
Ciència de Dades
Tipologia i cicle de vida de les dades

Guillem LLuch Moll

PRACTICA 2

curs 2017-2018

Índex

1	Descripció del dataset	3
1.1	Les dades	3
1.2	Preguntes interessants	3
2	Neteja de dades	4
2.1	Selecció de les dades d'interès a analitzar	4
2.2	Gestió de les dades	4
3	Anàlisi de les dades	4
3.1	Selecció dels grups de dades	4
3.2	Normalització	6
3.2.1	ANOVAs	8
3.3	Normalització	9
3.4	Proves estadístiques	12
4	Representació	19
5	Resolució	42

1 Descripció del dataset

1.1 Les dades

El dataset que es tractarà és PC_Games de Hakob Sukiasyan (<https://www.kaggle.com/sukiasyan/pc-games>). No hi ha una descripció del dataset, però els noms dels camps són prou clars per tenir una certa idea de la informació que contenen.

No sabem de la qualitat de les dades, de la fiabilitat de les fonts utilitzades ni de la seva cobertura.

Cada fila representa la ventes d'un vídeo joc.

##	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales
## 1	Wii Sports	Wii	2006	Sports	Nintendo	41.36
## 2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08
## 3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.68
##	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count
## 1	28.96	3.77	8.45	82.53	76	51
## 2	3.58	6.81	0.77	40.24	NA	NA
## 3	12.76	3.79	3.29	35.52	82	73
##	User_Score	User_Count	Developer	Rating		
## 1	8	322	Nintendo	E		
## 2		NA				
## 3	8.3	709	Nintendo	E		

El dataset inicial té 16719 files i 16 columnes. Les columnes presents són

##	[1]	"Name"	"Platform"	"Year_of_Release"
##	[4]	"Genre"	"Publisher"	"NA_Sales"
##	[7]	"EU_Sales"	"JP_Sales"	"Other_Sales"
##	[10]	"Global_Sales"	"Critic_Score"	"Critic_Count"
##	[13]	"User_Score"	"User_Count"	"Developer"
##	[16]	"Rating"		

La llicència és public domain, per tant es pot reutilitzar lliurement.

1.2 Preguntes interessants

L'autor ha contestat a algunes preguntes interessants a <https://www.kaggle.com/sukiasyan/games-ranking> com ara les plataformes que han venut més jocs o la relació entre la puntuació de la crítica i dels usuaris.

La indústria de l'entreteniment digital, i dels vídeos jocs en particular, està en constant creixement i evolució. El 2010 globalment, suposaria un volum de negoci de 25'1 bilions de dolars (https://en.wikipedia.org/wiki/Video_game_industry)

Ens podríem interessar per la relació entre la puntuació de la crítica i les vendes, l'efecte del rating en les vendes o en l'opinió de la crítica, el gènere més cultivat per a cada editor (i a la inversa), buscar jocs que s'han venut bé en una zona geogràfica però malament en les altres, o com ha afectat l'aparició de cada consola en la venda de cada gènere, etc.

La pregunta que vull respondre és quin ha estat el **gènere preferit** segons cada **any de llançament**. Es vol respondre conjuntament i **per plataformes**. També es vol tenir un coneixement general de les vendes per gènere i plataforma.

Val a dir que estaria bé tenir moltes més dades com ara el número d'unitats venudes i el seu preu i la data en que s'han fet les vendes (encara que fos per mesos).

2 Neteja de dades

2.1 Selecció de les dades d'interès a analitzar

Ens interessen: plataforma, l'any, el gènere i global sales.

2.2 Gestió de les dades

Hi ha dues files que no tenen el nom ni el gènere (files 660 i 14247). Dels camps considerats, els que tenen problemes en les dades son *Year of release* ja que hi ha 269 jocs **sense l'any** de llançament especificat (1% del total). El valor que hi figura és N/A. La informació que falta es podria completar des de l'API de IGDB (<https://igdb.github.io/api/endpoints/game/>).

3 Anàlisi de les dades

3.1 Selecció dels grups de dades

La neteja que es farà consisteix en treure les columnes innecessàries, llevar files que no tenen el gènere definit (cal destacar que ja hi ha un gènere "misc") o bé no tenen l'any establert. A continuació es conservaran els casos que no tenen cap NA en cap camp.

```

#File read.r
#function which prepares the data
#return: a data set without missing values

getGames<-function(){

  games <- read.csv("input/games.csv")

  initialRows<-nrow(games)
  initialCols<-ncol(games)

  #Remove unsued columns
  games$NA_Sales<-NULL
  games$EU_Sales<-NULL
  games$JP_Sales<-NULL
  games$Other_Sales<-NULL

  games$User_Score<-NULL
  games$User_Count<-NULL

  games$Rating<-NULL

  games$Developer<-NULL

  games$Name<-NULL
  games$Publisher<-NULL
  games$Critic_Score<-NULL
  games$Critic_Count<-NULL

  #Delete rows without a proper Genre
  index<-which(games$Genre=="")
  games<-games[-index,]
  games$Genre<-factor(games$Genre)

  #Delete rows without the release year
  index<-which(games$Year_of_Release=="N/A")
  games<-games[-index,]

  #Set year of release as a date.
  #dates$Year<-as.Date(games$Year_of_Realese,"%Y")

```

```

#Set proper names
names(games)[2]<-"Year"
names(games)[4]<-"Sales"

games<-games[complete.cases(games), ]

message(initialRows= nrow(games)," rows and "
        ,initialCols = ncol(games)," cols deleted" )

games
}

games<-getGames()

## 271 rows and 12 cols deleted

```

Així al data frame *games* hi ha el conjunt de dades a considerar per a la pregunta.

El conjunt final considerat té 16448 files i 4 columnes , tots amb valors vàlids.

3.2 Normalització

Un dels tests més emprats per analitzar la variància és l'ANOVA. Serveix per saber si les mitjanes d'una variable són diferents segons una determinada agrupació. En aquest cas ens pot interessar per saber les vendes segons els grups definits per l'any, el gènere i la plataforma. Això sí, per a fer el test, les dades han d'estar distribuïdes normalment.

La distribució, però de les vendes no es aproximable per una normal, com es pot veure a la imatge

Es podria normalitzar aplicant el logaritme nepelià, però tampoc l'ajust seria prou bo.

```
qqnorm(log(games$Sales))
```

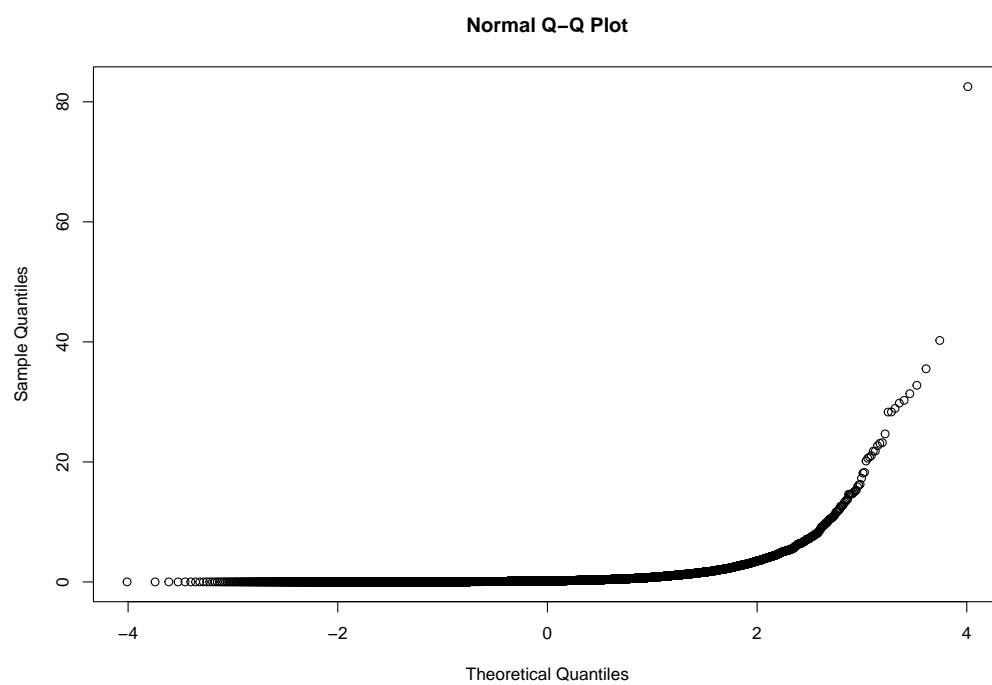
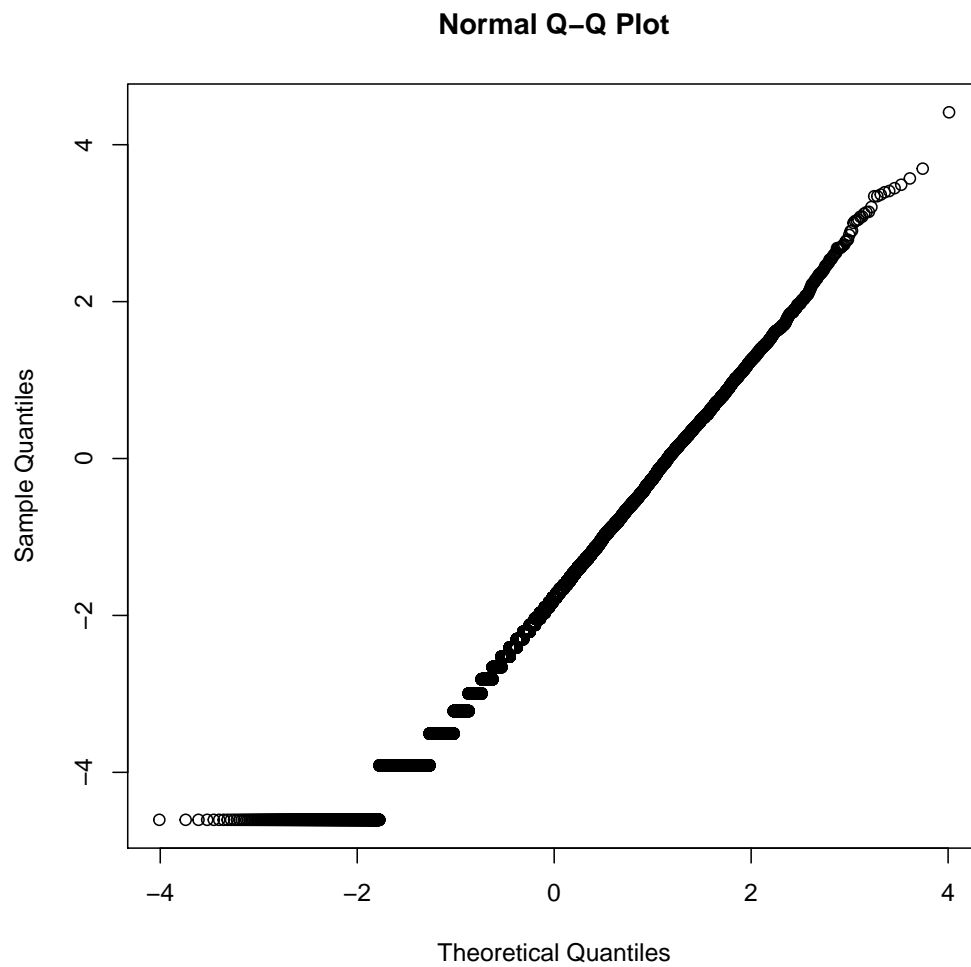


Figura 1: Gràfic q-qnorm per a les vendes



Encara que no segueixi una distribució normal, els resultats de molts test solen ser indicatius.

3.2.1 ANOVAs

Sales-Year

```
av.aov<-aov(games$Sales~games$Year)
summary(av.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
games\$Year	38	1016	26.727	11.27	<2e-16 ***
Residuals	16409	38931	2.373		

```
## ---
```



```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'últim paràmetre, el p valor, indica, donat que és gairebé zero, que les mitjanes de les vendes són molt diferents segons l'any.

Es pot veure la comparativa entre els diferents anys amb TukeyHSD(av.aov)

Plataforma-Vendes

```
pv.aov<-aov(games$Sales~games$Platform)
summary(pv.aov)

##              Df Sum Sq Mean Sq F value Pr(>F)
## games$Platform    30    1452    48.40   20.64 <2e-16 ***
## Residuals      16417   38494     2.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

També en aquest cas les mitjanes de les vendes són molt diferents segons la plataforma.

Gènere-Vendes

```
gv.aov<-aov(games$Sales~games$Genre)
summary(gv.aov)

##              Df Sum Sq Mean Sq F value Pr(>F)
## games$Genre     11     495    45.03   18.76 <2e-16 ***
## Residuals     16436   39451     2.40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Aquest, el gènere, també és un factor que condiciona les mitjanes.

3.3 Normalització

Hi ha relativament pocs jocs en els primers anys dels que es tenen constància, com es pot veure a la taula:

```
table(games$Year)

##
## 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994
```

##	9	46	36	17	14	14	21	16	15	17	16	41	43	60	121
##	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
##	219	263	289	379	338	350	482	829	775	762	939	1006	1197	1427	1426
##	2010	2011	2012	2013	2014	2015	2016	2017	2020	N/A					
##	1255	1136	653	544	581	606	502	3	1	0					

Donat que estem tractant els anys com un factor, podem crear un nou nivell que sigui 1980-1994. L'únic joc del 2020 és *Imagine: Makeup Artist* per a DS i la data correcta és 2009 (es corregirà). De l'any 2017, només hi ha tres jocs, la qual cosa deu significar que la recolecció de les dades s'ha acabat a principis d'aquell any. En tot cas, es descartaran.

```
#File clean.R
cleanGames<-function(games){
  #Change one error
  index<-which(games$Year=="2020")
  games[index,"Year"]="2009"

  #group the firsts years 1980-1994
  fy<-seq(1980,1994)
  fyears<-which(games$Year %in% fy)
  games$Year<-as.character(games$Year)
  games[fyears,"Year"]<-"1980-1994"

  #Delete the 3 games from 2017
  y2017=which(games$Year=="2017")
  games<-games[-y2017,]

  games$Year<-as.factor(games$Year)

  games
}

games<-cleanGames(games)
```

A l'hora d'agregar les vendes, el podem utilitzar és la proporció corresponent per a cada any de cada categoria i plataforma. Una altra dada que també podria ser d'utilitat és la resultant de dividir les vendes del gènere entre el nombre de jocs, que seria la rendabilitat mitjana per any i gènere.

En principi s'emprarà la proporció. Cal destacar que d'aquesta manera no importa saber les unitats en que estan les vendes.

```
library("reshape2")
games1<-dcast(games,games$Year+games$Platform~games$Genre, sum)

## Using Sales as value column: use value.var to override.

names(games1)[1]<-"Year"
names(games1)[2]<-"Device" #platform is a genre!!!

#Creating a new column with the sum of all sales
#in all genres per year.
games1$Sales<-rowSums(games1[,3:length(games1)])

#Proportion and then percentages
games2<-sweep(games1[,3:length(games1)],1,games1[,length(games1)],"/")
games2<-games2*100 #more readable

#Adding some columns
games2$Year<-games1$Year
games2$Device<-games1$Device
games2$Sales<-NULL

head(games2,4)

##      Action Adventure Fighting      Misc Platform      Puzzle      Racing
## 1  30.484001  0.4620538 0.8894536 4.135382 15.32864 15.76759 2.679912
## 2   0.000000  0.0000000 0.0000000 0.000000  0.00000 100.00000 0.000000
## 3 100.000000  0.0000000 0.0000000 0.000000  0.00000  0.00000 0.000000
## 4   3.396787 11.2993364 0.0000000 3.030038 32.68425 37.17255 3.973105
##  Role-Playing  Shooter Simulation  Sports Strategy      Year Device
## 1      0.00000 26.533441  0.5198106 3.199723 0.0000000 1980-1994 2600
## 2      0.00000  0.000000  0.0000000 0.000000 0.0000000 1980-1994   3D0
## 3      0.00000  0.000000  0.0000000 0.000000 0.0000000 1980-1994    DS
## 4      1.95599  1.047852  0.0000000 5.282920 0.1571778 1980-1994    GB

#Construim un altra dataframe només pels anys i gènere,
#independent de la plataforma

gamesg1<-dcast(games,games$Year ~ games$Genre, sum)
```

```
## Using Sales as value column: use value.var to override.

names(gamesg1)[1]<-"Year"
gamesg1$Sales<-rowSums(gamesg1[,2:length(gamesg1)])
gamesg2<-sweep(gamesg1[,2:(length(gamesg1)-1)],1,gamesg1[,length(gamesg1)],"/")
gamesg2<-gamesg2*100
gamesg2$Year<-gamesg1$Year

#A games1 hi ha les vendes per any i device vs genere en format absolut
#A games2 hi ha les vendes per any i device vs genere en percentatge
#A gamesg1 hi ha les vendes per any vs gènere en format absolut
#A games2 hi ha les vendes per any vs gènere en percentatge
```

3.4 Proves estadístiques

Una de les primeres proves d'interés és saber fins a quin punt les vendes totals i de cada gènere al llarg dels anys estan relacionades. Les dades són numèriques, decimals, i una de les millors mesures és la correlació. Ho podem fer amb `round(cor(gamesg1matrix),2)` o amb `round(cor(gamesg1matrix,method="spearman"),2)`

```
gamesg1matrix<-gamesg1[,2:length(gamesg1)]
round(cor(gamesg1matrix),2)
```

##	Action	Adventure	Fighting	Misc	Platform	Puzzle	Racing
## Action	1.00	0.62	0.21	0.69	0.07	0.16	0.28
## Adventure	0.62	1.00	0.61	0.73	0.50	0.64	0.46
## Fighting	0.21	0.61	1.00	0.30	0.64	0.64	0.53
## Misc	0.69	0.73	0.30	1.00	0.02	0.20	0.51
## Platform	0.07	0.50	0.64	0.02	1.00	0.93	0.21
## Puzzle	0.16	0.64	0.64	0.20	0.93	1.00	0.23
## Racing	0.28	0.46	0.53	0.51	0.21	0.23	1.00
## Role-Playing	0.65	0.56	0.22	0.65	0.16	0.13	0.27
## Shooter	0.85	0.62	0.08	0.53	0.23	0.31	0.02
## Simulation	0.53	0.62	0.34	0.81	0.02	0.25	0.71
## Sports	0.61	0.69	0.46	0.82	0.17	0.26	0.46
## Strategy	0.05	0.34	0.36	0.37	0.02	0.03	0.45
## Sales	0.79	0.90	0.60	0.82	0.50	0.60	0.55
##	Role-Playing	Shooter	Simulation	Sports	Strategy	Sales	
## Action	0.65	0.85	0.53	0.61	0.05	0.79	

## Adventure	0.56	0.62	0.62	0.69	0.34	0.90
## Fighting	0.22	0.08	0.34	0.46	0.36	0.60
## Misc	0.65	0.53	0.81	0.82	0.37	0.82
## Platform	0.16	0.23	0.02	0.17	0.02	0.50
## Puzzle	0.13	0.31	0.25	0.26	0.03	0.60
## Racing	0.27	0.02	0.71	0.46	0.45	0.55
## Role-Playing	1.00	0.59	0.28	0.58	0.28	0.69
## Shooter	0.59	1.00	0.29	0.43	-0.16	0.71
## Simulation	0.28	0.29	1.00	0.64	0.28	0.68
## Sports	0.58	0.43	0.64	1.00	0.25	0.81
## Strategy	0.28	-0.16	0.28	0.25	1.00	0.25
## Sales	0.69	0.71	0.68	0.81	0.25	1.00

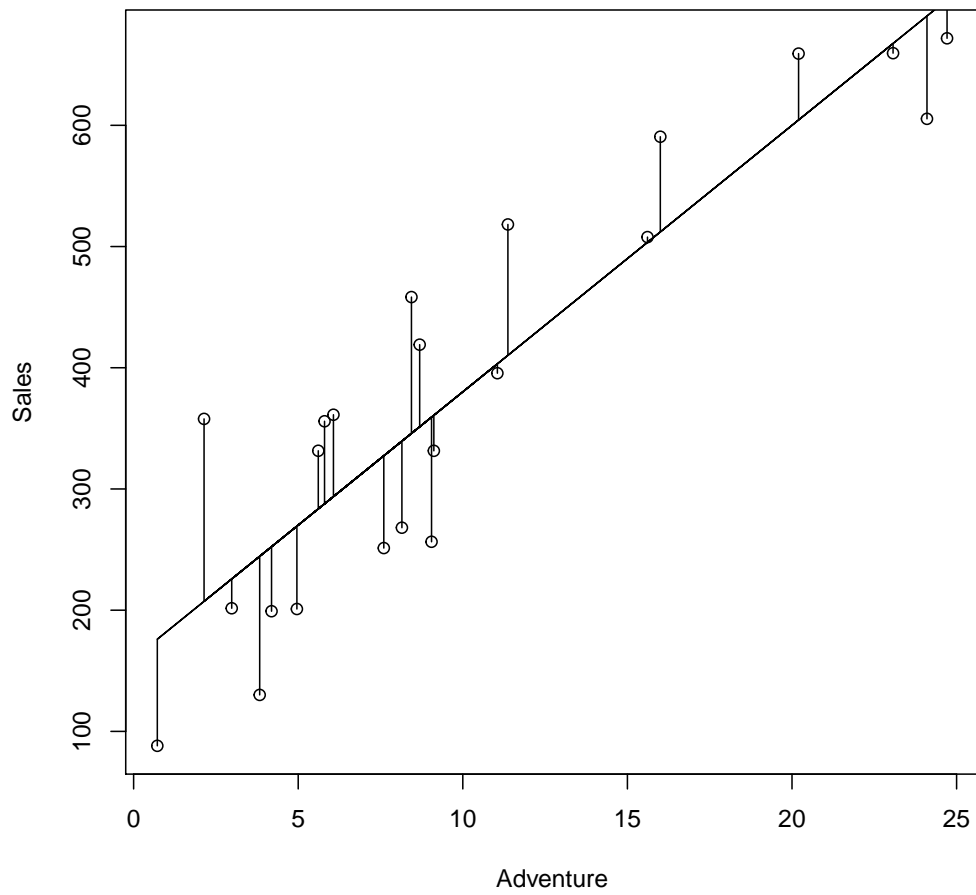
El més interessant dels resultats és que no tots els gèneres estan fortament relacionats linealment.

De la matriu de correlacions podem veure agrupacions de gèneres que segueixen un patró de vendes similars: shooter - action, misc-simulation, misc-sports, platform - puzzle. Si el gènere fos multivaluat, hi hauria molts jocs que serien alhora shooter i action o plataforma i puzzle, però també es podrien esperar altres combinacions que no es donen com ara Role-playing i adventure o sports i racing.

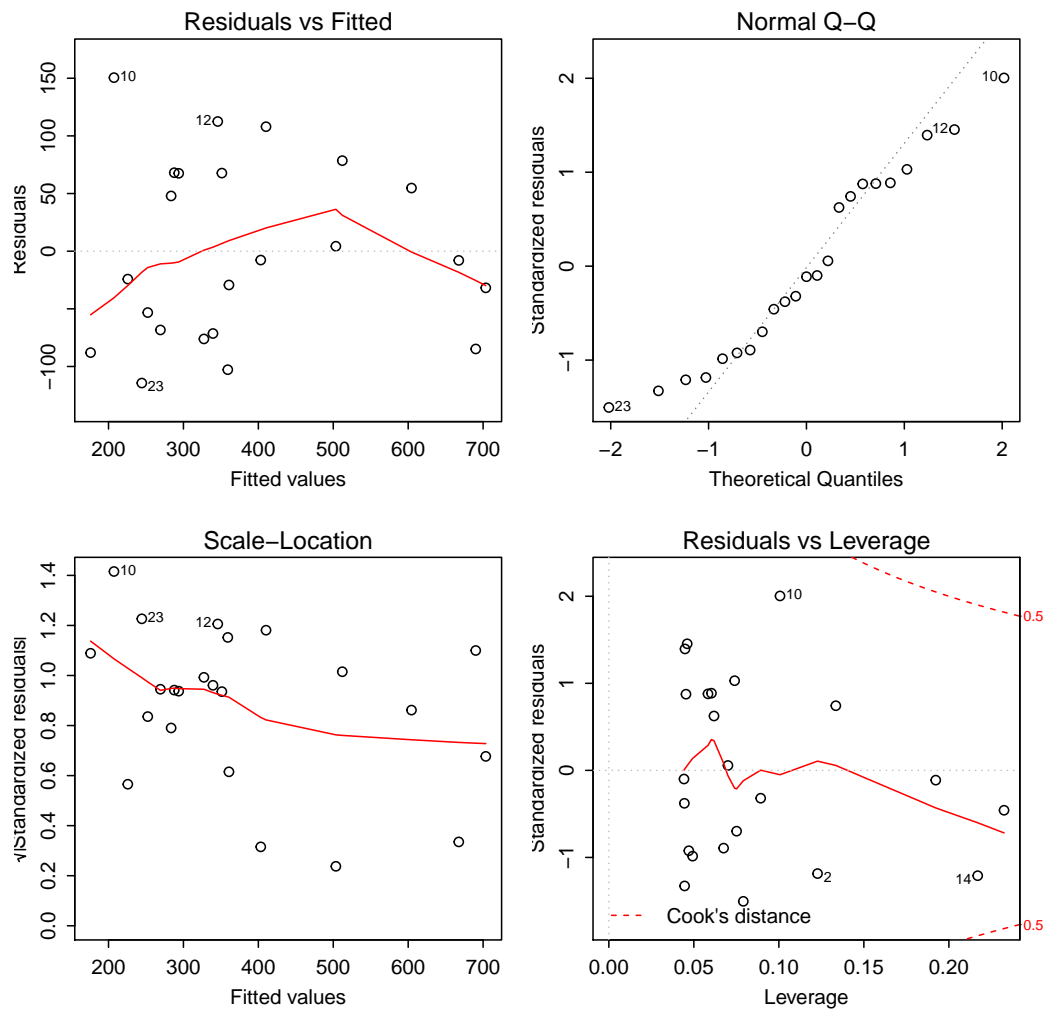
Una de les columnes o files, segons com es miri, que mereixen més atenció és la de Sales, on hi ha el coeficient de correlació amb cada un dels gèneres. Segons podem veure les vendes totals les podríem predir a partir de les d'aventura, misc o sports, però no d'estratègia. També es podria fer a la inversa i tractar de predir quines han estat les vendes d'un gènere a partir de les vendes totals.

Podem analitzar amb més detall la relació entre el gènere aventura i sales. Per fer-ho, la regressió lineal. Podem veure l'adjust:

```
attach(gamesg1)
lm.sales<-lm(Sales~Adventure)
plot(Adventure,Sales)
lines(Adventure,fitted(lm.sales))
segments(Adventure,fitted(lm.sales),+Adventure,Sales)
```



```
par(mfrow=c(2,2), mex=0.6)  
plot(lm.sales)
```



La fila més enfora de la seva predicció és la 10.

```
subset(gamesg1, select=c(3,14))
```

```
##      Adventure Sales
## 1      23.07 659.53
## 2       0.72  88.11
## 3       4.19 199.15
## 4       4.96 200.98
## 5       9.05 256.45
## 6       7.60 251.25
## 7       2.98 201.58
## 8       9.12 331.47
## 9      11.05 395.51
```

```
## 10      2.14 357.80
## 11      8.69 419.05
## 12      8.44 458.31
## 13     11.37 518.22
## 14     24.10 605.37
## 15     24.71 671.79
## 16     20.20 659.17
## 17     16.00 590.59
## 18     15.61 507.79
## 19      5.80 355.84
## 20      6.07 361.24
## 21      5.61 331.51
## 22      8.15 268.05
## 23      3.83 130.10
```

Si comparam la fila 10 amb la 7 podem veure que, a igual vendes de jocs d'aventura, les vendes totals han estat prou diferents. El resultat inusual és el de la fila 10 que correspon al 2003:

```
##      Year Action Adventure Fighting Misc Platform Puzzle Racing
## 10 2003  67.93      2.14    23.73 23.82    42.89   2.42  52.19
##      Role-Playing Shooter Simulation Sports Strategy Sales
## 10      30.26   27.13    21.19  56.11      7.99 357.8
```

També podríem fer un model lineal emprant més d'un gènere per predir les vendes totals:

```
mlm.sales<-lm(Sales~Action+Adventure+Misc+
              `Role-Playing`+Shooter+Simulation+Sports)
anova(mlm.sales)

## Analysis of Variance Table
##
## Response: Sales
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## Action         1 421626  421626 150.4802 3.198e-09 ***
## Adventure       1 179587  179587  64.0953 8.532e-07 ***
## Misc           1  10074   10074   3.5954 0.07737 .
## `Role-Playing`  1   2776    2776   0.9908 0.33534
## Shooter        1    401     401   0.1431 0.71052
## Simulation      1   2613    2613   0.9325 0.34953
```



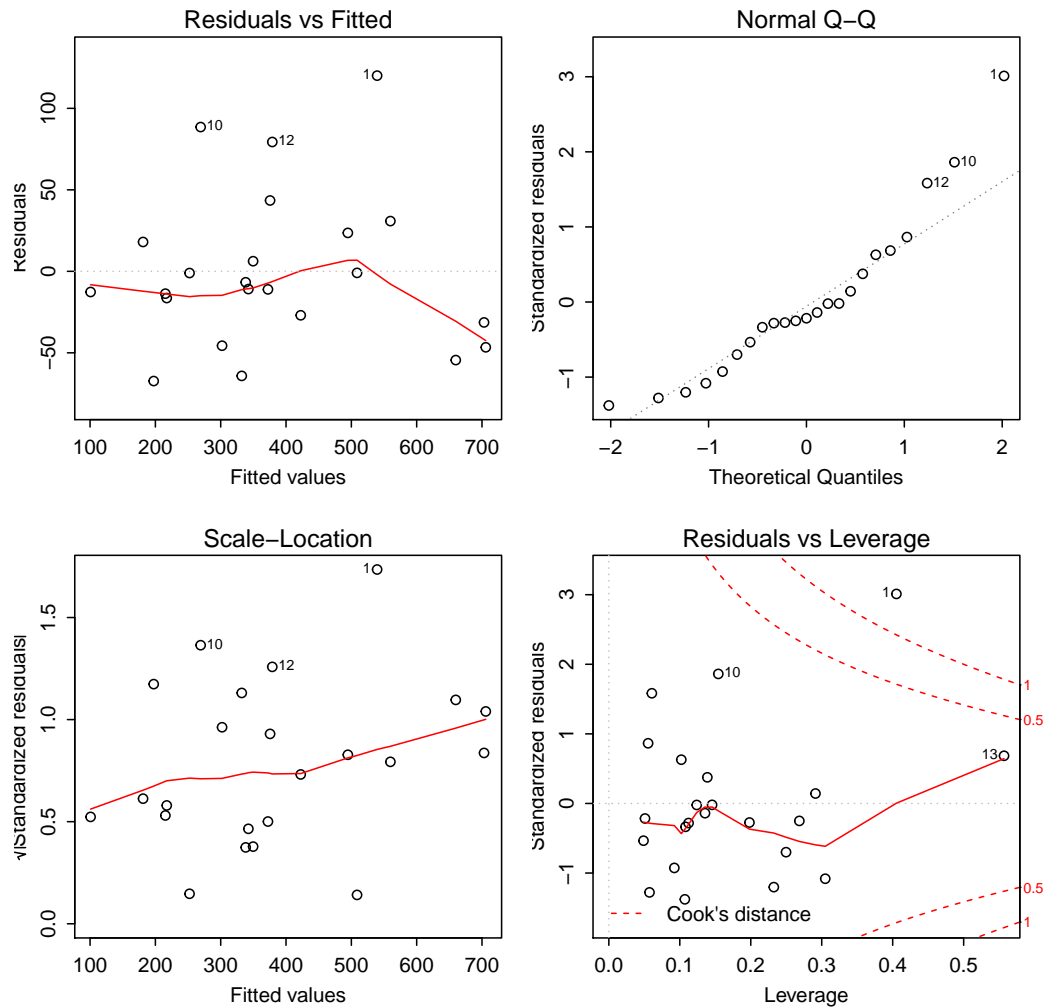
```
## Sports          1  15060   15060   5.3750   0.03496 *
## Residuals       15  42028    2802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Del resultat s'extreu que els únics gèneres necessaris són Action, Adventure i Sports

```
mlm.sales<-lm(Sales~Action+Adventure+Sports)
anova(mlm.sales)

## Analysis of Variance Table
##
## Response: Sales
##          Df Sum Sq Mean Sq  F value    Pr(>F)
## Action      1 421626   421626 157.7099 1.197e-10 ***
## Adventure    1 179587   179587  67.1747 1.168e-07 ***
## Sports       1  22156    22156   8.2876 0.009615 **
## Residuals   19  50795     2673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

par(mfrow=c(2,2), mex=0.6)
plot(mlm.sales)
```



Si enlloc de mirar el valors absoluts, miram el percentatges,

```
gamesg2matrix<-gamesg2[,1:(length(gamesg2)-1)]
round(cor(gamesg2matrix),2)
```

##	Action	Adventure	Fighting	Misc	Platform	Puzzle	Racing
## Action	1.00	-0.05	-0.66	-0.02	-0.62	-0.47	-0.50
## Adventure	-0.05	1.00	-0.28	0.12	-0.04	0.29	-0.12
## Fighting	-0.66	-0.28	1.00	-0.26	0.49	0.16	0.30
## Misc	-0.02	0.12	-0.26	1.00	-0.38	-0.04	-0.07
## Platform	-0.62	-0.04	0.49	-0.38	1.00	0.78	0.19
## Puzzle	-0.47	0.29	0.16	-0.04	0.78	1.00	0.00
## Racing	-0.50	-0.12	0.30	-0.07	0.19	0.00	1.00
## Role-Playing	-0.15	-0.28	0.30	-0.27	-0.08	-0.45	0.04

## Shooter	0.72	0.14	-0.61	-0.20	-0.47	-0.29	-0.68
## Simulation	-0.26	0.00	-0.02	0.53	-0.12	0.10	0.48
## Sports	-0.01	0.11	-0.14	0.26	-0.36	-0.24	-0.03
## Strategy	-0.62	-0.15	0.73	-0.09	0.25	-0.06	0.45
##	Role-Playing	Shooter	Simulation	Sports	Strategy		
## Action	-0.15	0.72	-0.26	-0.01	-0.62		
## Adventure	-0.28	0.14	0.00	0.11	-0.15		
## Fighting	0.30	-0.61	-0.02	-0.14	0.73		
## Misc	-0.27	-0.20	0.53	0.26	-0.09		
## Platform	-0.08	-0.47	-0.12	-0.36	0.25		
## Puzzle	-0.45	-0.29	0.10	-0.24	-0.06		
## Racing	0.04	-0.68	0.48	-0.03	0.45		
## Role-Playing	1.00	-0.04	-0.32	-0.24	0.45		
## Shooter	-0.04	1.00	-0.41	-0.07	-0.61		
## Simulation	-0.32	-0.41	1.00	-0.03	0.06		
## Sports	-0.24	-0.07	-0.03	1.00	-0.16		
## Strategy	0.45	-0.61	0.06	-0.16	1.00		

llavors, veiem que no tenim correlació (la que hi ha deu ser deguda a l'atzar)

4 Representació

Podem veure un grafic del gènere més popular per any.

```
#Take the position of the greatest value for each row
index<-apply(gamesg2matrix,1,which.max)

#create a new dataframe to save the results
yres<-as.data.frame(matrix(3,23))

#Get the names of the genre
yres$genre<-names(gamesg2matrix)[index]

#Get the values
yres$value<-apply(gamesg2matrix,1,max)
yres$sv<-paste(round(yres$value,0),"%",sep="")

#Get the years
yres$year<-gamesg2$Year
```

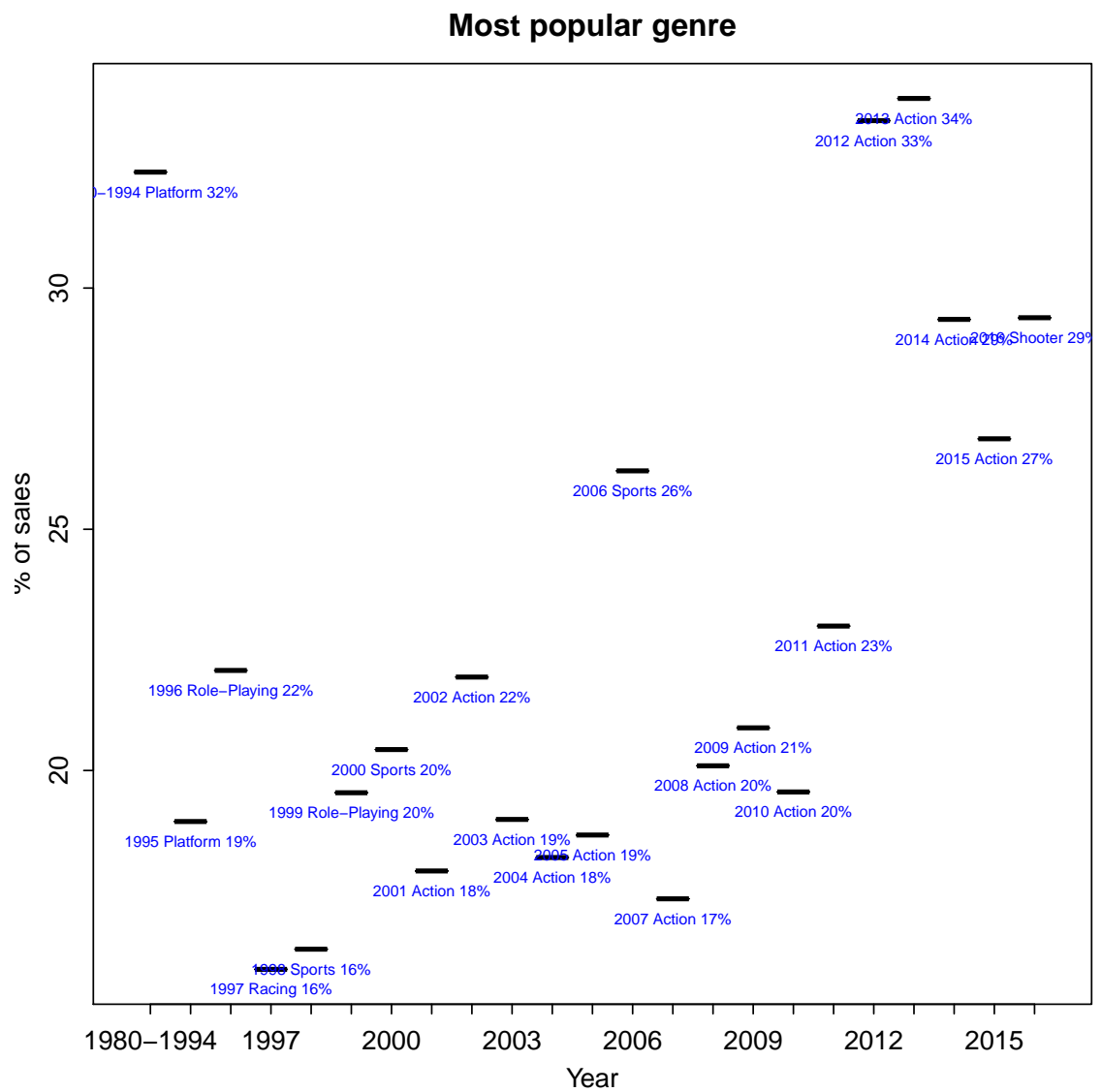
```

#Concat year and platform
yres$label<-paste(yres$year,yres$genre, sep=" ")
yres$label<-paste(yres$label,yres$sv,sep=" ")

yres$V1<-NULL

par(mfrow=c(1,1), mex=0.6)
plot(yres$year,yres$value, xlab="Year",ylab="% of sales",main="Most popular genre",
text(yres$year,yres$value,yres$label, cex=0.6, pos=1, col="blue")

```



I aquí tenim els gràfics de cada una de les consoles principals:

```
mSales=0.5
par(mfrow=c(1,1), mex=0.6)
#Delete rows with a low sales (~24 rows)
games2<-games2[which(games1$Sales>mSales),]

for (device in levels(games2$Device)){
  nr=nrow(games2[which(games2$Device==device ),])
  if (nr>1){
    plaf=as.data.frame(matrix(1,nr))
    plaf<-(games2[which(games2$Device==device ),])
    if(nrow(plaf)>1){
      plafmatrix<-plaf[,1:(length(games2)-2)]
      #Take the position of the greatest value for each row
      index<-apply(plafmatrix,1,which.max)

      #create a new dataframe to save the results
      yres<-as.data.frame(matrix(3,nr))

      #Get the names of the genre
      yres$genre<-names(plafmatrix)[index]

      #Get the values
      yres$value<-apply(plafmatrix,1,max)
      yres$sv<-paste(round(yres$value,0),"%",sep=" ")

      #Get the years
      yres$year<-factor(plaf$Year)

      #Concat year and platform
      yres$label<-paste(yres$year,yres$genre, sep=" ")
      yres$label<-paste(yres$label,yres$sv,sep=" ")

      yres$V1<-NULL

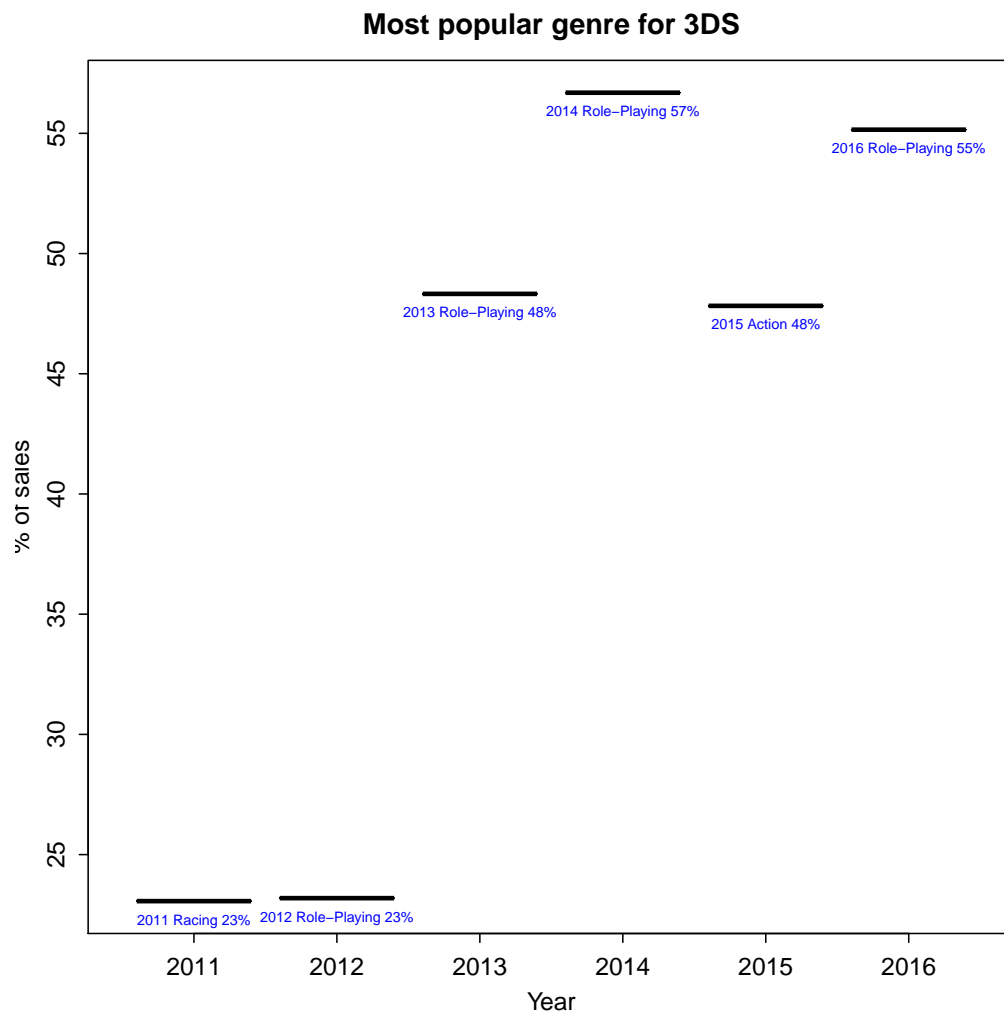
      tit=paste("Most popular genre for",device, sep=" ")
      fil=paste("img/",device,sep="")
      fil=paste(fil,".png",sep="")
      #png(fil,width = 800, height = 600)
```

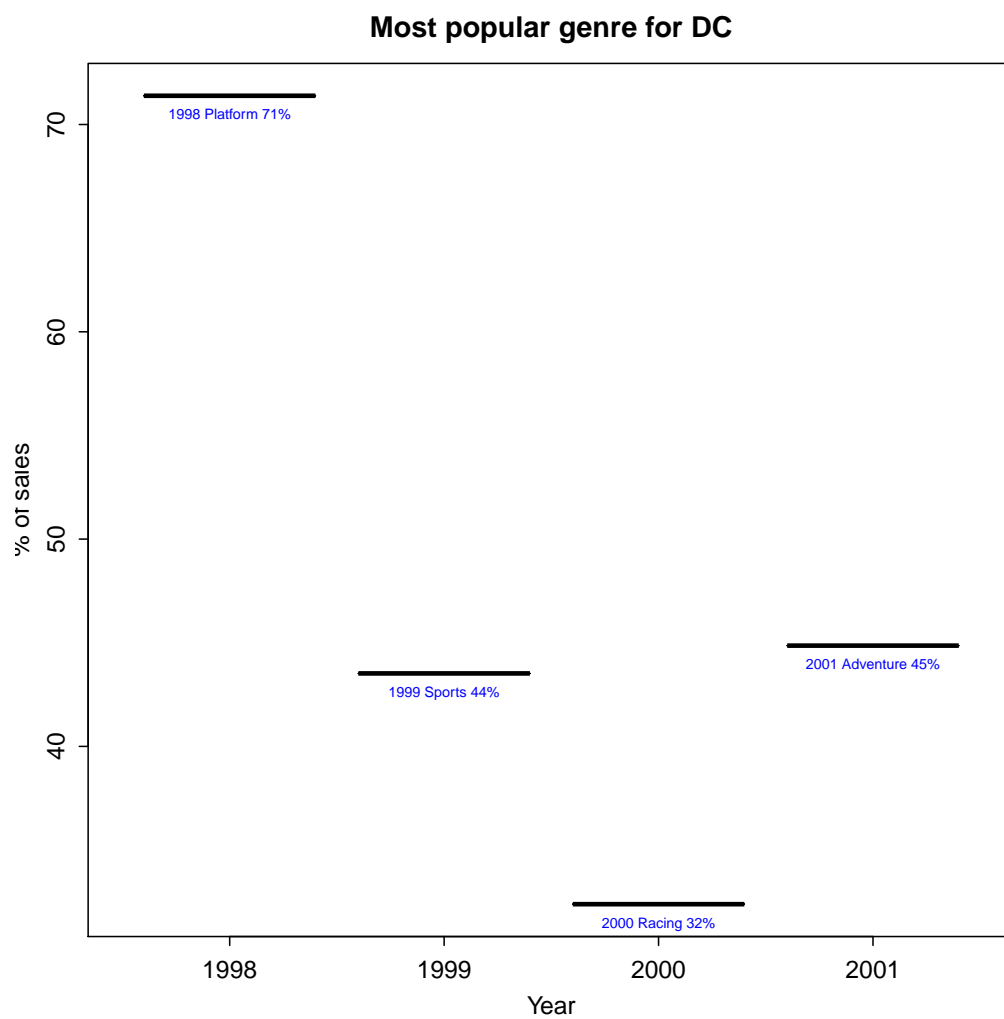
```

par(mfrow=c(1,1), mex=0.6)
plot(yres$year,yres$value, xlab="Year",ylab="% of sales",main=tit)
text(yres$year,yres$value,yres$label, cex=0.6, pos=1, col="blue")

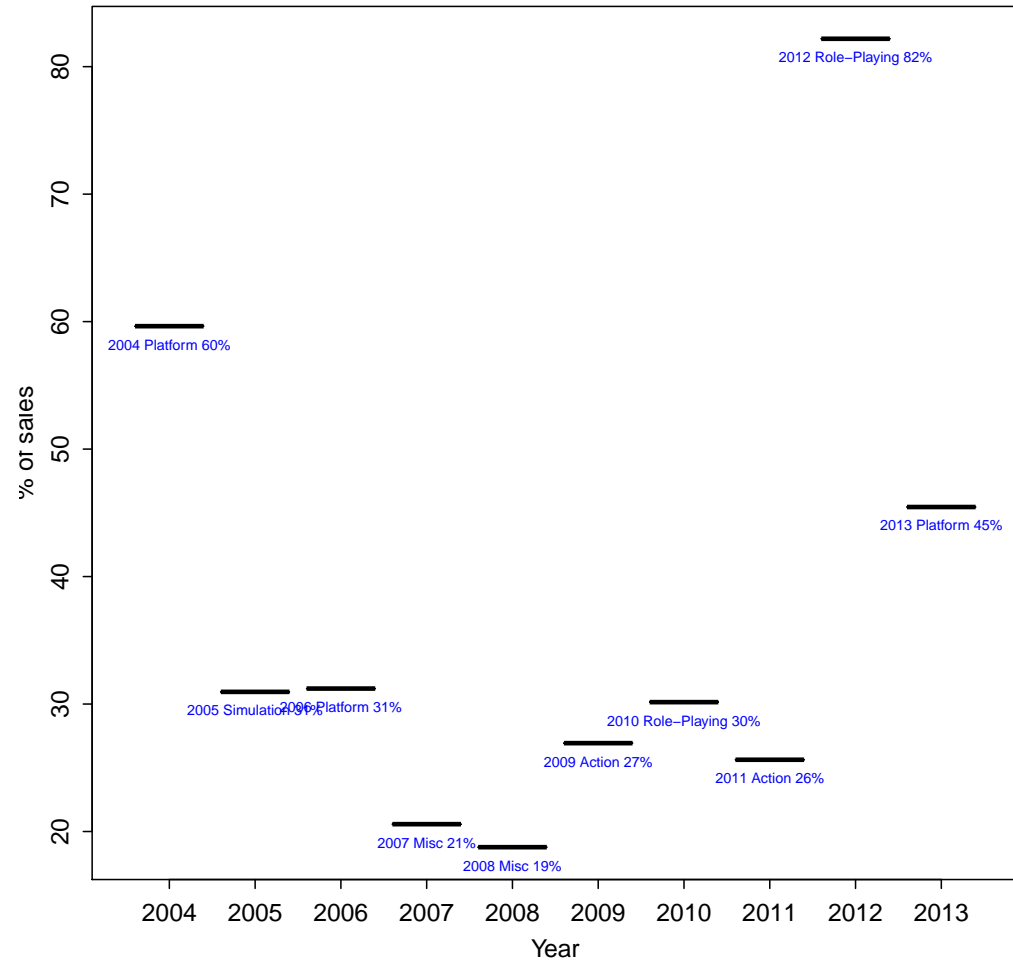
#dev.off()
}
}
}

```

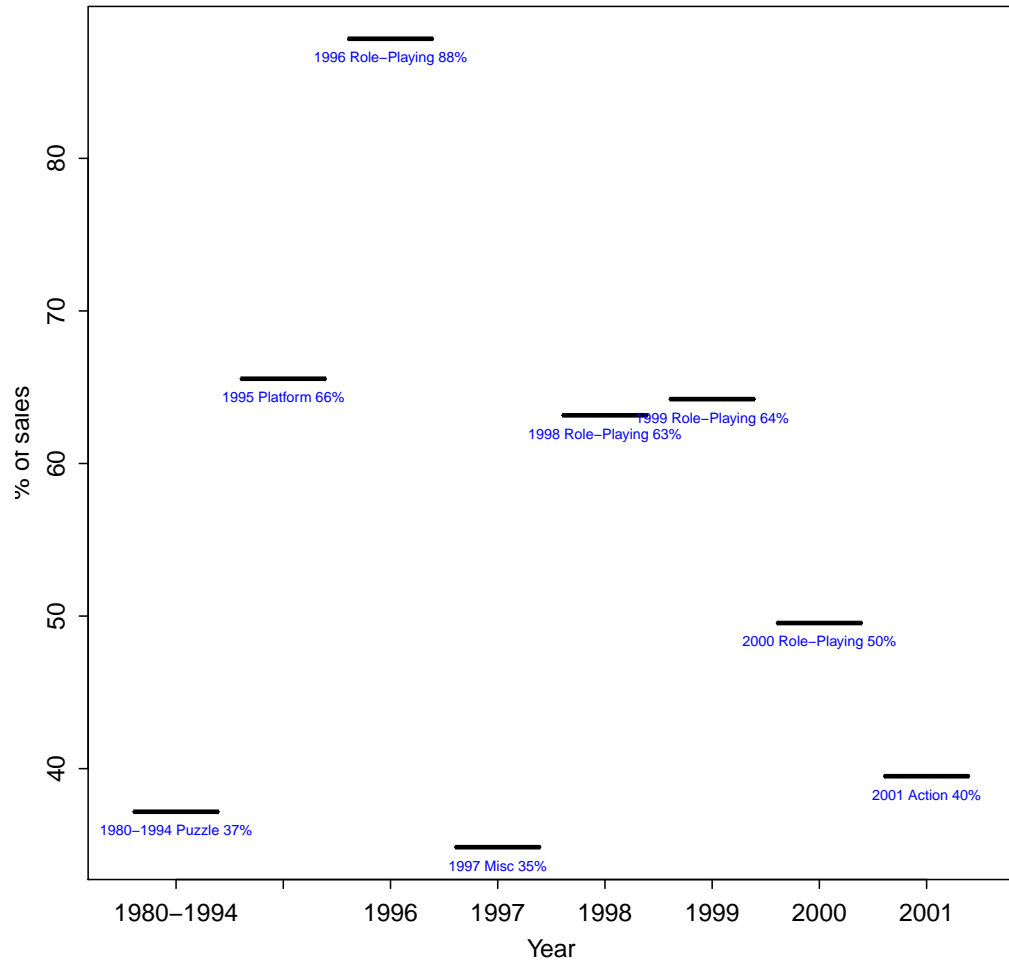




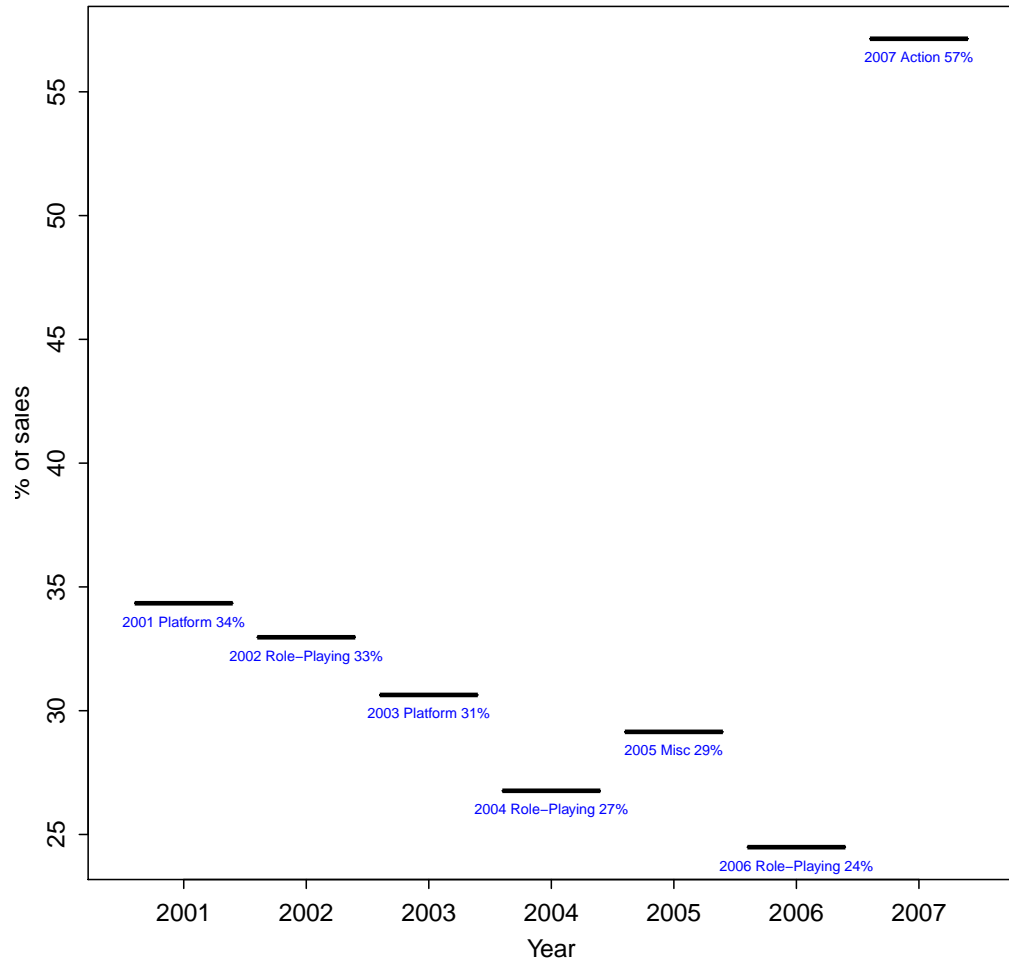
Most popular genre for DS



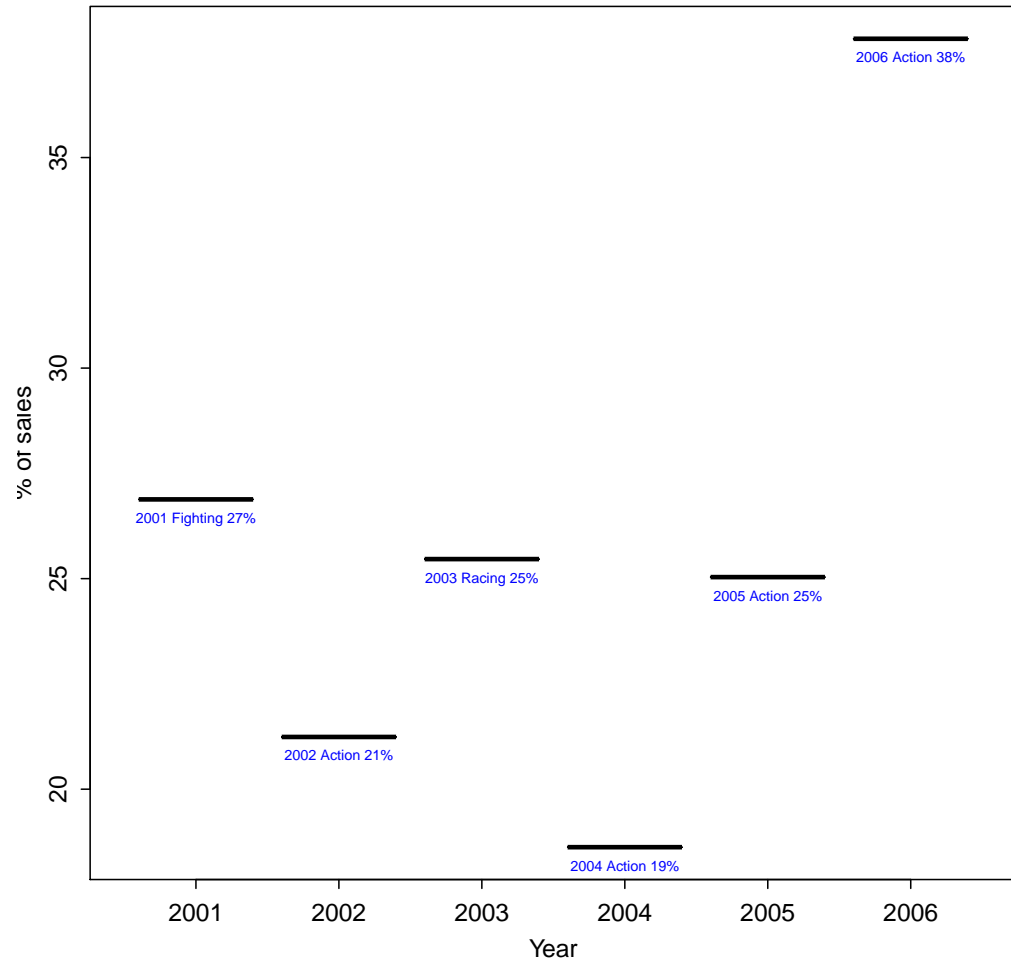
Most popular genre for GB



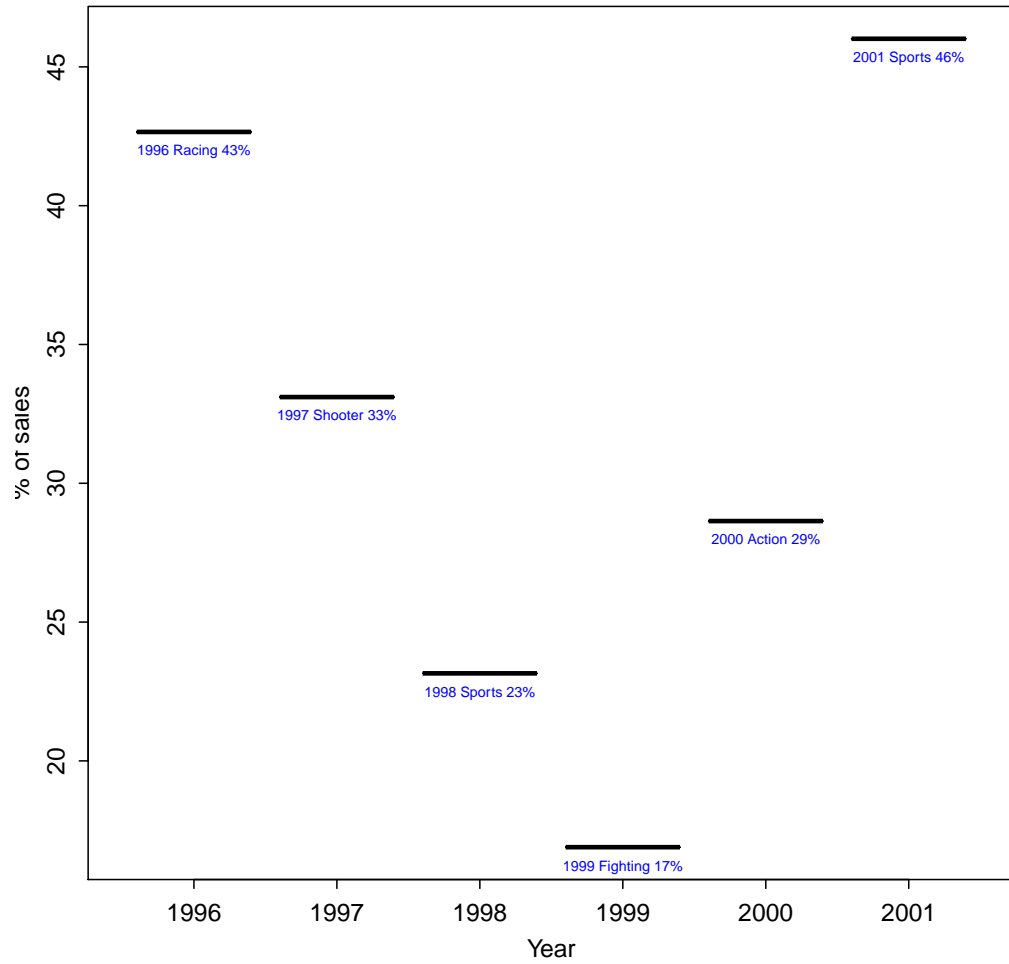
Most popular genre for GBA

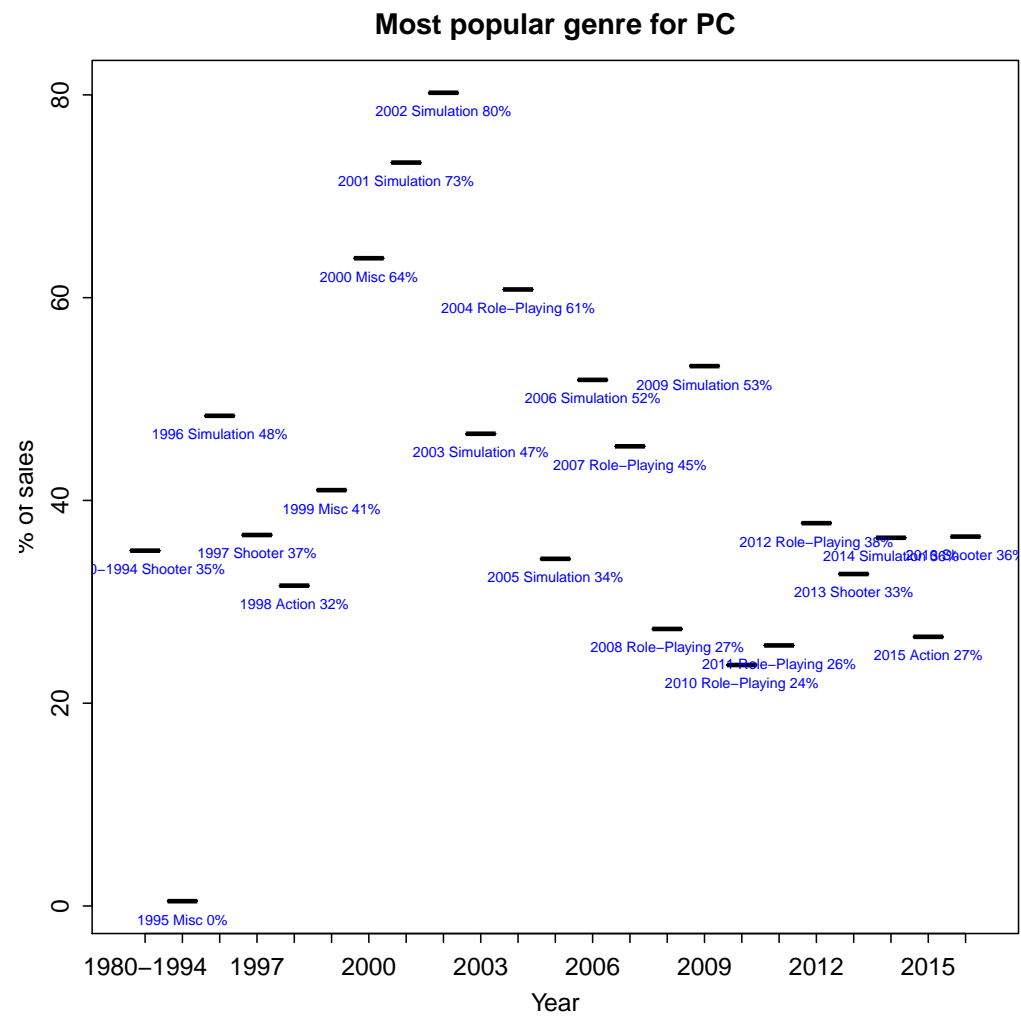


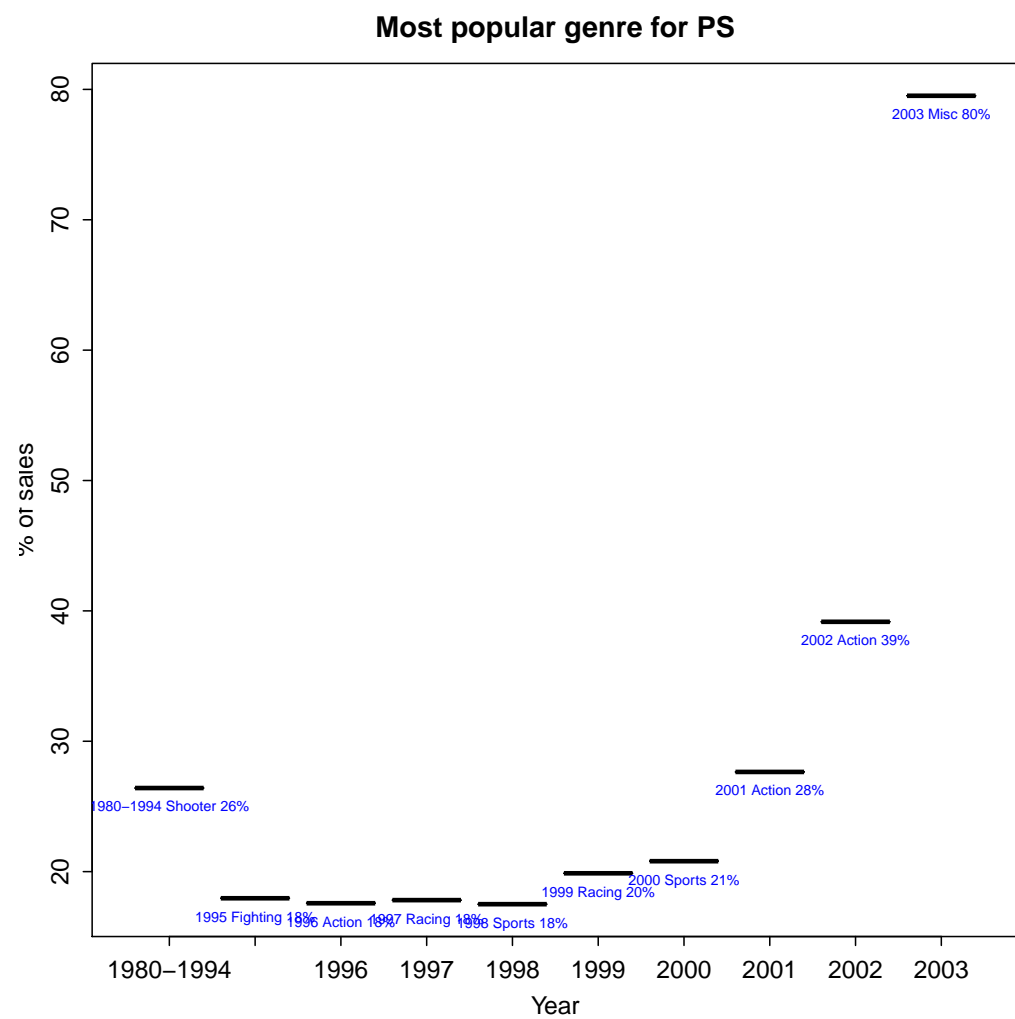
Most popular genre for GC



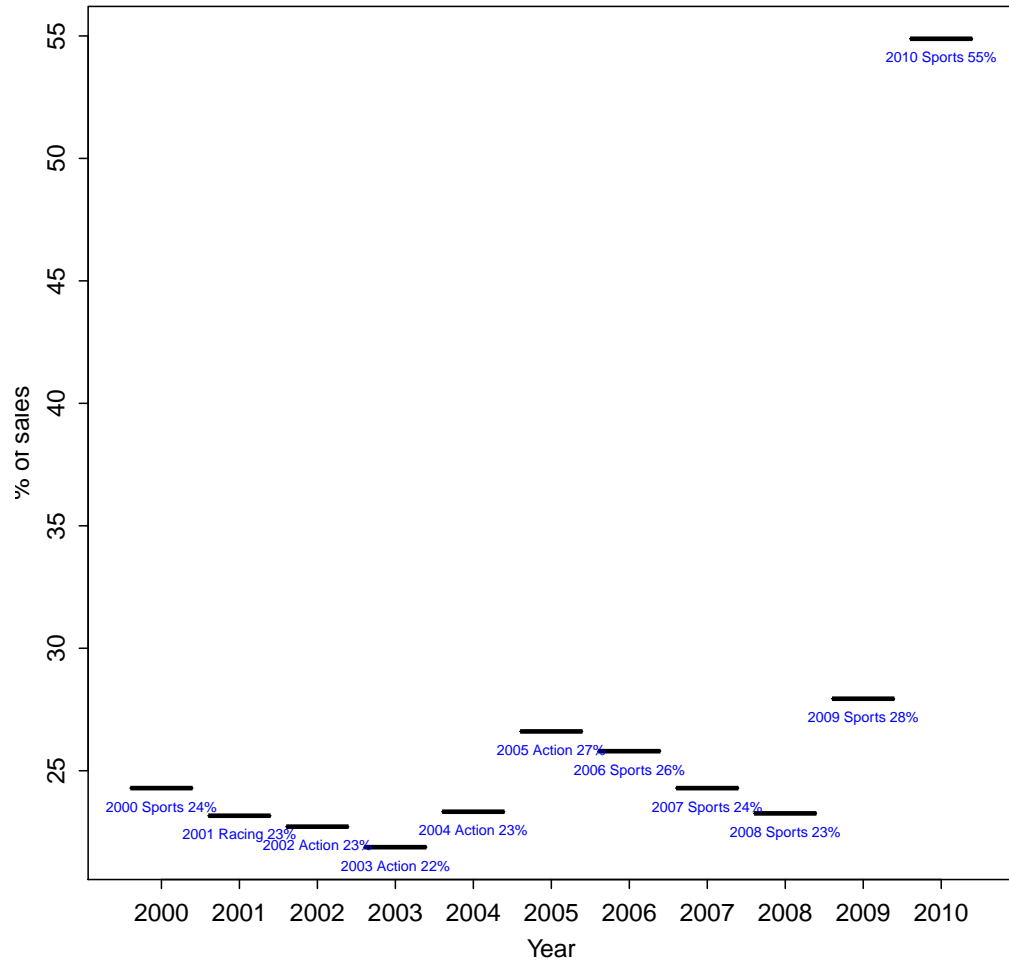
Most popular genre for N64



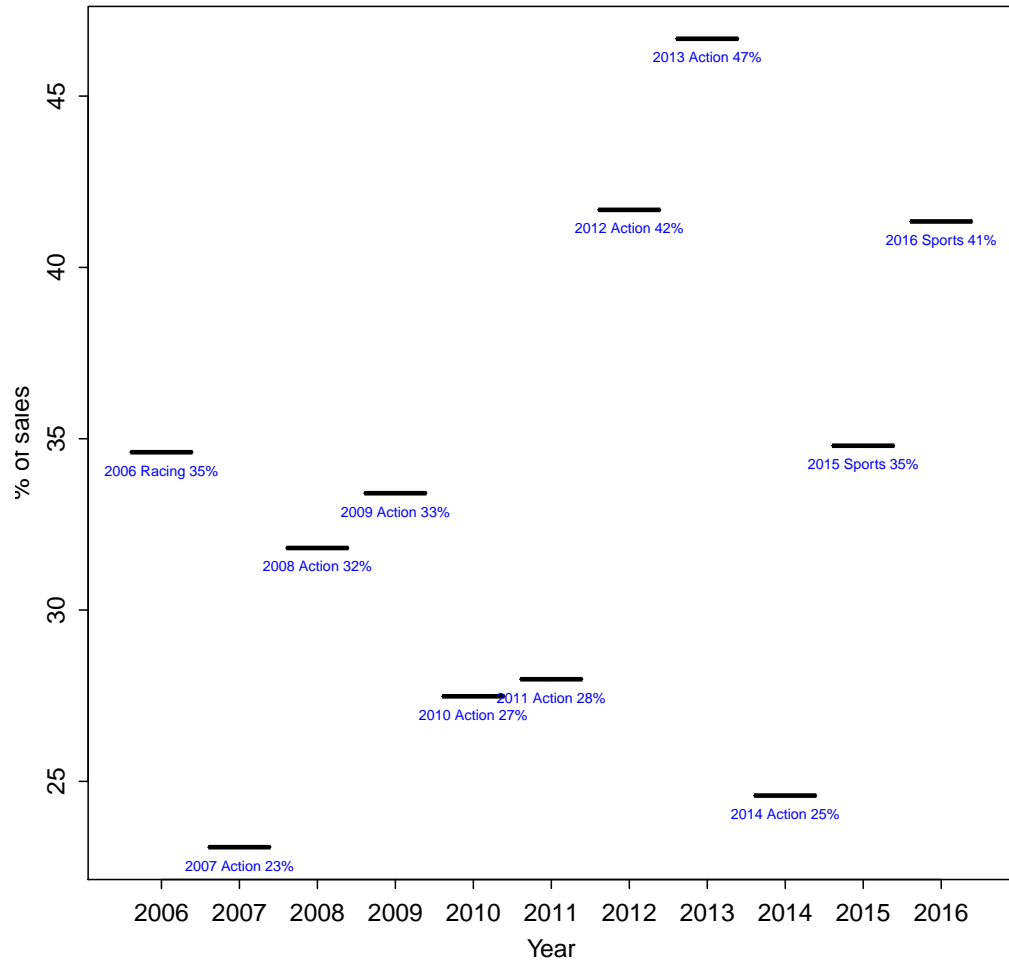


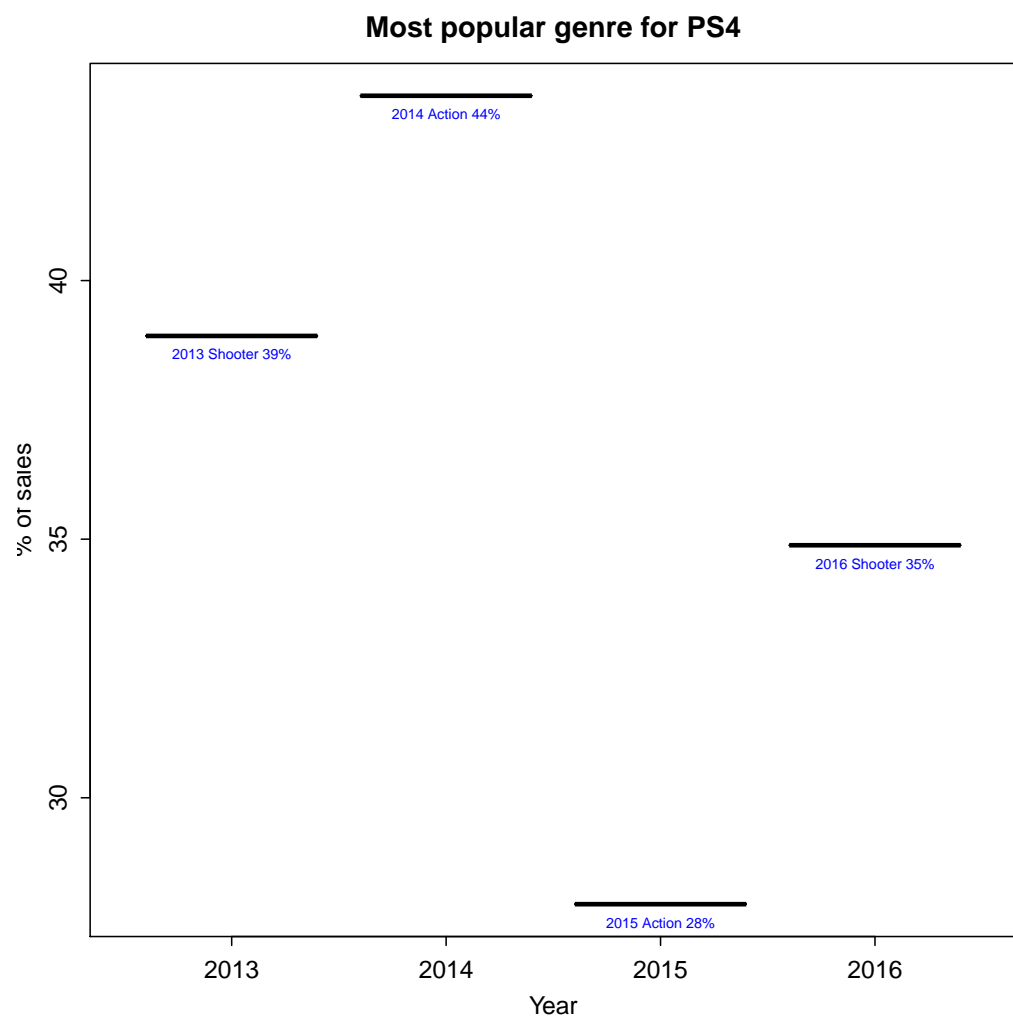


Most popular genre for PS2

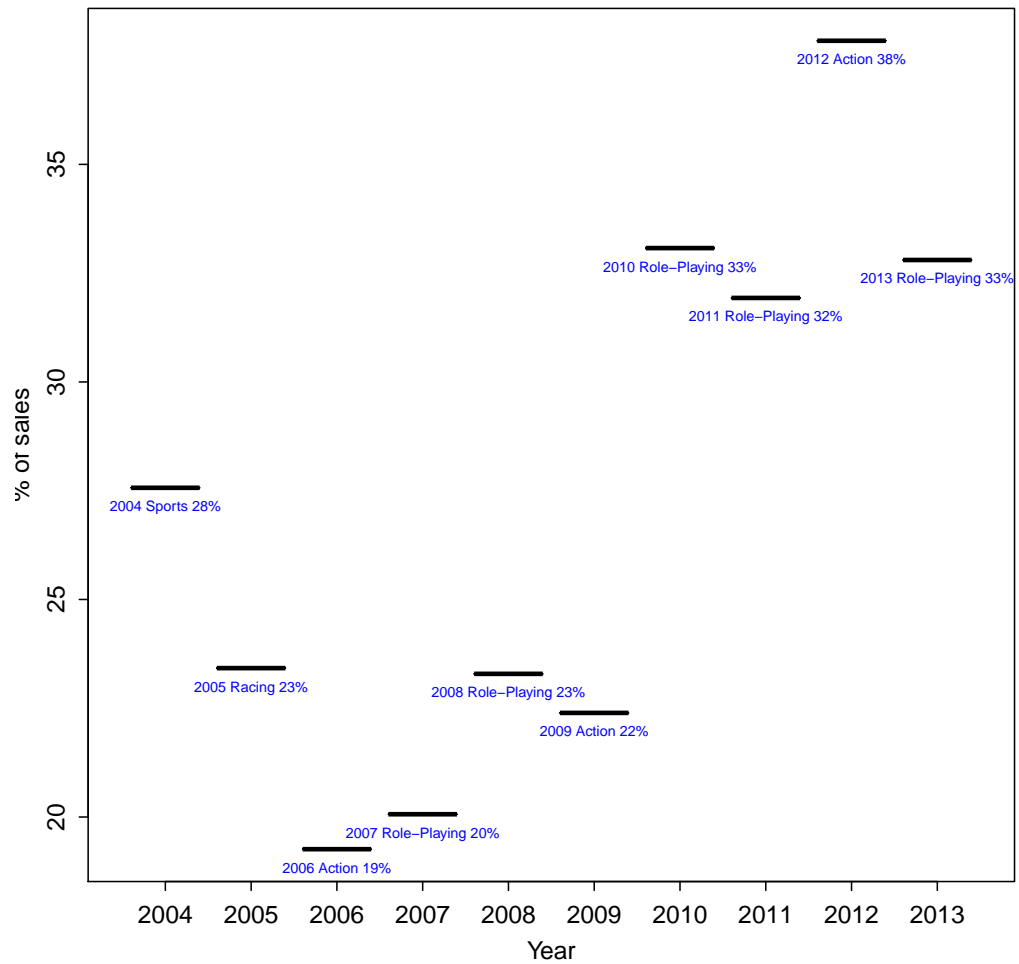


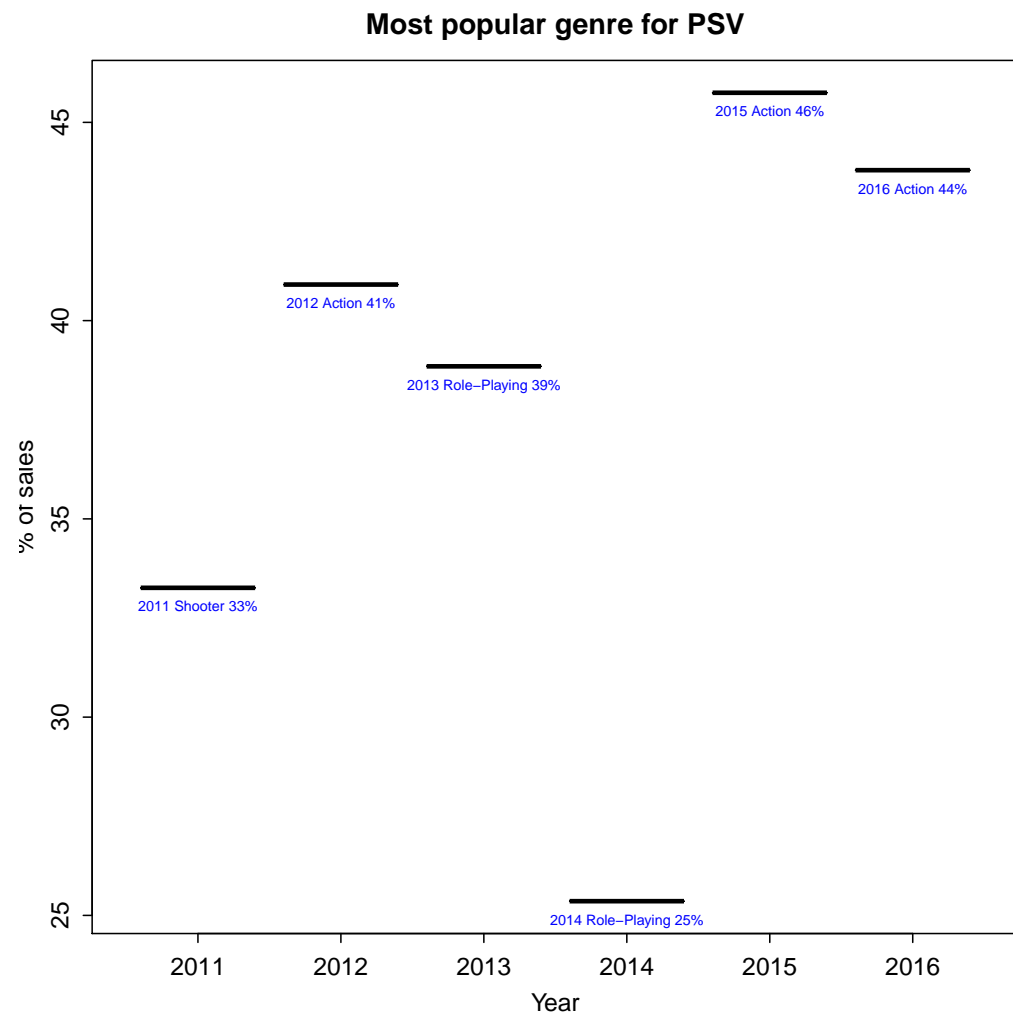
Most popular genre for PS3



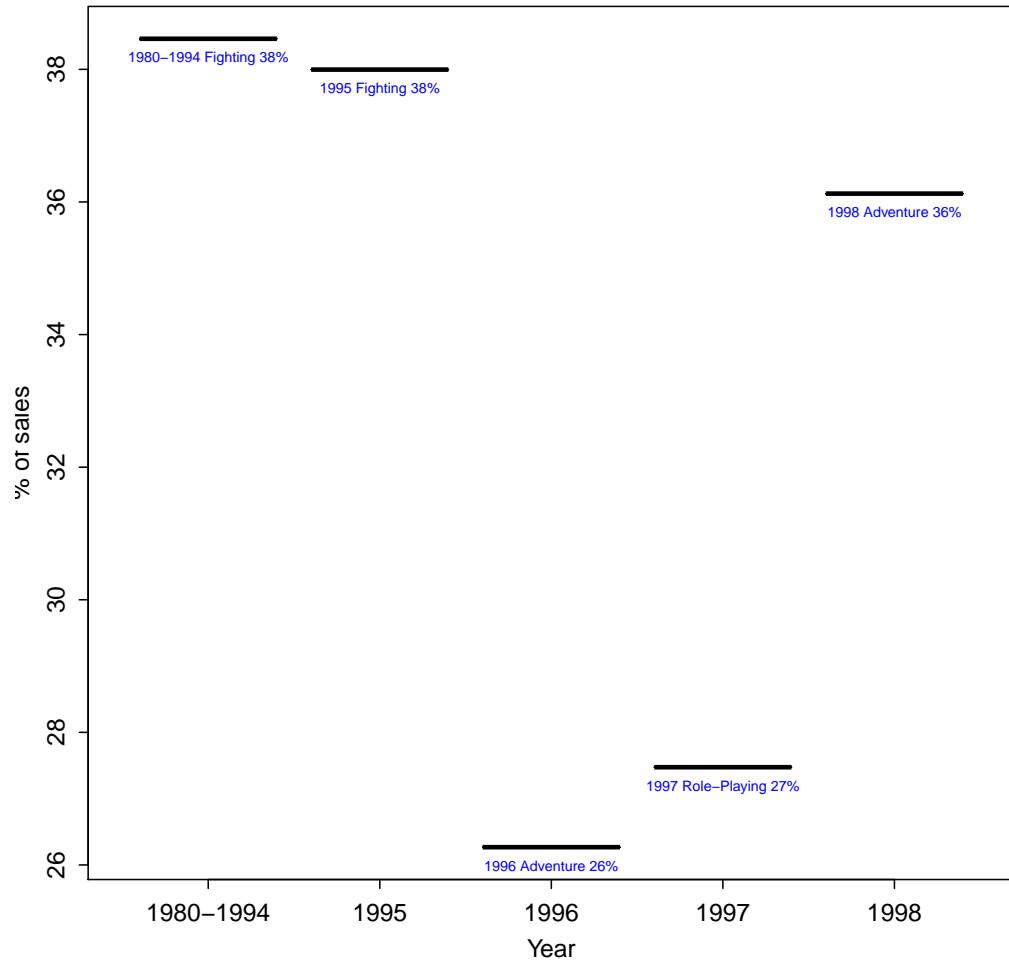


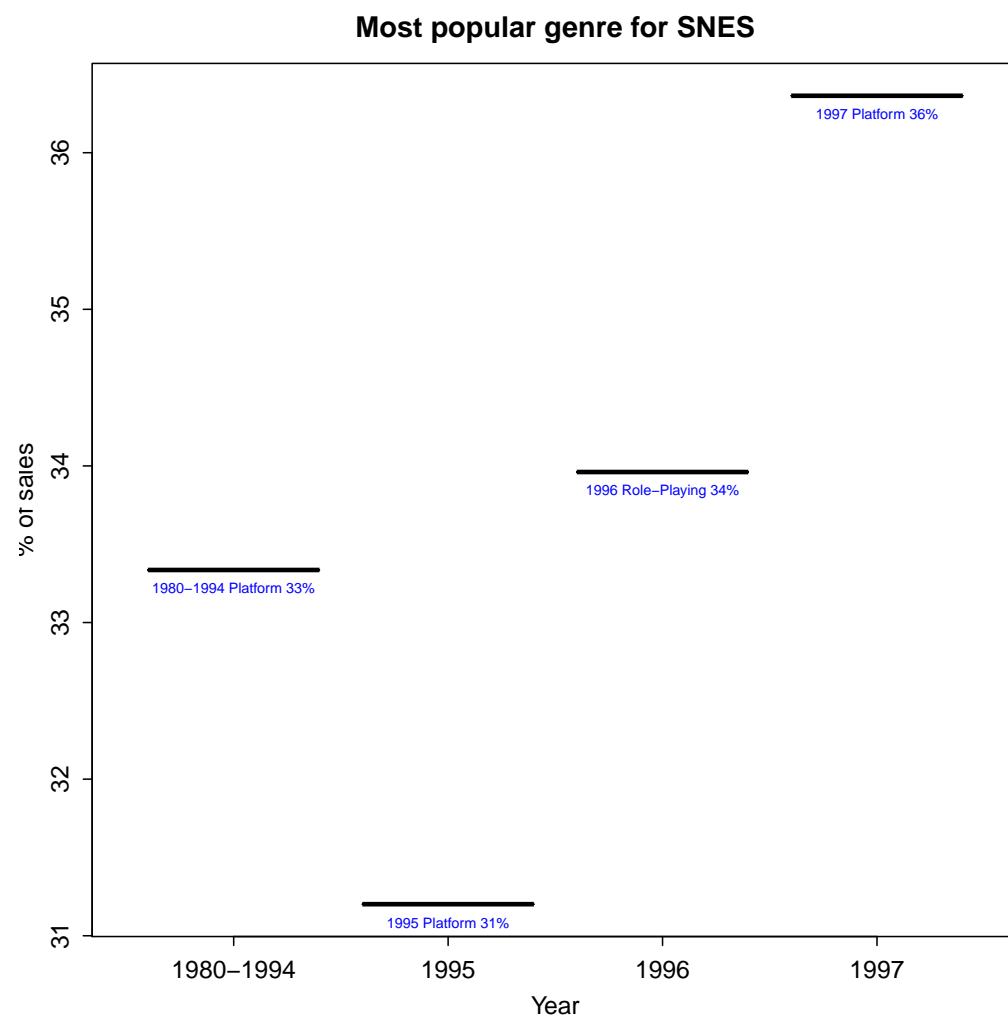
Most popular genre for PSP



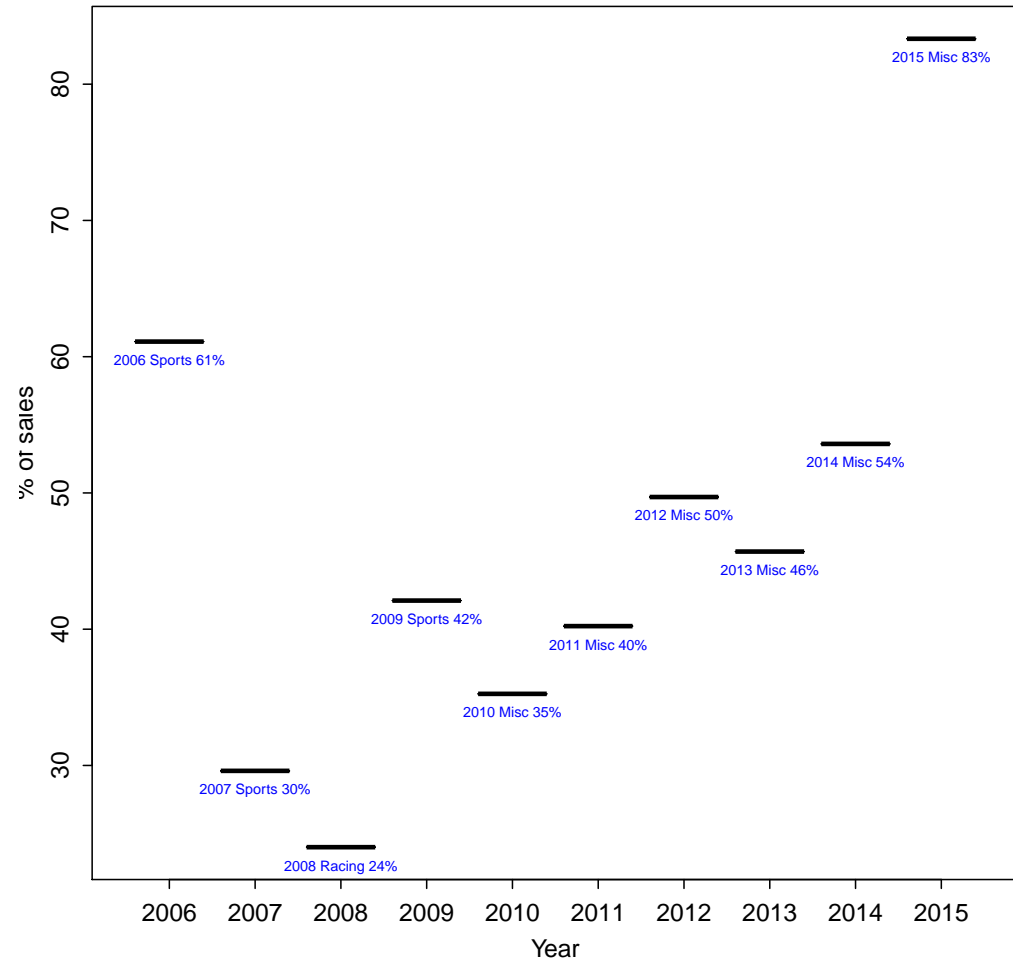


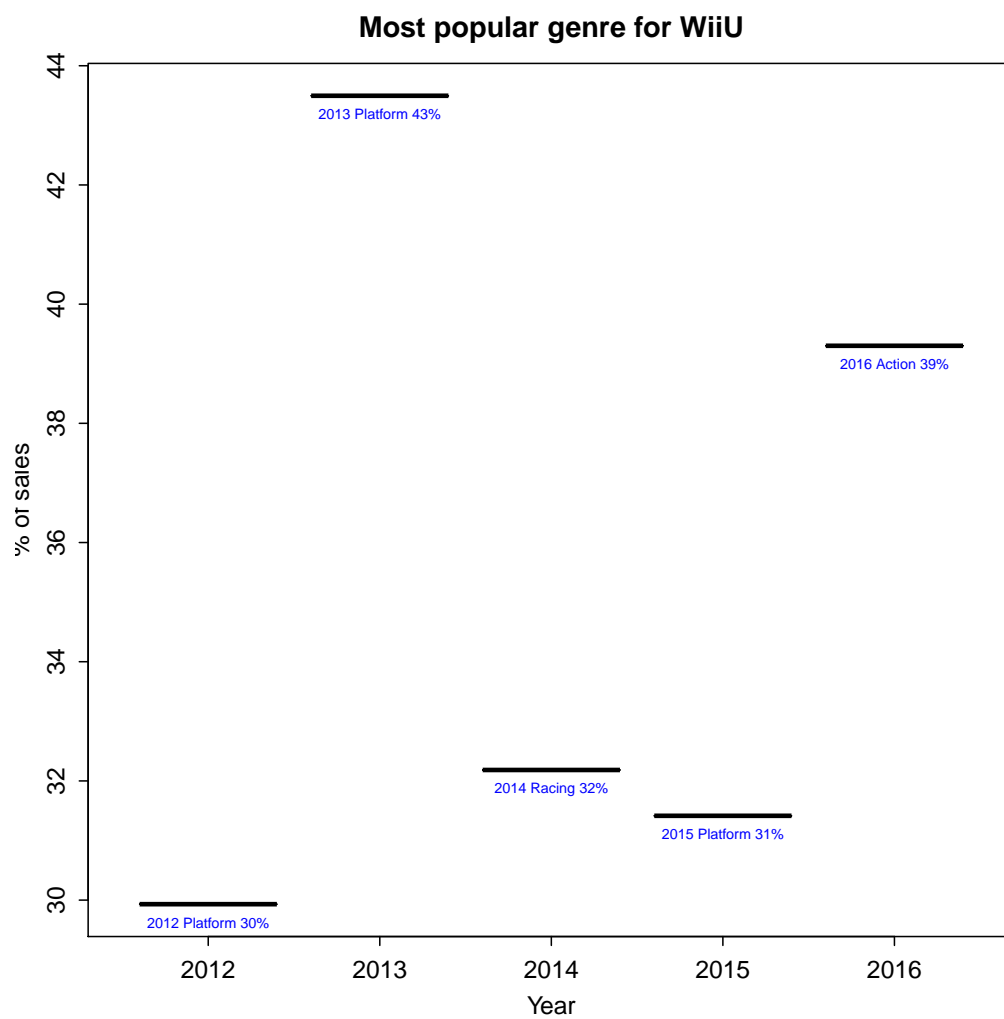
Most popular genre for SAT



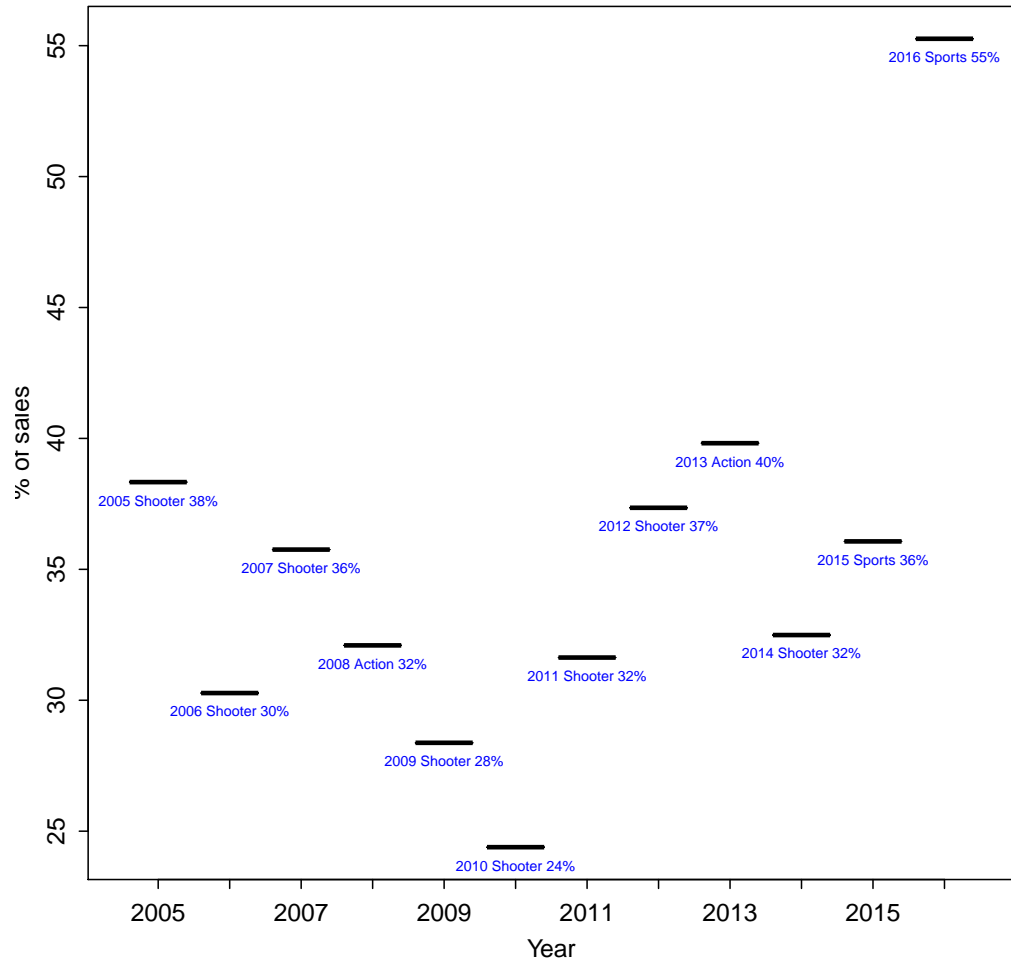


Most popular genre for Wii

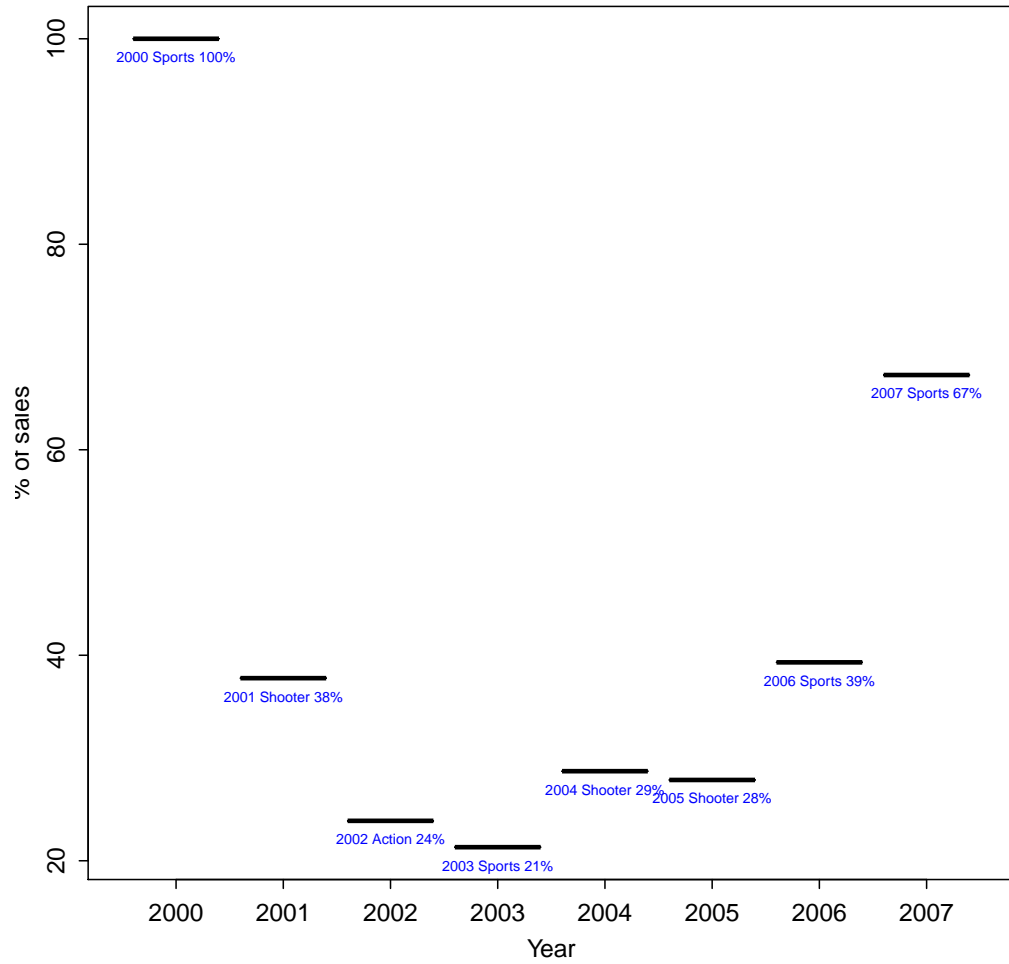


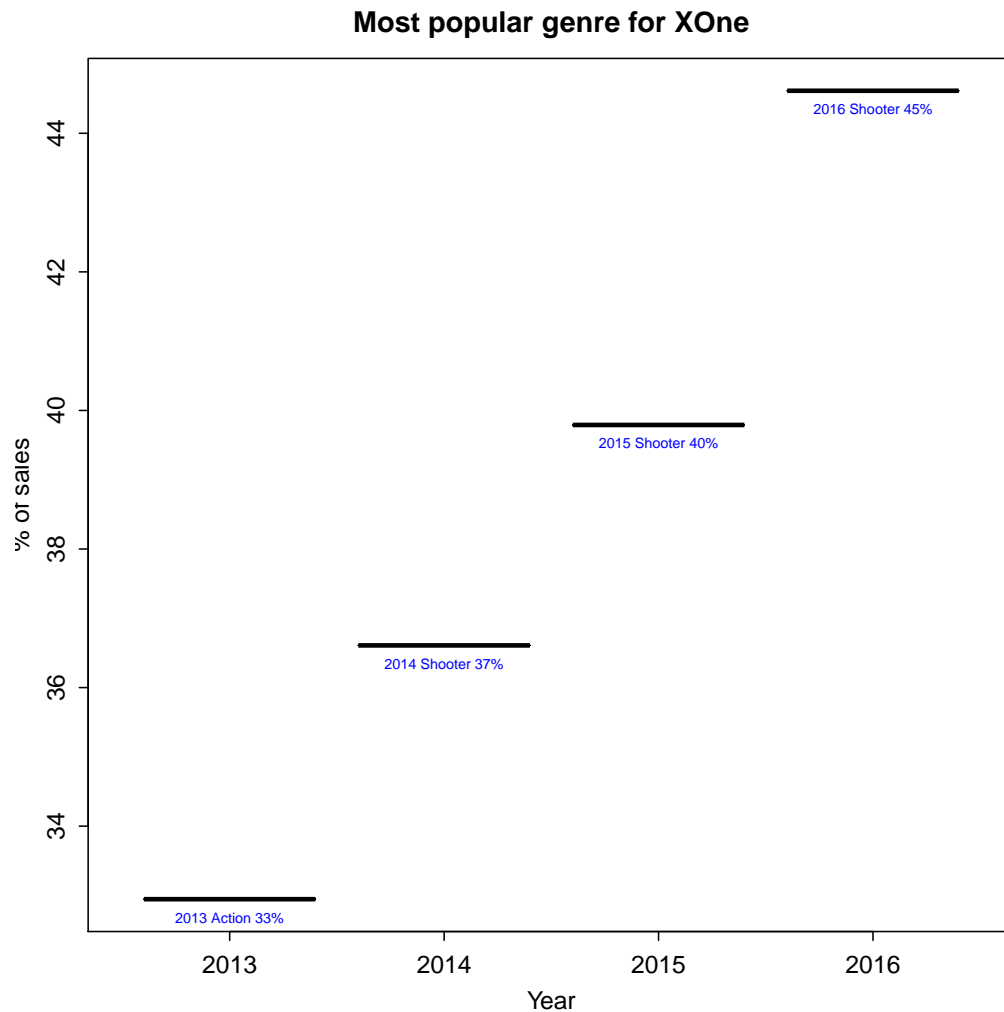


Most popular genre for X360



Most popular genre for XB





5 Resolució

Les vendes globals estan linealment relacionades amb les d'*Action*, *Adventure* i *Sports* i no amb les de estratègia. S'ha vist que la mitjana de vendes per cada plataforma i cada gènere és diferent de la global.

Les vendes han estat dominades pel gènere acció des del 2001 fins el 2015, excepte el 2006 (que és l'any del llançament del *wii sports* que va ser un gran èxit de vendes). Els valors van des del 16% fins al 32%

De totes les platatformes que s'han considerat, hi ha prou dades de 21. En part, això es degut a ajuntar els anys 1980-1994, altrament l'atari 2600 hauria estat considerada. L'anàlisi d'aquells anys podria ser un altre estudi.

D'entre les considerades, els gèneres dominats són

```

##
## Plataforma: 3DS. Temps de vida: 2011 fins 2016.
## Gènere dominant: Role-Playing
##
## Plataforma: DC. Temps de vida: 1998 fins 2001.
## Gènere dominant: Sports
##
## Plataforma: DS. Temps de vida: 2004 fins 2013.
## Gènere dominant: Platform
##
## Plataforma: GB. Temps de vida: 1995 fins 2001.
## Gènere dominant: Role-Playing
##
## Plataforma: GBA. Temps de vida: 2001 fins 2007.
## Gènere dominant: Role-Playing
##
## Plataforma: GC. Temps de vida: 2001 fins 2006.
## Gènere dominant: Action
##
## Plataforma: N64. Temps de vida: 1996 fins 2001.
## Gènere dominant: Sports
##
## Plataforma: PC. Temps de vida: 1995 fins 2016.
## Gènere dominant: Simulation
##
## Plataforma: PS. Temps de vida: 1995 fins 2003.
## Gènere dominant: Action
##
## Plataforma: PS2. Temps de vida: 2000 fins 2010.
## Gènere dominant: Sports
##
## Plataforma: PS3. Temps de vida: 2006 fins 2016.
## Gènere dominant: Action
##
## Plataforma: PS4. Temps de vida: 2013 fins 2016.
## Gènere dominant: Shooter
##
## Plataforma: PSP. Temps de vida: 2004 fins 2013.
## Gènere dominant: Role-Playing
##

```

Plataforma: PSV. Temps de vida: 2011 fins 2016.
Gènere dominant: Action

Plataforma: SAT. Temps de vida: 1995 fins 1998.
Gènere dominant: Adventure

Plataforma: SNES. Temps de vida: 1995 fins 1997.
Gènere dominant: Platform

Plataforma: Wii. Temps de vida: 2006 fins 2015.
Gènere dominant: Misc

Plataforma: WiiU. Temps de vida: 2012 fins 2016.
Gènere dominant: Platform

Plataforma: X360. Temps de vida: 2005 fins 2016.
Gènere dominant: Shooter

Plataforma: XB. Temps de vida: 2000 fins 2007.
Gènere dominant: Sports

Plataforma: XOne. Temps de vida: 2013 fins 2016.
Gènere dominant: Shooter