

Relations entre streaming et ventes de jeux vidéo

Basé sur les diffusions en direct
de Twitch.tv en février 2015

Présenté par :
Jawhara CHAFI
Mathys GENET
Rémi GOMES MOREIRA
Imane SAHNOUNE

Objectifs :

Réaliser une analyse exploratoire

Les bases de données (publiques)

- Ventes de jeux en fin d'année 2015, récolté sur vgchartz.com
- Diffusion en direct sur Twitch pendant février 2015, récolté toutes les 5 minutes sur une période d'un mois via l'api Twitch (disponible [ici](#))

Un peu de vocabulaire

Stream :

Diffusion audiovisuel en ligne et en direct
(ici sur la plateforme Twitch.tv)

Twitch :

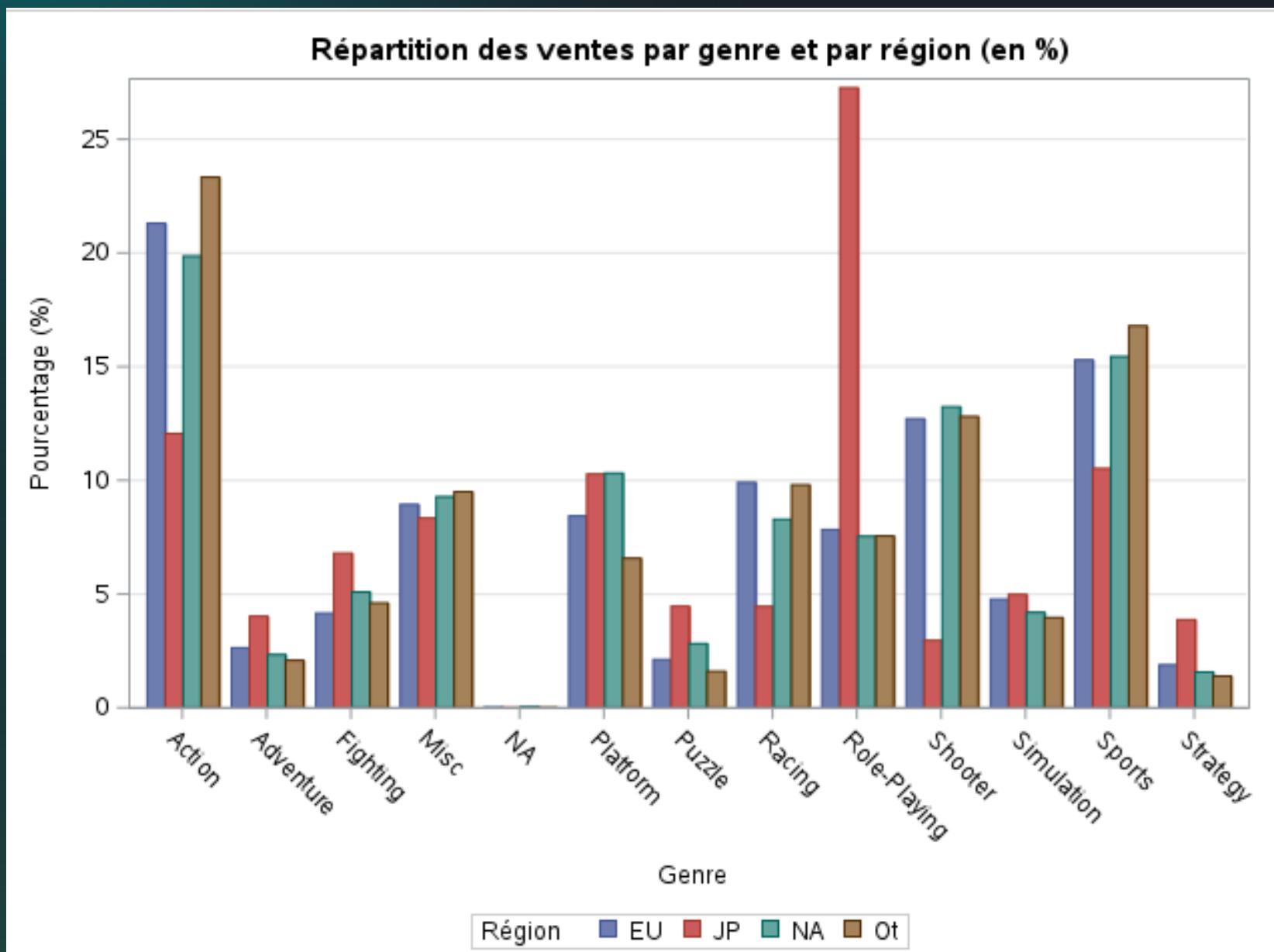
Plateforme en ligne de diffusion en direct de contenu audiovisuel

Limites :

Grande perte d'informations due à la suppression de données :

- jeux qui n'ont pas été diffusés
- diffusion dont le jeu n'est pas dans la base des ventes

Présentation des ventes de jeux

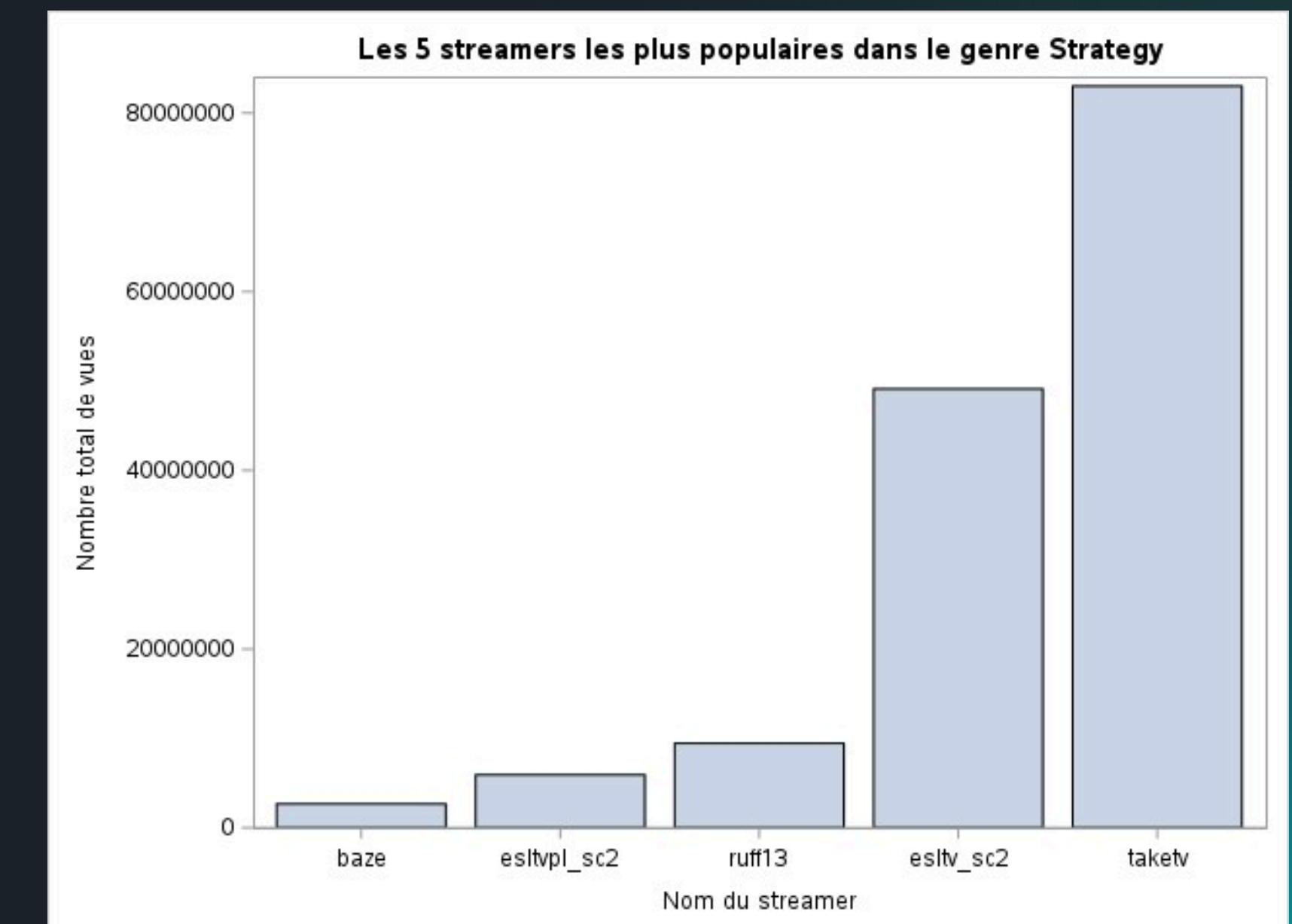
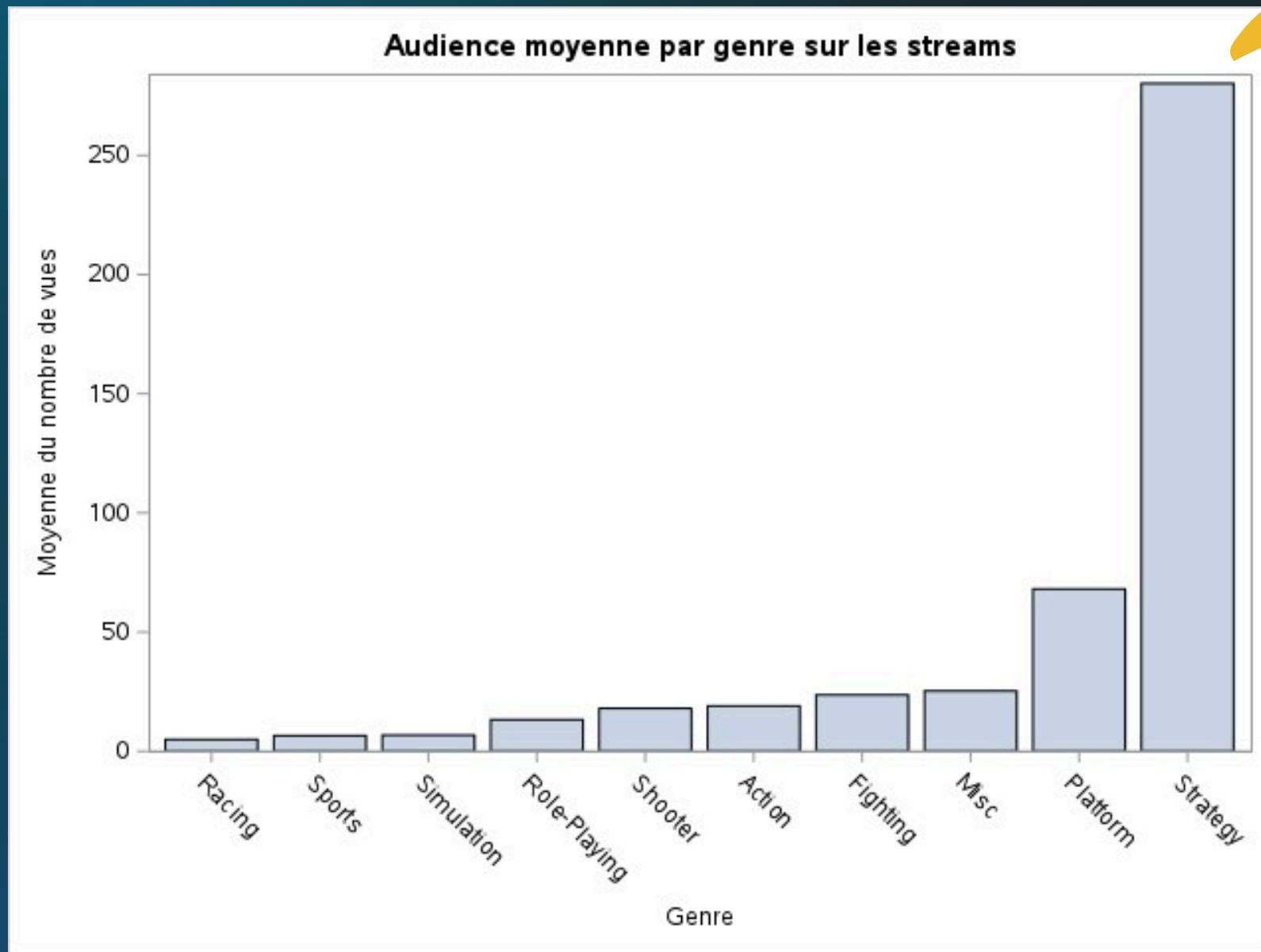


En 2015, le jeu vidéo a engendré

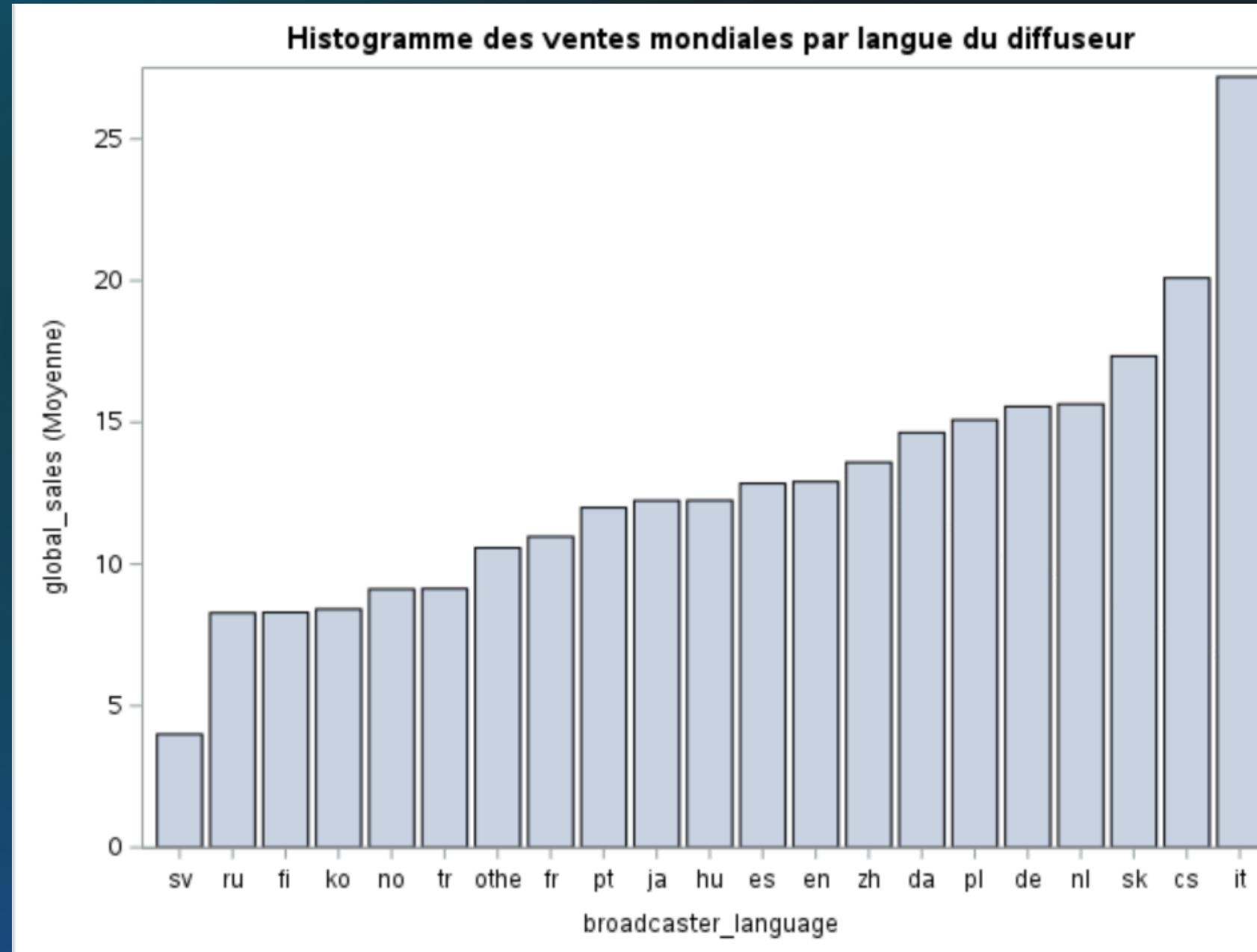
8.6 Md \$

Ventes globales par genre de jeu		
Obs.	Genre	Total_Global_Sales
1	Action	1687.55
2	Sports	1286.79
3	Shooter	1003.64
4	Role-Playing	912.42
5	Platform	822.65
6	Misc	788.12
7	Racing	720.76
8	Fighting	438.19
9	Simulation	386.22
10	Puzzle	240.17
11	Adventure	229.58
12	Strategy	172.14
13	NA	2.42

Quelles sont les **facteurs** impactant le nombre de vues d'un stream ?



L'impact des différences nationales sur les ventes de jeux populaires

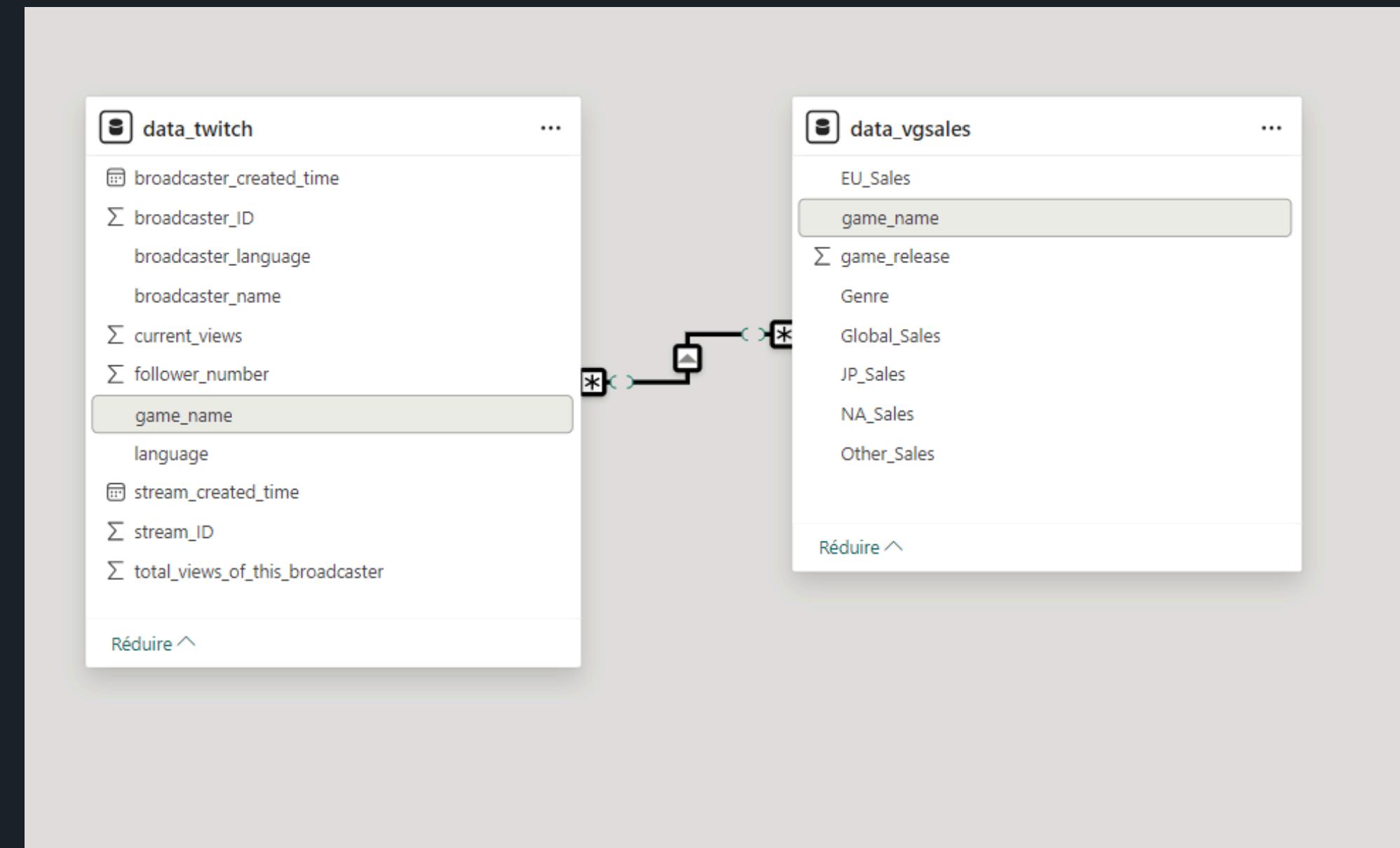


Niveau de broadcaster_language	global_sales		
	N	Moyenne	Ec-type
cs	3	20.0933333	31.6367829
da	10	14.6380000	16.3863112
de	77	15.5587013	14.5647687
en	4906	12.9051998	14.3931396
es	48	12.8437500	12.7102265
fi	4	8.2925000	9.5592447
fr	72	10.9665278	13.4998363
hu	16	12.2450000	17.8021710
it	3	27.1900000	27.1560012
ja	53	12.2388679	12.9071450
ko	14	8.4092857	9.6379366
nl	7	15.6457143	9.0569124
no	3	9.1166667	12.5473118
othe	17	10.5700000	15.2426511
pl	32	15.0856250	15.4259783
pt	16	11.9887500	8.9433222
ru	95	8.2773684	11.8660240
sk	1	17.3400000	-
sv	17	3.9976471	4.1590497
tr	4	9.1300000	7.6789366
zh	210	13.5884762	15.4369726

Merci pour votre attention

Annexe 1

Modélisation des données :



Lien git: https://github.com/Jawharachf/Projet_SAS.git

Annexe 2

Les jeux et le genre les plus populaires dans notre base de données

game_name	nb_jeux
Dying Light	1187
Destiny	545
Minecraft	440
Call of Duty: Advanced Warfare	406
Grand Theft Auto V	397
Battlefield 4	180
Madden NFL 15	180
FIFA 15	173
NBA 2K15	166
Far Cry 4	93
ArmA III	81
World of Tanks	74
StarCraft II: Heart of the Swarm	73
The Last of Us	71
Halo: The Master Chief Collection	69
Call of Duty: Black Ops II	59
Resident Evil	53
Dragon Age: Inquisition	43
Guild Wars 2	34
The Elder Scrolls V: Skyrim	34
World of Warcraft	33

Genre	nb_jeux
Action	2007
Shooter	1746
Sports	593
Misc	455
Role-Playing	425
Strategy	118
Racing	87
Fighting	63
Platform	61
Simulation	32
Adventure	14
Puzzle	7

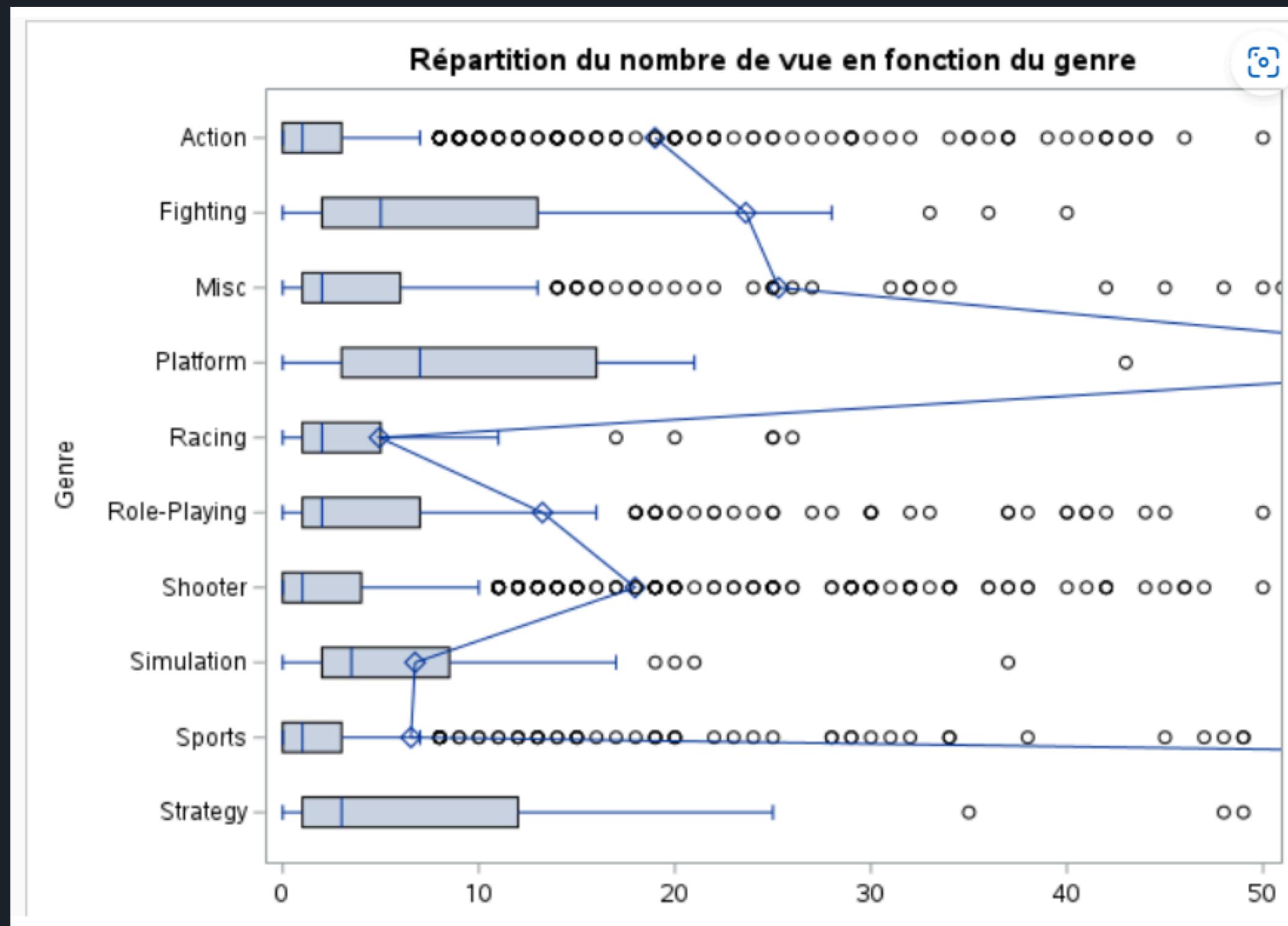
Annexe 3

Les statistiques descriptives de la variable exprimant le nombre de vues par genre

Statistique descriptives du nombre de vues par genre						
La procédure MEANS						
Variable d'analyse : current_views						
Genre	N obs	N	Moyenne	Ec-type	Minimum	Maximum
Action	2007	2007	18.9910314	364.3459332	0	15252.00
Adventure	14	14	2.9285714	3.8122129	0	12.0000000
Fighting	63	63	23.6190476	64.9164298	0	371.0000000
Misc	455	455	25.2945055	205.3998600	0	3893.00
Platform	61	61	68.0655738	299.5662903	0	2301.00
Puzzle	7	7	1.8571429	3.6709931	0	10.0000000
Racing	87	87	4.9310345	9.1978171	0	54.0000000
Role-Playing	425	425	13.2470588	71.0493294	0	1295.00
Shooter	1746	1746	17.9696449	293.7047944	0	12037.00
Simulation	32	32	6.7500000	8.1636488	0	37.0000000
Sports	593	593	6.5396290	32.4353129	0	629.0000000
Strategy	118	118	280.2033898	2515.51	0	27293.00

Action, Sports et Strategy se distinguent le plus avec une plus grande variabilité sur les jeux actions

Annexe 4



Beaucoup d'outliers ainsi qu'une grande différence de moyenne

Annexe 5

Test Anova sur le nombre de vues et le genre

Test des moyennes du nombre de vues en fonction du genre

La procédure GLM

Test de Levene pour l'homogénéité de la variance current_views ANOVA des écarts au carré des moyennes de groupe					
Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Genre	11	4.444E15	4.04E14	3.75	<.0001
Erreur	5596	6.034E17	1.078E14		

Il existe une non homogénéité des variances donc on ne peut pas procéder à l'analyse des moyennes

Annexe 5

Méthode de Tunkey pour l'analyse des moyennes

Moyennes des moindres carrés pour l'effet Genre Pr > t pour H0: LSMean(i)=LSMean(j)												
Variable dépendante : current_views												
i/j	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	<.0001
2	1.0000		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5997
3	1.0000	1.0000		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0184
4	1.0000	1.0000	1.0000		0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	<.0001
5	0.9996	1.0000	1.0000	0.9999		1.0000	0.9996	0.9994	0.9996	1.0000	0.9978	0.1327
6	1.0000	1.0000	1.0000	1.0000	1.0000		1.0000	1.0000	1.0000	1.0000	1.0000	0.9248
7	1.0000	1.0000	1.0000	1.0000	0.9996	1.0000		1.0000	1.0000	1.0000	1.0000	0.0014
8	1.0000	1.0000	1.0000	1.0000	0.9994	1.0000	1.0000		1.0000	1.0000	1.0000	<.0001
9	1.0000	1.0000	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000		1.0000	1.0000	<.0001
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		1.0000	0.1141
11	1.0000	1.0000	1.0000	1.0000	0.9978	1.0000	1.0000	1.0000	1.0000	1.0000		<.0001
12	<.0001	0.5997	0.0184	<.0001	0.1327	0.9248	0.0014	<.0001	<.0001	0.1141	<.0001	

La procédure GLM
Moyennes des moindres carrés
Ajustement pour comparaisons multiples : Tukey-Kramer

Genre	current_views LSMEAN	Nombre LSMEAN
Action	18.991031	1
Adventure	2.928571	2
Fighting	23.619048	3
Misc	25.294505	4
Platform	68.065574	5
Puzzle	1.857143	6
Racing	4.931034	7
Role-Playing	13.247059	8
Shooter	17.969645	9
Simulation	6.750000	10
Sports	6.539629	11
Strategy	280.203390	12

Le genre Stratégie se distingue beaucoup plus, cela peut être dû aux outliers.

Annexe 5

Analyse de la corrélation avec current_views

La procédure CORR							
1 Avec les variables :	current_views						
7 Variables :	current_views total_views_of_this_broadcaster Global_Sales JP_Sales NA_Sales EU_Sales follower_number						
Statistiques simples							
Variable	N	Moyenne	Ec-type	Somme	Minimum	Maximum	
current_views	5608	23.16512	461.26930	129910	0	27293	
total_views_of_this_broadcaster	5608	87097	1690109	488439740	1.00000	83006063	
Global_Sales	5608	12.83053	14.35904	71954	0.01000	56.57000	
JP_Sales	5608	0.38962	0.51952	2185	0	5.12000	
NA_Sales	5608	5.92460	6.17251	33225	0	23.84000	
EU_Sales	5608	4.85522	6.10385	27228	0	23.42000	
follower_number	5608	1060	13639	5943676	1.00000	603355	
Coefficients de corrélation de Pearson, N = 5608 Proba > r sous H0: Rho=0							
current_views	current_views 1.00000	follower_number 0.57767 <.0001	total_views_of_this_broadcaster 0.55257 <.0001	Global_Sales -0.01599 0.2313	EU_Sales -0.01574 0.2386	NA_Sales -0.01572 0.2393	JP_Sales -0.00556 0.6772

Le nombre de vues sur le stream est corrélé avec le nombre d'abonnés et le nombre total du stream

Annexe 5

Test 2-way anova de nombre de vues sur le stream en fonction du nombre total de vues et du genre

La procédure GLM					
Variable dépendante : current_views Nombre de vues					
Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Modèle	10	364962771	36496277	245.77	<.0001
Erreur	5576	828025875	148498		
Total sommes corrigées	5586	1192988646			

R-carré	Coef de var	Racine MSE	current_views Moyenne
0.305923	1657.972	385.3546	23.24253

Source	DDL	Type I SS	Carré moyen	Valeur F	Pr > F
total_views_of_this_	1	364248246.3	364248246.3	2452.88	<.0001
Genre	9	714524.8	79391.6	0.53	0.8503

Source	DDL	Type III SS	Carré moyen	Valeur F	Pr > F
total_views_of_this_	1	356716284.1	356716284.1	2402.16	<.0001
Genre	9	714524.8	79391.6	0.53	0.8503

Le genre n'est plus significatif et donc aucune différence de moyenne n'existe