2005	Simulation	Nintendo	9.07	11.00
12	12	Mario Kart DS	DS	2005	Racing	Nintendo	9.81	7.57
13	13	Pokemon Gold/Pokemon Silver	GB	1999	Role-Playing	Nintendo	9.00	6.18

1-10 of 10 rows | 1-9 of 12 columns

结论: 发行商全为Nintendo，此游戏发行商制作的游戏质量很高，多数游戏的主机为Wii。 Wii平台的游戏为动作感应类，耐玩性，可玩性都很高。

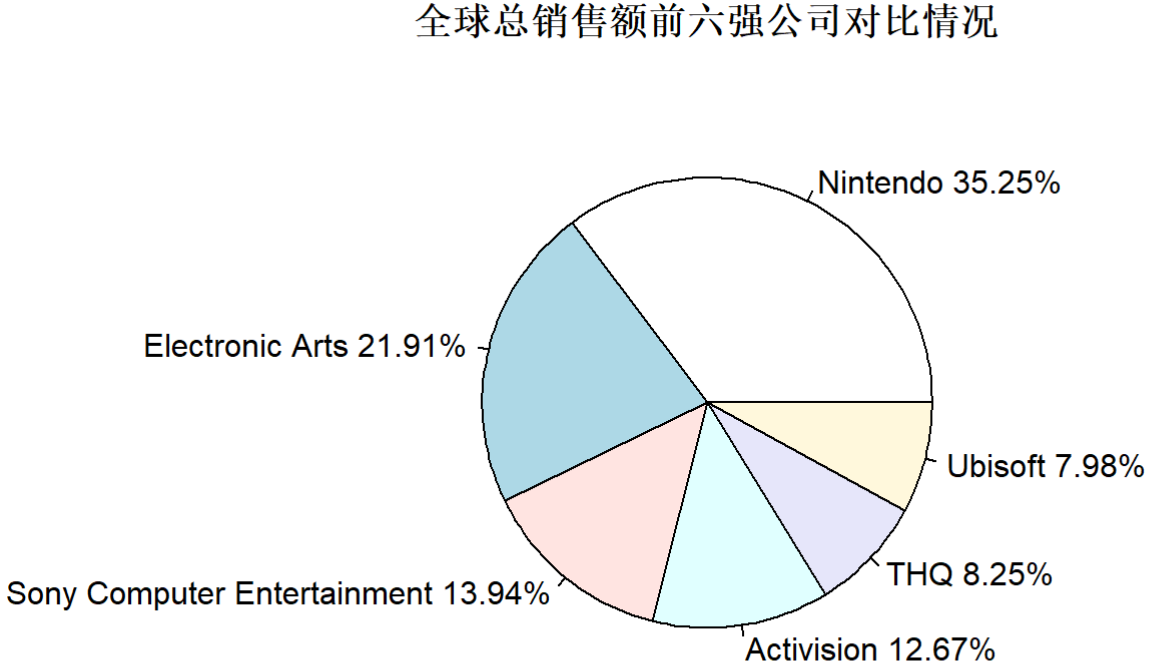
```
#3.不同发布商全球总销售额对比情况

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

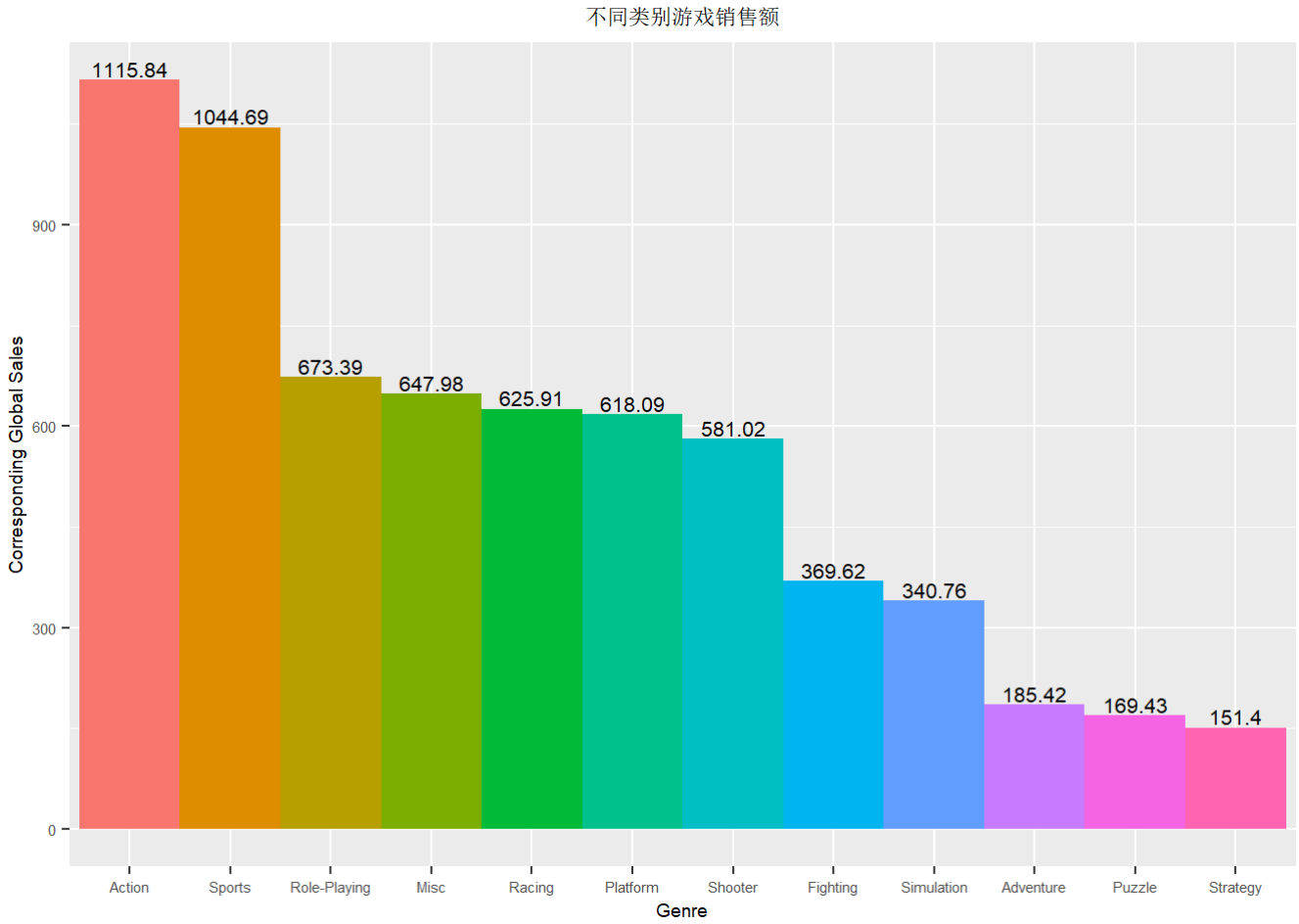
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
sales_by_group = aggregate(data1$Global_Sales,list(c(data1$Publisher)),sum)
sales_by_group=sales_by_group[order(-sales_by_group$x),]
sales_by_group_top6 = sales_by_group[1:6, ]

pie.number<- sales_by_group_top6$x
pie.labels<- sales_by_group_top6$Group.1
pie.pct <- round(100*sales_by_group_top6$x/sum(sales_by_group_top6$x), 2)
pie.labels <- paste(pie.labels," ",pie.pct,"%" , sep="")
pie(pie.number, pie.labels,main = '全球总销售额前六强公司对比情况')
```



结论： 在前6家大型游戏厂商中，任天堂占据了最大份额的收益，两家日本游戏厂家（Nintendo 和 Sony Computer Entertainment）在前6家大厂种占据几乎50%的份额，可见日本的游戏 产业十分强劲。

```
# 所有厂家不同类别游戏销售额对比情况

library(ggplot2)
sales_by_genre = aggregate(data1$Global_Sales,list(c(data1$Genre)),sum)
sales_by_genre = sales_by_genre[order(-sales_by_genre$x),]
genre = c(sales_by_genre$Group.1)
sales_genre = c(sales_by_genre$x)

sales_df = data.frame(sales_genre,genre)

sales_df$genre=factor(sales_df$genre,levels=genre)

par(pin=c(1,1))
ggplot(sales_df, aes(x=genre, y=sales_genre,fill=genre)) + geom_bar(stat="identity",width = 1)+geom_text(aes(label=sales_gen
re), vjust=-0.2, size = 2.8)+ labs(x='Genre',y='Corresponding Global Sales',title='不同类别游戏销售额')+theme(plot.title=elem
ent_text(hjust=0.5),text=element_text(size=7))+guides(fill=FALSE)
```




```
#与全球销售额前十的游戏做对比
head(data1[order(-data1$Global_Sales),'Genre'],n=10)
```

```
##  [1] "Sports"       "Racing"       "Sports"       "Role-Playing" "Platform"
##  [6] "Misc"         "Platform"     "Simulation"   "Racing"       "Role-Playing"
```

结论： 可以看出动作、运动以及射击类游戏是最受玩家欢迎的游戏类别，而策略类是受众最小的游戏类型。值得注意的是，这与全球销售额最大的10大游戏反映出的不同，任天堂上榜的游戏中，没有动作和射击类游戏，全球最畅销的游戏大多为运动，竞赛，和竞速。

```
#观察任天堂公司的热卖游戏类别
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.0.3
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
data2 =data.table(data1)
#data2[Publisher=='Nintendo',sum(Global_Sales),by = Genre]
Nintendo_df = data.frame(data2[Publisher=='Nintendo',sum(Global_Sales),by = Genre])
Nintendo_df[order(-Nintendo_df$V1),]
```

	Genre <chr>	V1 <dbl>
4	Platform	279.18
3	Role-Playing	218.87
1	Sports	192.31
5	Misc	156.68
2	Racing	122.35
9	Action	79.42
7	Puzzle	75.14
6	Simulation	62.12
8	Fighting	35.26

	Genre <chr>	V1 <dbl>
11	Shooter	33.05

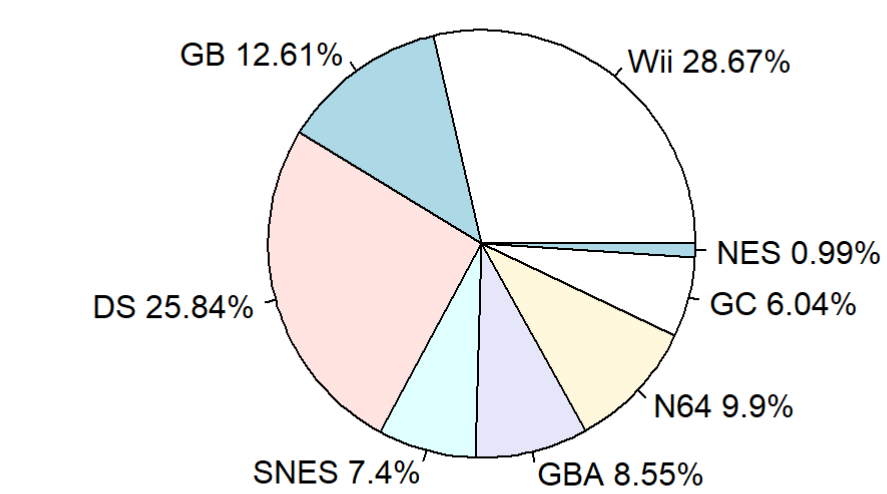
1-10 of 12 rows Previous **1** 2 Next

```
#Wii平台下销量游戏类别销量一览

Nintendo = data1[data1$Publisher=='Nintendo',]




library(data.table)
Nintendo_table = data.table(Nintendo)
Nintendo_table[Nintendo$Platform=='Wii',sum(Global_Sales), by = Genre]
```

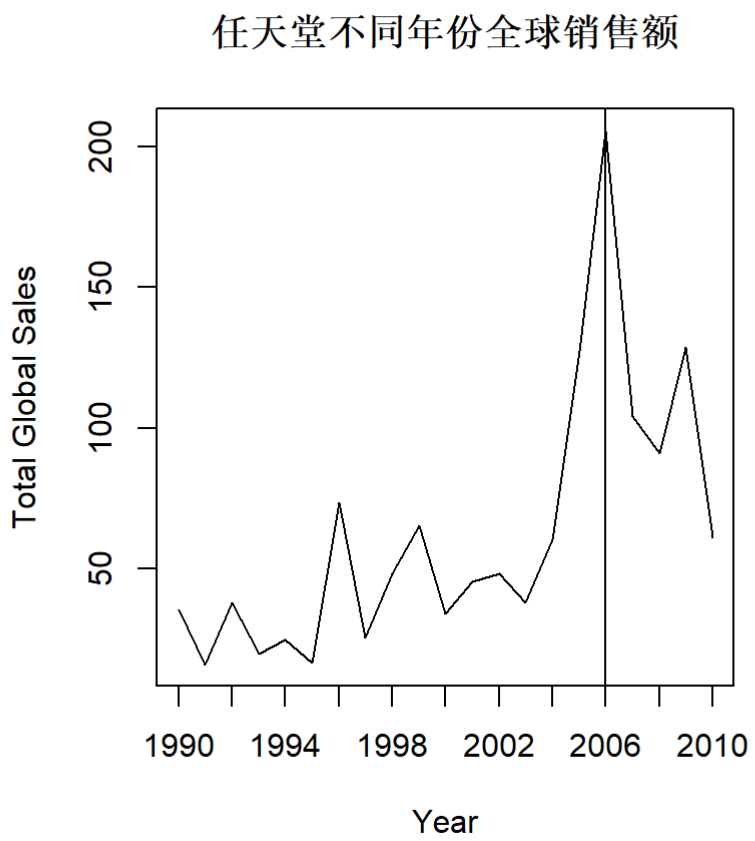
Genre <chr>	V1 <dbl>
Sports	169.19
Racing	37.28
Misc	53.35
Platform	64.69
Fighting	13.56
Action	9.84
Shooter	7.89
Simulation	6.21
Puzzle	3.64
Role-Playing	4.87

1-10 of 12 rows Previous **1** 2 Next

```
#Wii平台销售额最好的是运动类
#Nintendo旗下主机全球对Nintendo全球销售额贡献情况

Nintendo_df = data.frame(Nintendo_table[,sum(Global_Sales), by = Platform])
par(pin=c(3,3))
pie.number<- Nintendo_df$V1
pie.labels<- Nintendo_df$Platform
pie.pct <- round(100*Nintendo_df$V1/sum(Nintendo$Global_Sales),2)
pie.labels <- paste(pie.labels," ",pie.pct,"%" , sep="")
pie(pie.number, pie.labels,main = 'Nintendo旗下主机全球对Nintendo全球销售额贡献情况')
```



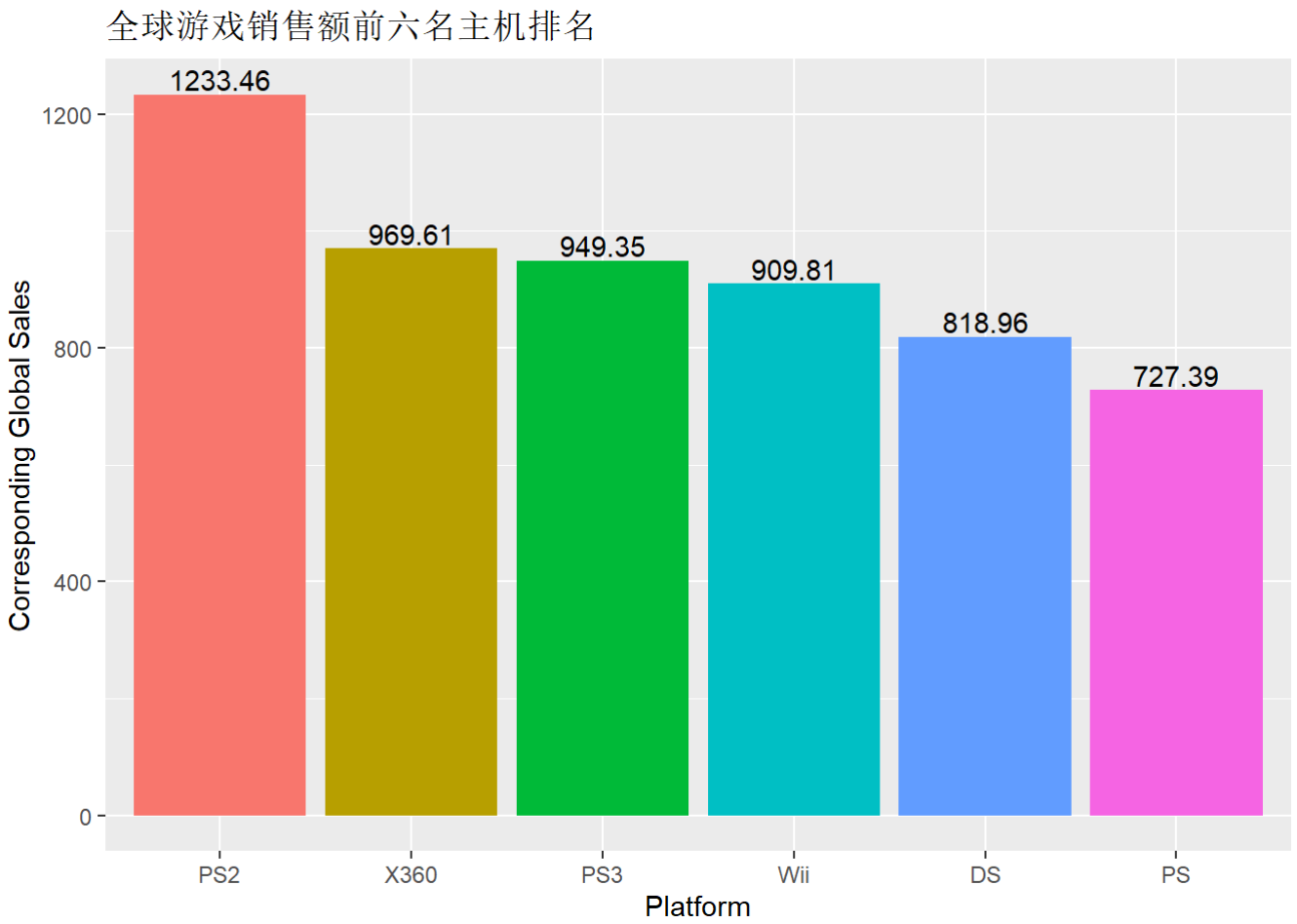

```
#任天堂不同年份总销售额对比情况
plot(sort(Nintendo[!duplicated(Nintendo$Year),'Year']),tapply(Nintendo$Global_Sales,Nintendo$Year,sum),type = 'l',xlab='Yea
r', ylab='Total Global Sales',main='任天堂不同年份全球销售额',xaxt="n")
axis(side=1, at=c(seq(from=1990,to=2010,by=2)))
abline(v=2006)
```




```
#5.不同平台总销售额对比情况
library(dplyr)


sort(tapply(data1$Global_Sales,data1$Platform,sum))
```

```
##    PCFX      GG     3DO    TG16      WS      NG     SCD      DC     NES     GEN
##    0.03    0.04    0.10    0.16    1.42    1.44    1.87   15.97   27.55   28.36
##     SAT      PC      GB      GC    SNES     N64      XB     PSP     GBA     PS3
##   33.59  159.32  188.01  197.14  200.05  218.21  252.09  262.51  313.56  491.25
##    X360      PS      DS     Wii     PS2
##  583.70  727.39  777.52  809.28 1232.99
```

```
platform_top6 = c('PS2','X360','PS3','Wii','DS','PS')

sales_top6 = c(1233.46,969.61,949.35,909.81,818.96,727.39)
top6 = data.frame(global_sales = sales_top6, platform = platform_top6)

top6$platform = factor(top6$platform, levels=c('PS2','X360','PS3','Wii','DS','PS'))

library(ggplot2)
ggplot(top6, aes(x=platform, y=global_sales,fill=platform)) + geom_bar(stat="identity")+geom_text(aes(label=global_sales), v
just=-0.2)+labs(x='Platform',y='Corresponding Global Sales',title='全球游戏销售额前六名主机排名')+guides(fill=FALSE)
```



结论： 任天堂在动作游戏领域的竞争效果并不理想，只在任天堂所有游戏类别的全球销售额占到第五名 销售额最多的游戏类别依次是:平台、角色扮演、竞 速（在之前的全球销售额排名中，Wii Sports系列的两部游戏和Super Mario Bros占据了全球销量前4） 任天堂非常善于使用Wii系列游戏主机，发售了许多竞速，运动,音乐相关的游戏，此类游戏 可以大大发挥出Wii平台动作感应的特点。而又得益于Wii平台可以多人互动的优势，上述游戏类别成了Wii平台玩家中意的游戏类型。而Wii主机也没有辜负任天堂的期许，在所有主机中为任天堂贡献21%的销售额，是任天堂旗下为任天堂盈利最多的主机。06年Wii的发售， 也让任天堂的全球销售额达到了顶峰，同时Wii在全球游戏主机的市场中也霸据了第四名的位置。