

So You Want to Produce the next “Game of the Year”

**What do the previous decades of gaming
tell us about the future game market.**

The Gamers:

Raj Agrawal

John Banowsky

Van Eisele



Data: Sales Of Video Games

Analyze sales data from more than 11,000 games.

<https://www.kaggle.com/datasets/arslanali4343/sales-of-video-games>

What are some factors that go into game sales?

1. Are there any regions that help dictate successful game sales?
2. Does genre have any impact on sales?
3. Do certain publishers do better in video game sales?
4. Are there any differences in gaming platforms sales?

But first

The boring stuff

Why: John and Van are both avid gamers and Raj wanted to learn more about the future of games after DigDug.

Data cleaning:

1. We looked at the data with .describe and .info
2. We found there were 271 rows missing years so we dropped null values.
3. We Binned the different platforms into the same platform company.
4. We took game release years and binned them into decades.
5. We binned similar game genres.

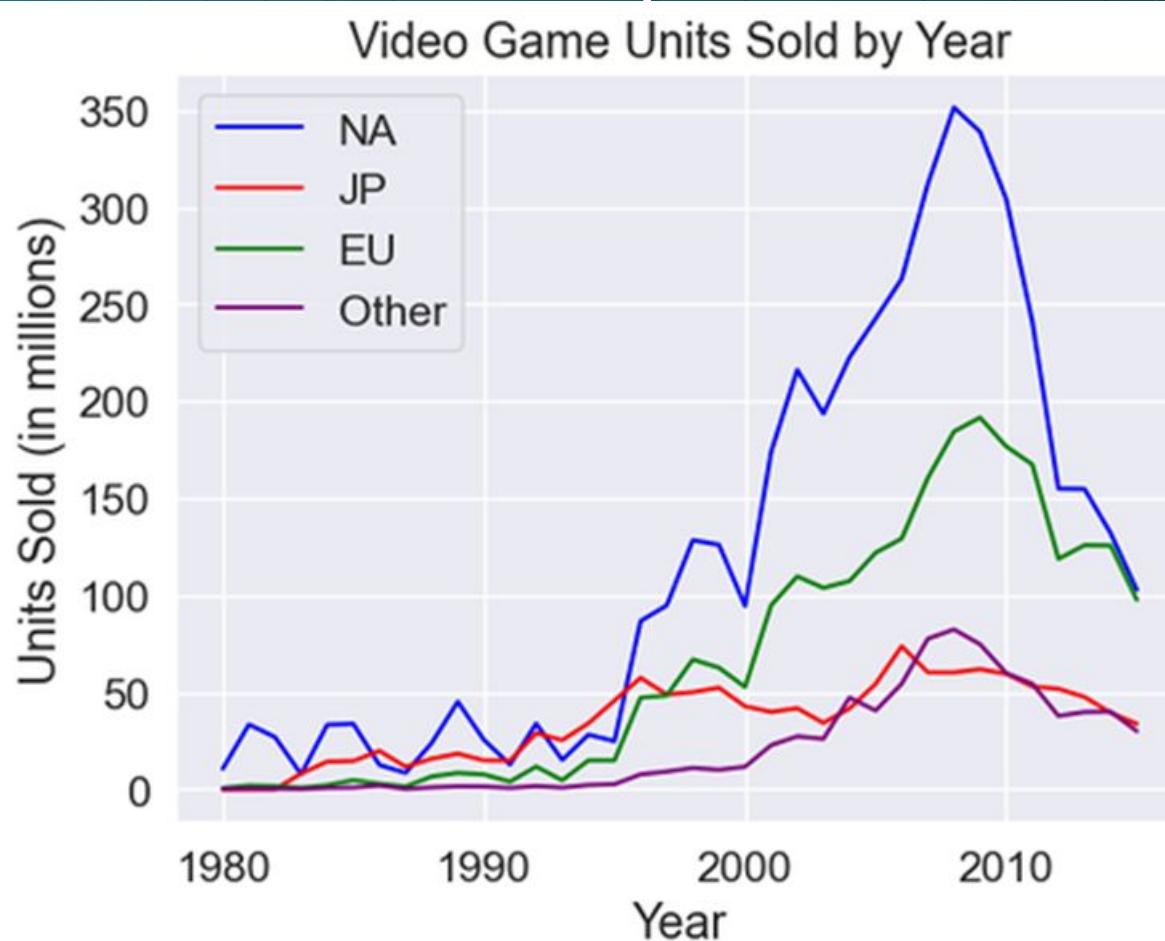
Code Excerpt

```
1 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["N64", "GC", "Wii", "WiiU", "NES", "GB", "DS", "SNES"])
2 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["PS", "PS2", "PS3", "PSV", "PSP", "PS4"], "Sony")
3 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["XOne", "XB", "X360"], "Microsoft")
4 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["GG", "SAT", "Sega", "GEN", "SCD", "DC"], "Sega")
5 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["2600"], "Atari")
6 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["WS"], "BANDAI")
7 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["NG"], "SNK")
8 sales_data["Platform Company"] = sales_data["Platform Company"].replace(["TG16", "PCFX"], "NEC")
9 sales_data["Platform Company"].unique()

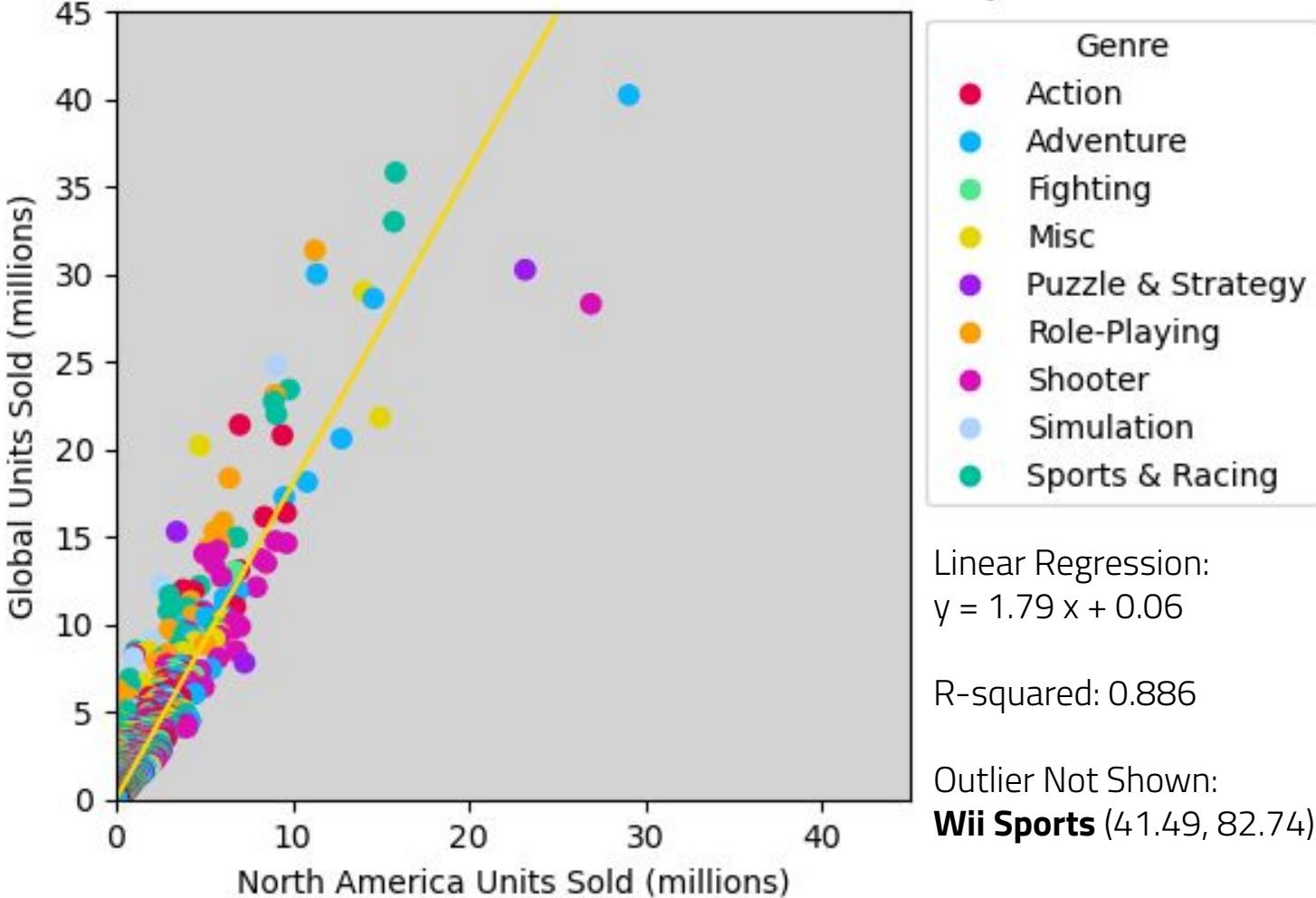
array(['Nintendo', 'Microsoft', 'Sony', 'PC', 'Atari', 'Sega', 'BANDAI',
       'SNK', 'NEC', '3DO'], dtype=object)

1 sales_data["Genre"] = sales_data["Genre"].replace(["Sports", "Racing"], "Sports & Racing")
2 sales_data["Genre"] = sales_data["Genre"].replace(["Puzzle", "Strategy"], "Puzzle & Strategy")
3 sales_data["Genre"] = sales_data["Genre"].replace(["Platform"], "Adventure")
```

What Can We Tell Globally



Video Game Units Sold in North America vs Globally



OLS Regression Results

=====
 Dep. Variable: Global_Sales R-squared: 0.886
 Model: OLS Adj. R-squared: 0.886
 Method: Least Squares F-statistic: 1.247e+05
 Date: Wed, 17 May 2023 Prob (F-statistic): 0.00
 Time: 19:05:34 Log-Likelihood: -12606.
 No. Observations: 15979 AIC: 2.522e+04
 Df Residuals: 15977 BIC: 2.523e+04
 Df Model: 1
 Covariance Type: nonrobust

=====

	coef	std err	t	P> t	[0.025	0.975]
const	0.0637	0.004	14.382	0.000	0.055	0.072
NA_Sales	1.7935	0.005	353.115	0.000	1.784	1.803

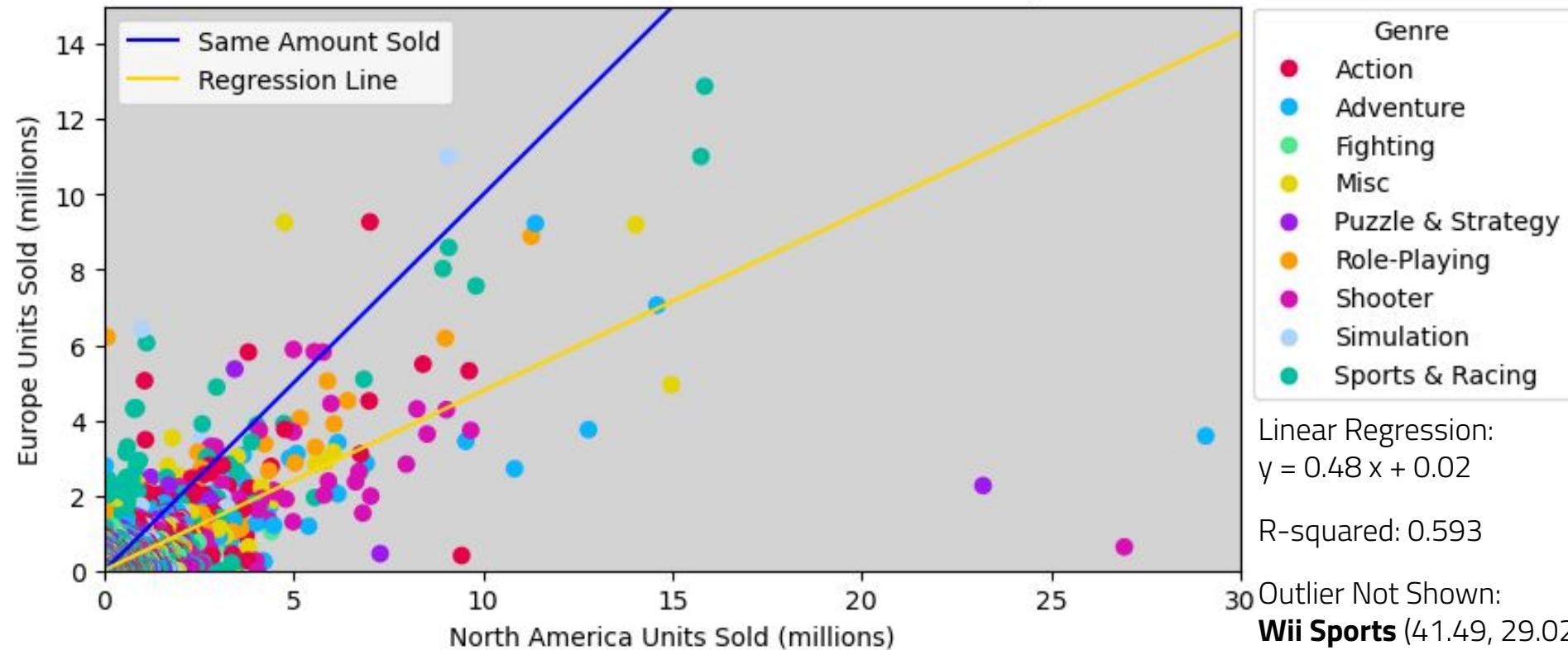
=====

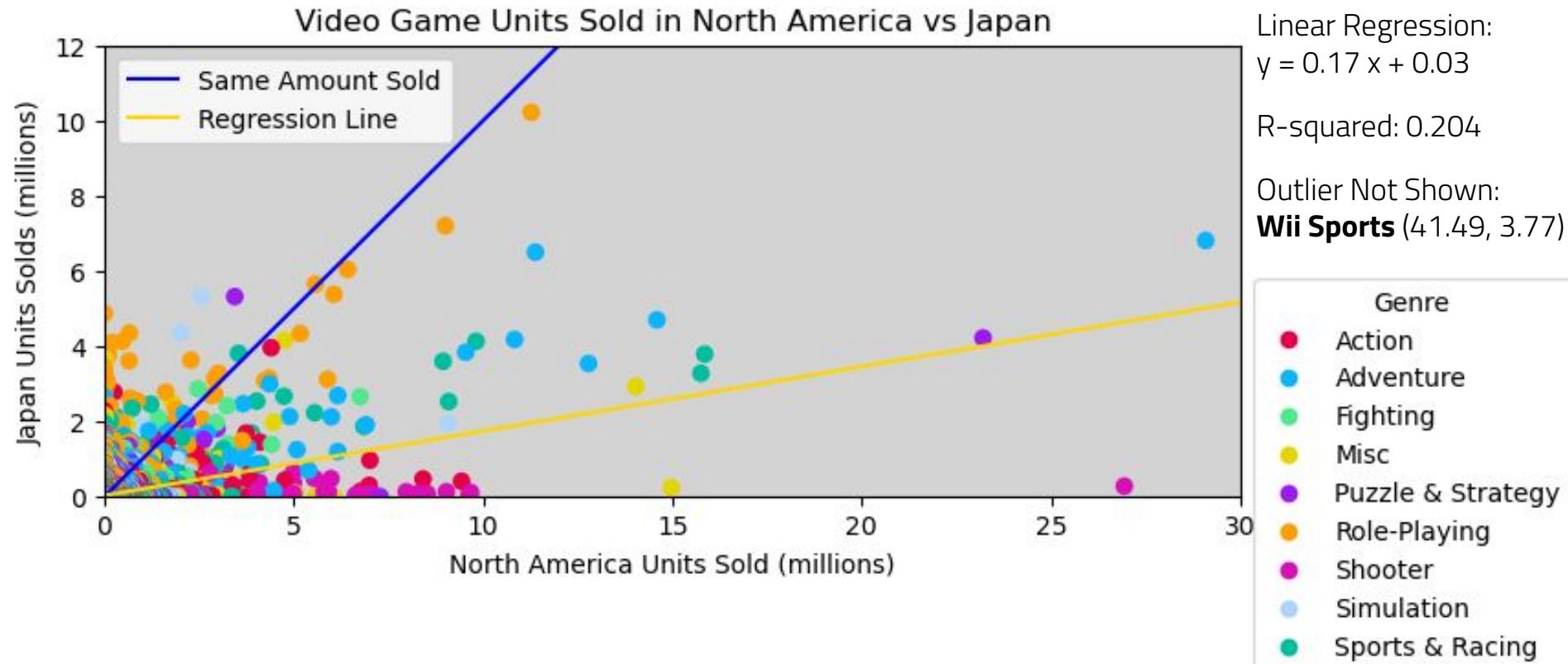
Omnibus:	9342.128	Durbin-Watson:	1.927
Prob(Omnibus):	0.000	Jarque-Bera (JB):	37165632.327
Skew:	1.074	Prob(JB):	0.00
Kurtosis:	239.256	Cond. No.	1.42

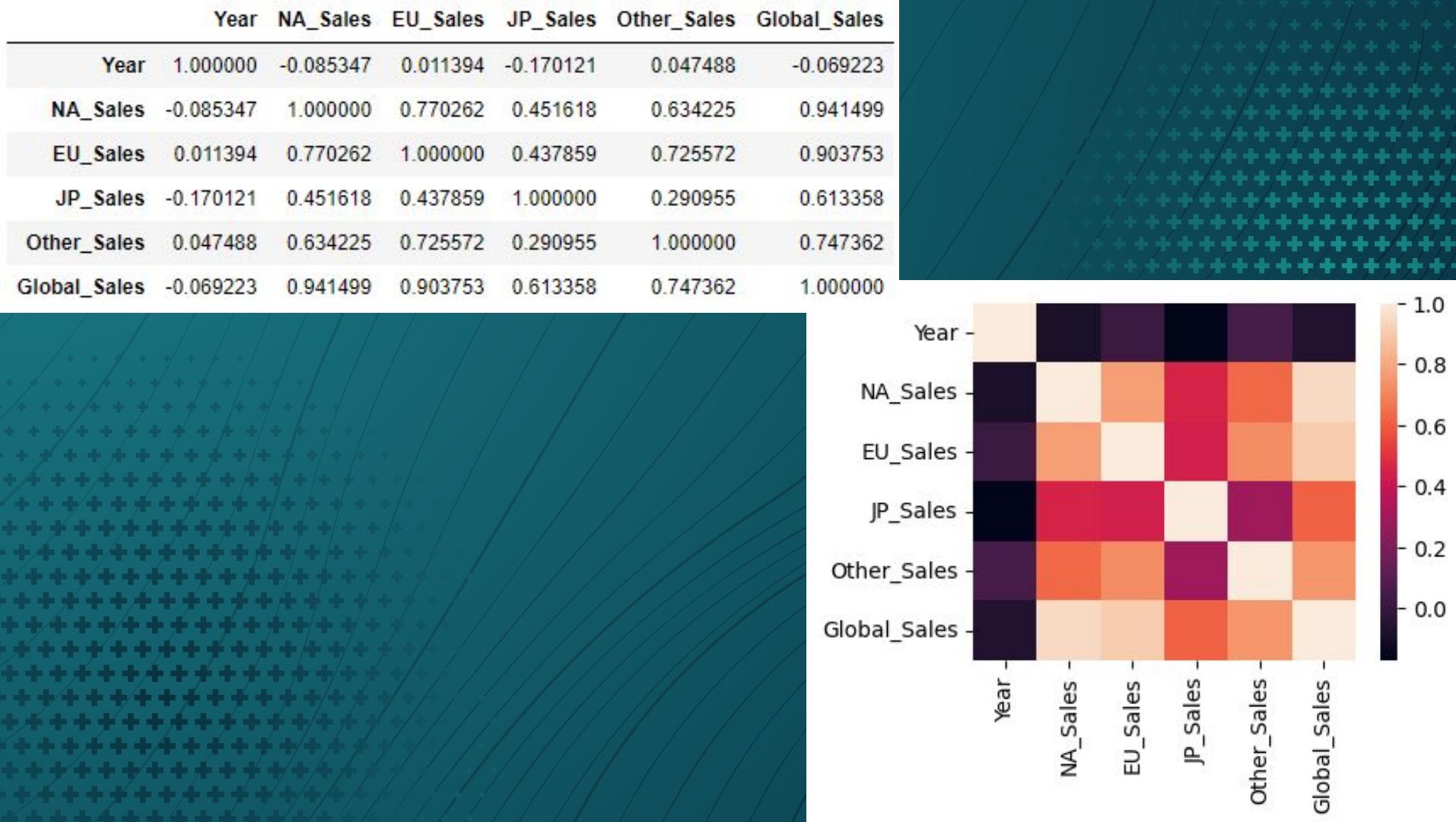
=====

Notes:
 [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

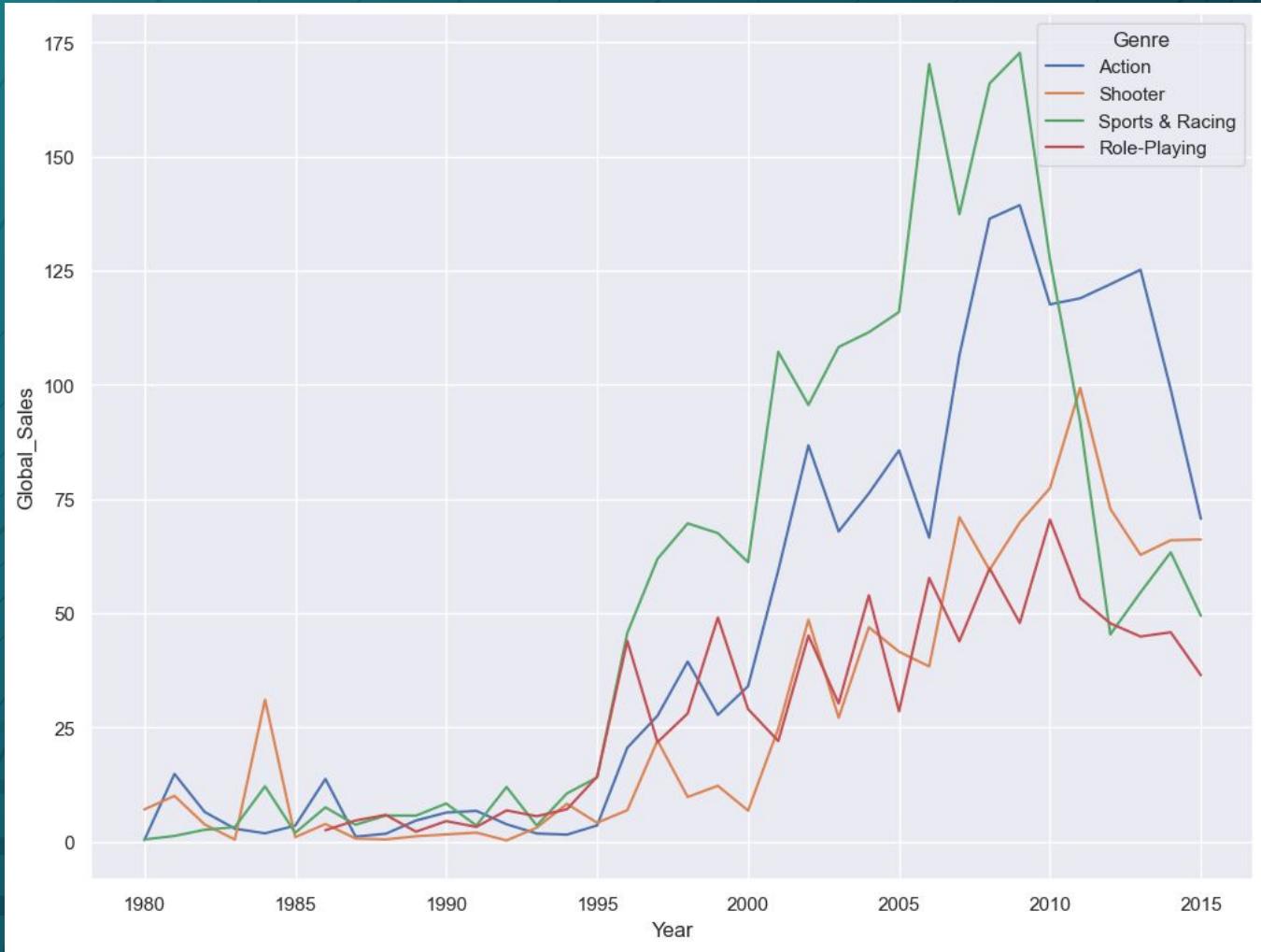
Video Game Units Sold in North America vs Europe



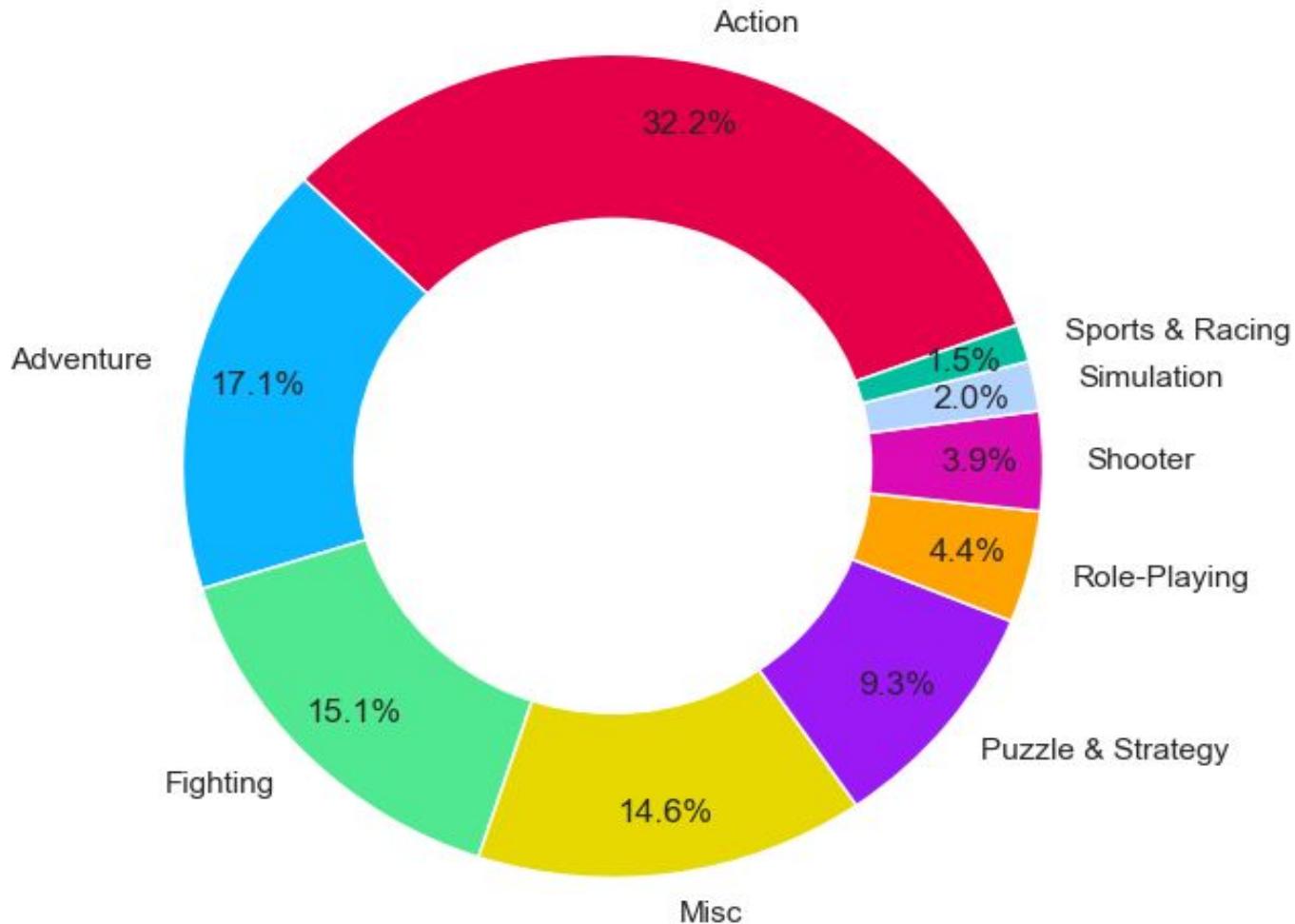




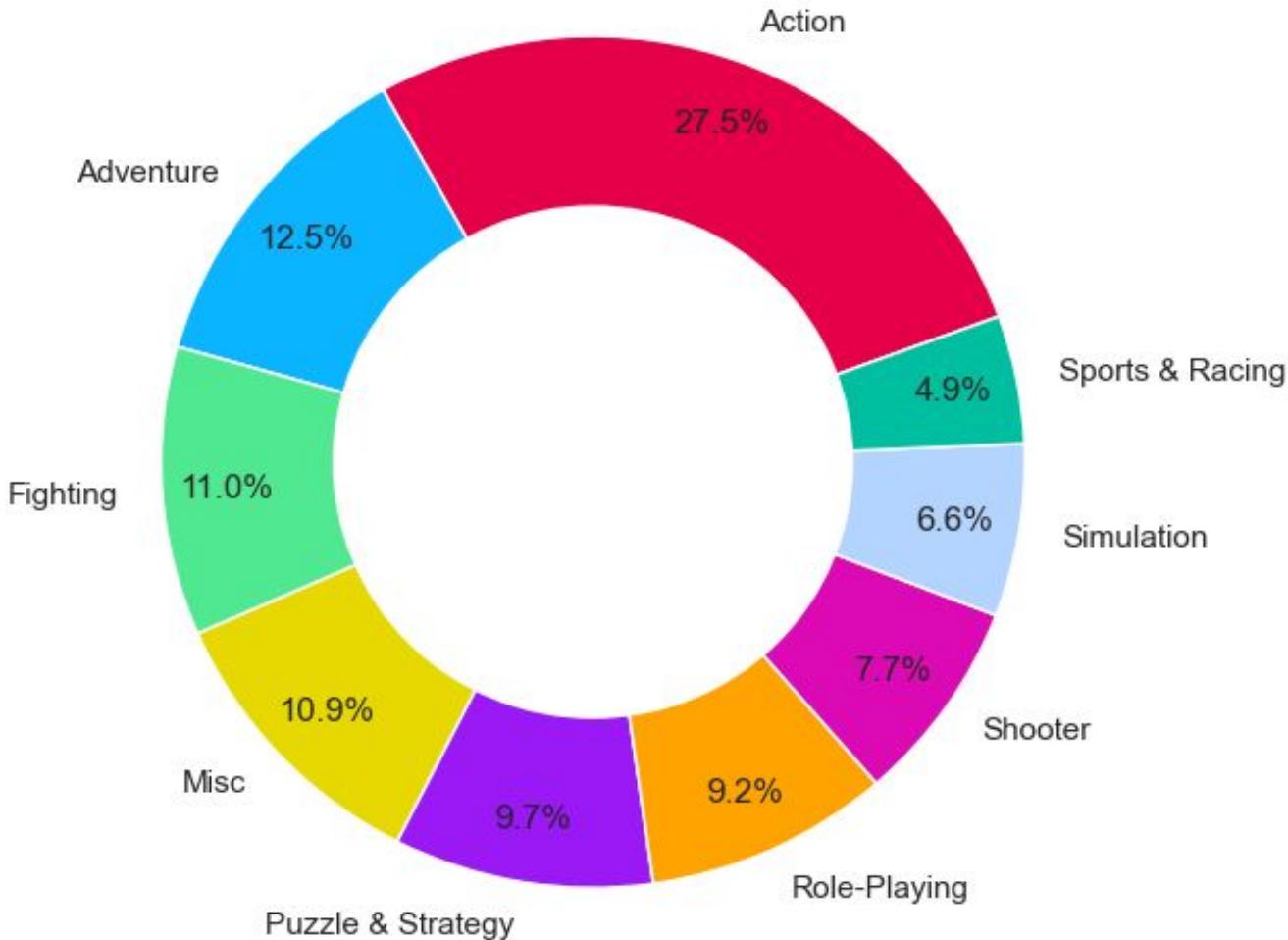
Genre Sales through the Decades



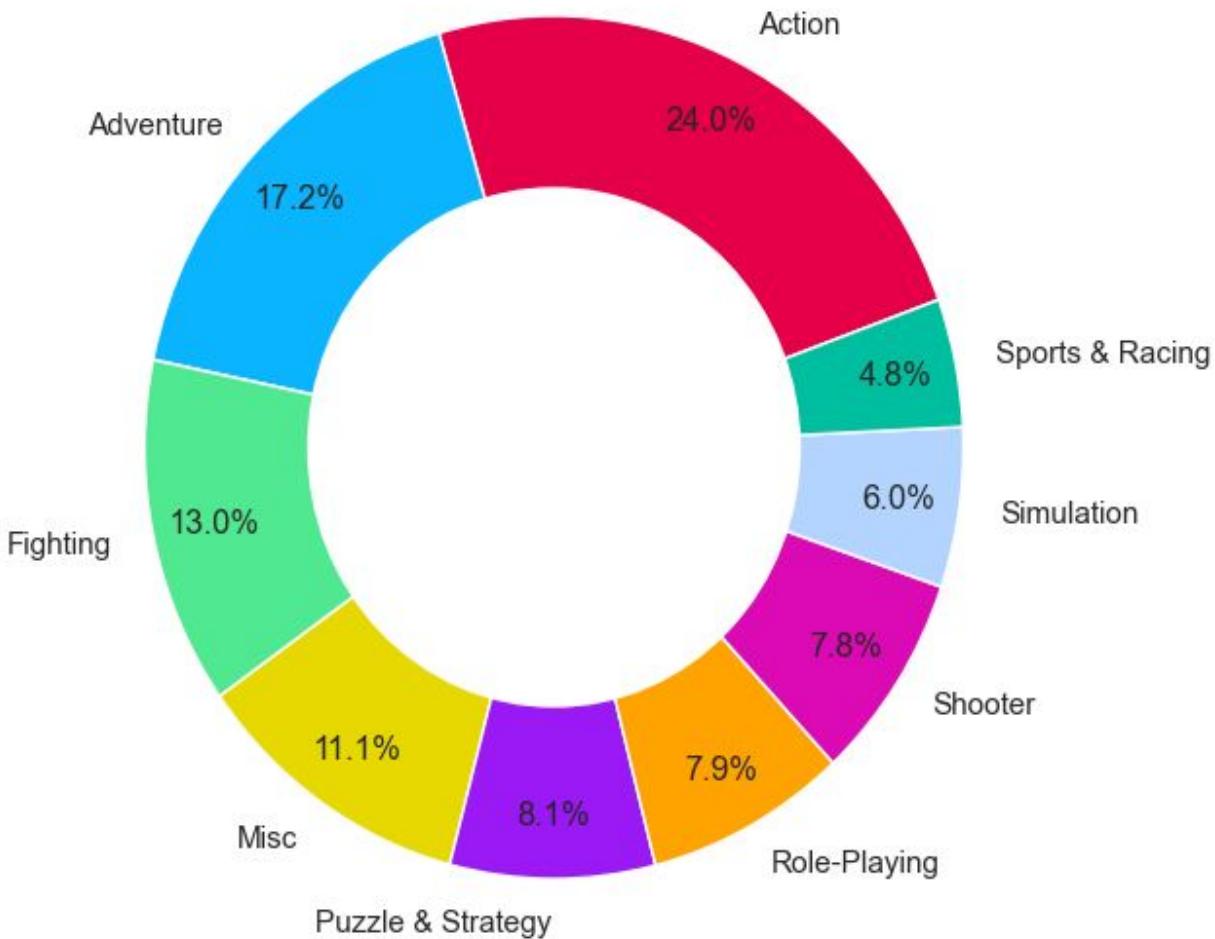
1980 - 1989 Videogame Genre



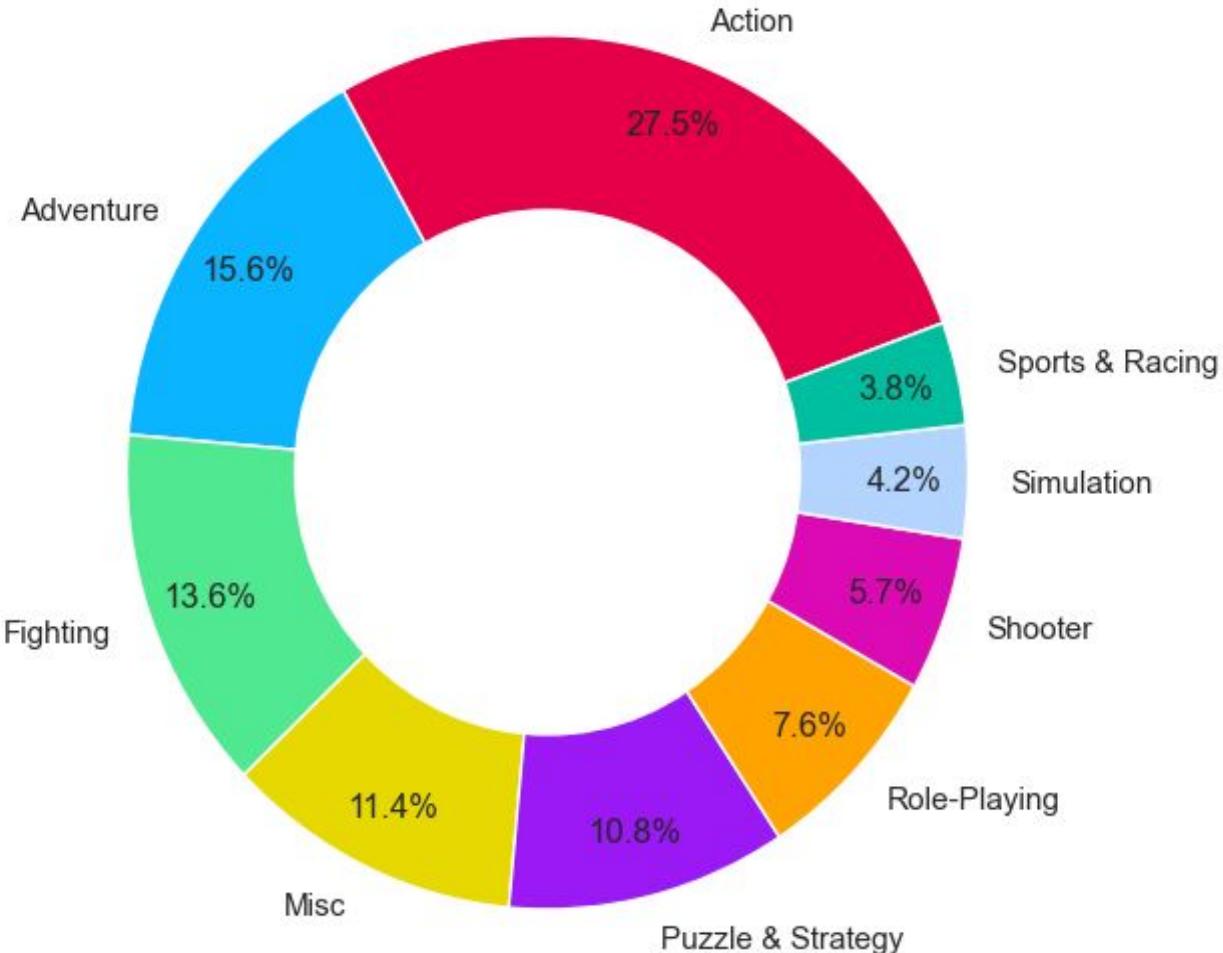
1990 - 1999 Videogame Genre



2000 - 2009 Videogame Genre



2010 - 2015 Videogame Genre



Outlier Exploration for 2010s

There were 1320 total Action games with 143 considered outliers. 10.8% were outliers selling greater than 0.5 million units.

There were 650 total Adventure games with 91 considered outliers. 14.0% were outliers selling greater than 0.18 million units.

There were 183 total Fighting games with 12 considered outliers. 6.6% were outliers selling greater than 0.58 units.

There were 545 total Misc games with 51 considered outliers. 9.4% were outliers selling greater than 0.45 units.

There were 274 total Puzzle & Strategy games with 23 considered outliers. 8.4% were outliers selling greater than 0.2 units.

**There were 516 total Role-Playing games with 64 considered outliers.
12.4% were outliers selling greater than 0.4 million units.**

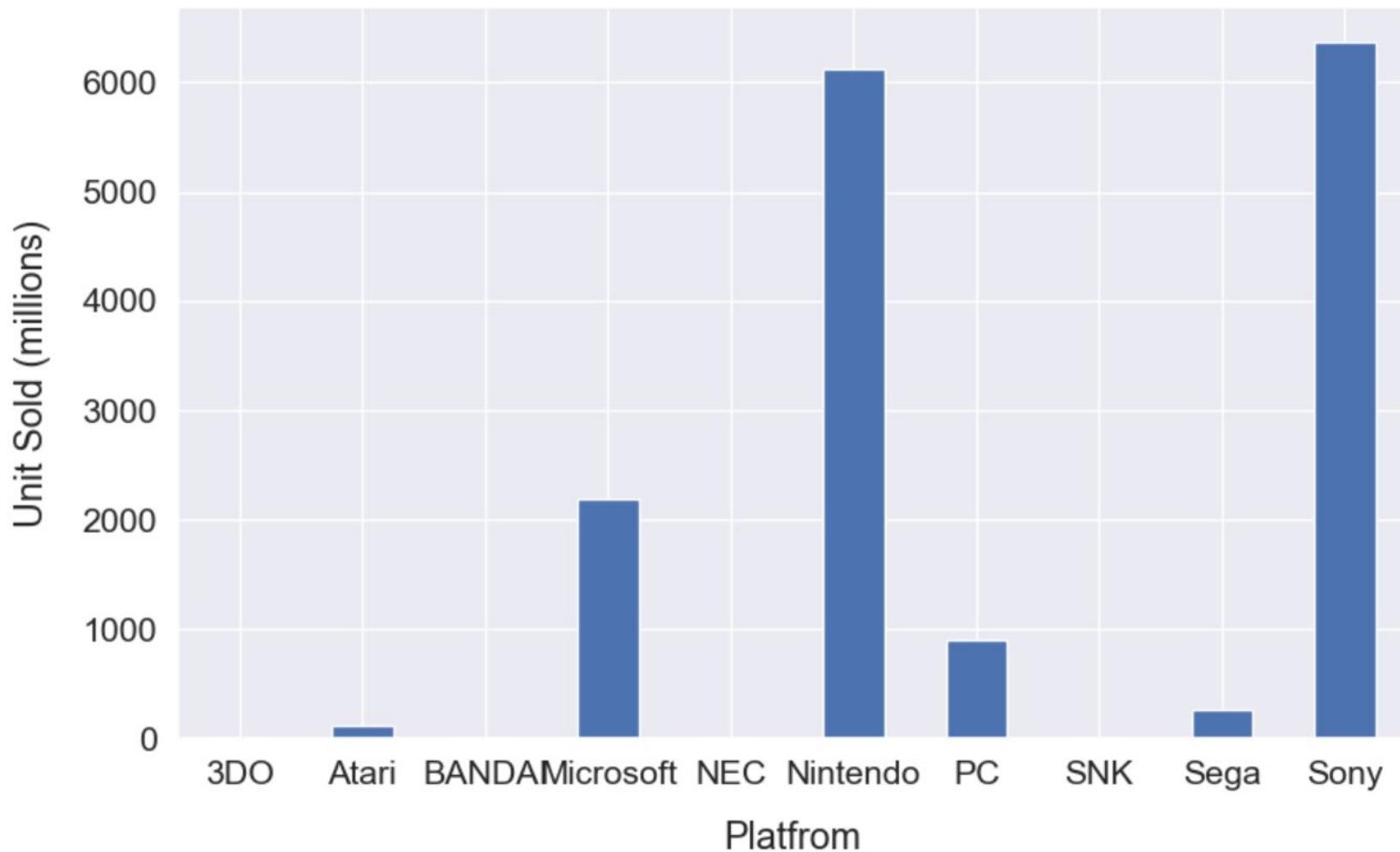
There were 363 total Shooter games with 45 considered outliers. 12.4% were outliers selling greater than 1.24 million units.

There were 200 total Simulation games with 24 considered outliers. 12.0% were outliers selling greater than 0.31 units.

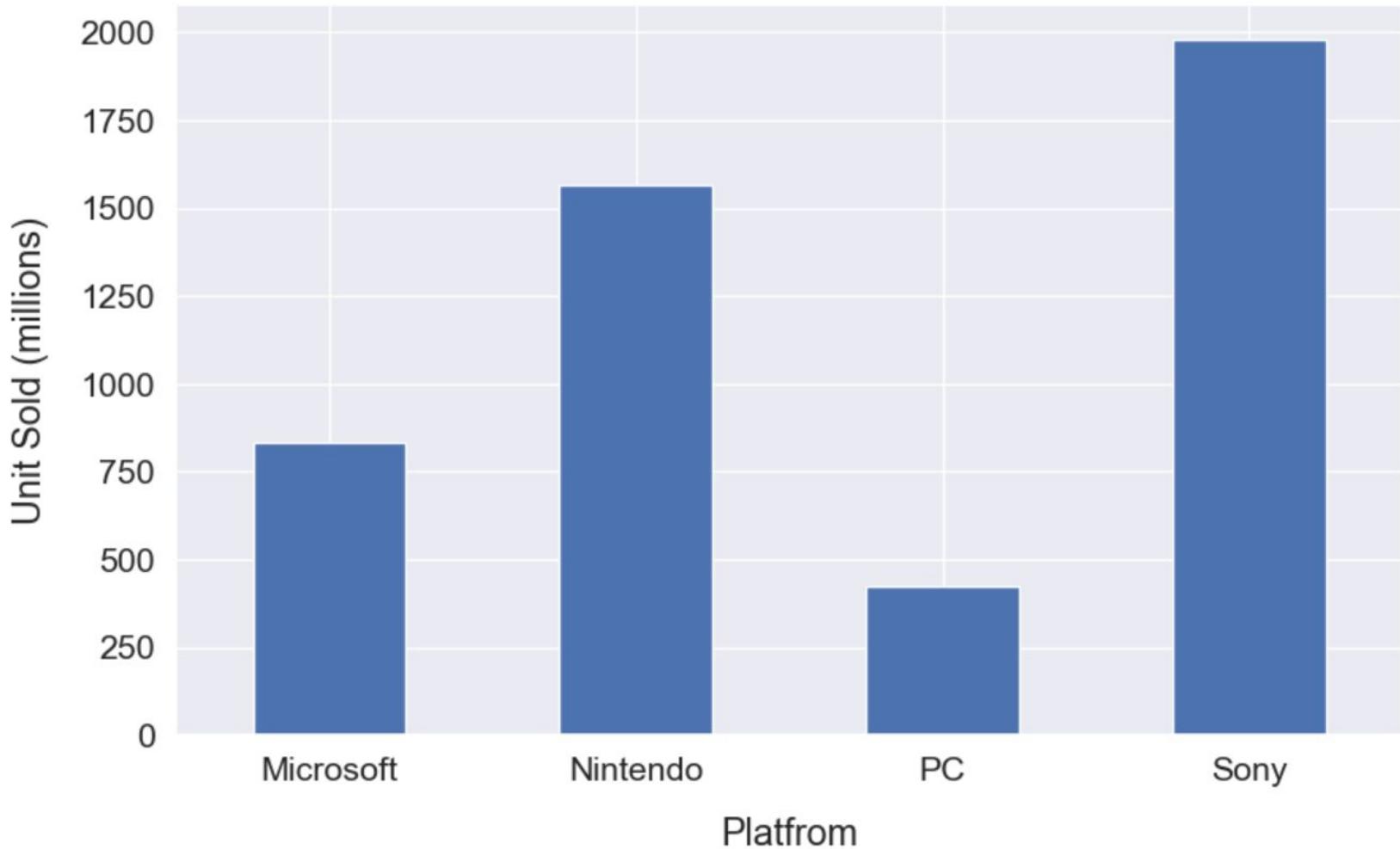
There were 746 total Sports & Racing games with 78 considered outliers. 10.5% were outliers selling greater than 0.67 million units.

Genre By Platform Company

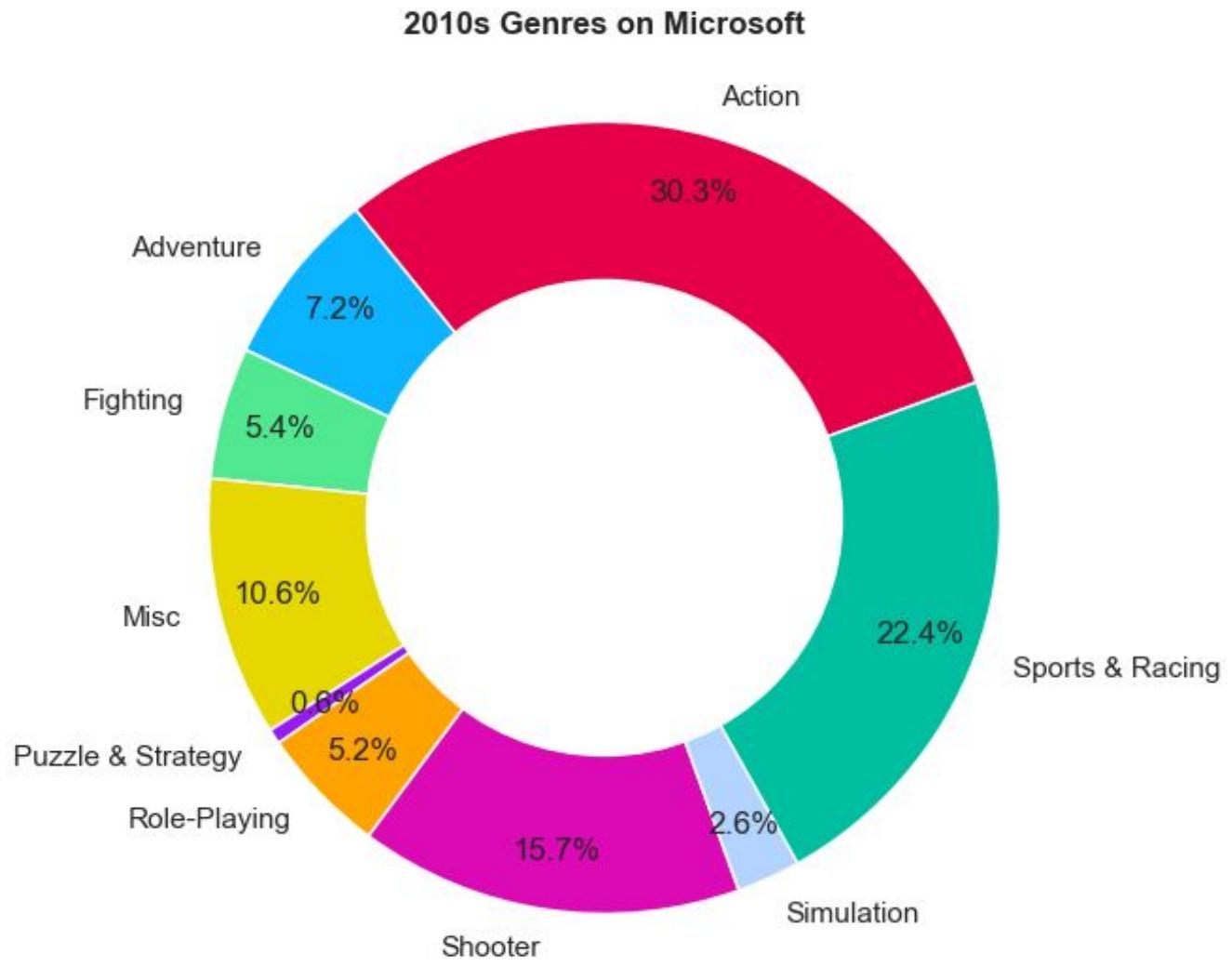
Unit Sold by the Platfrom

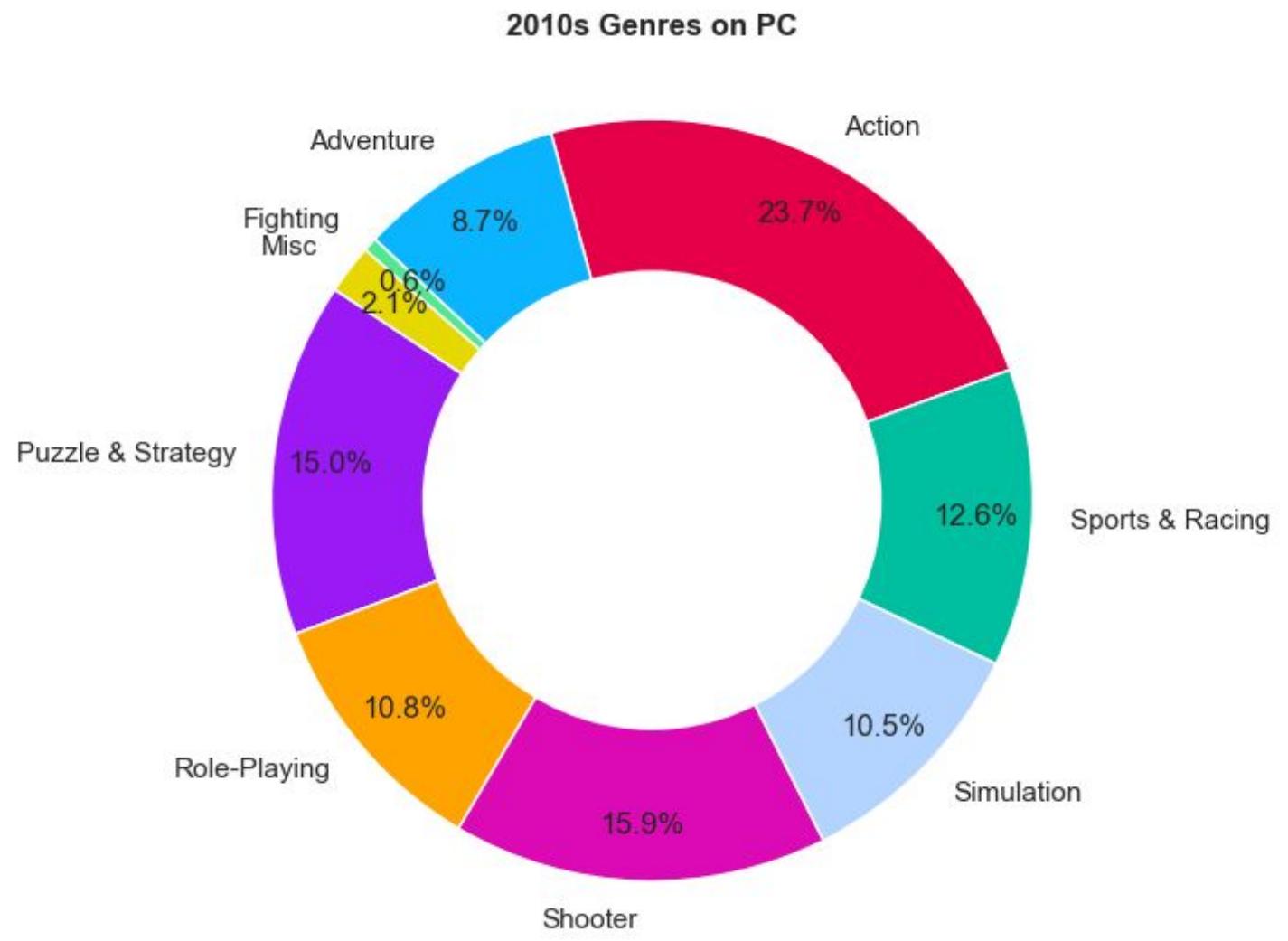


Unit Sold by the Platfrom - 2010s

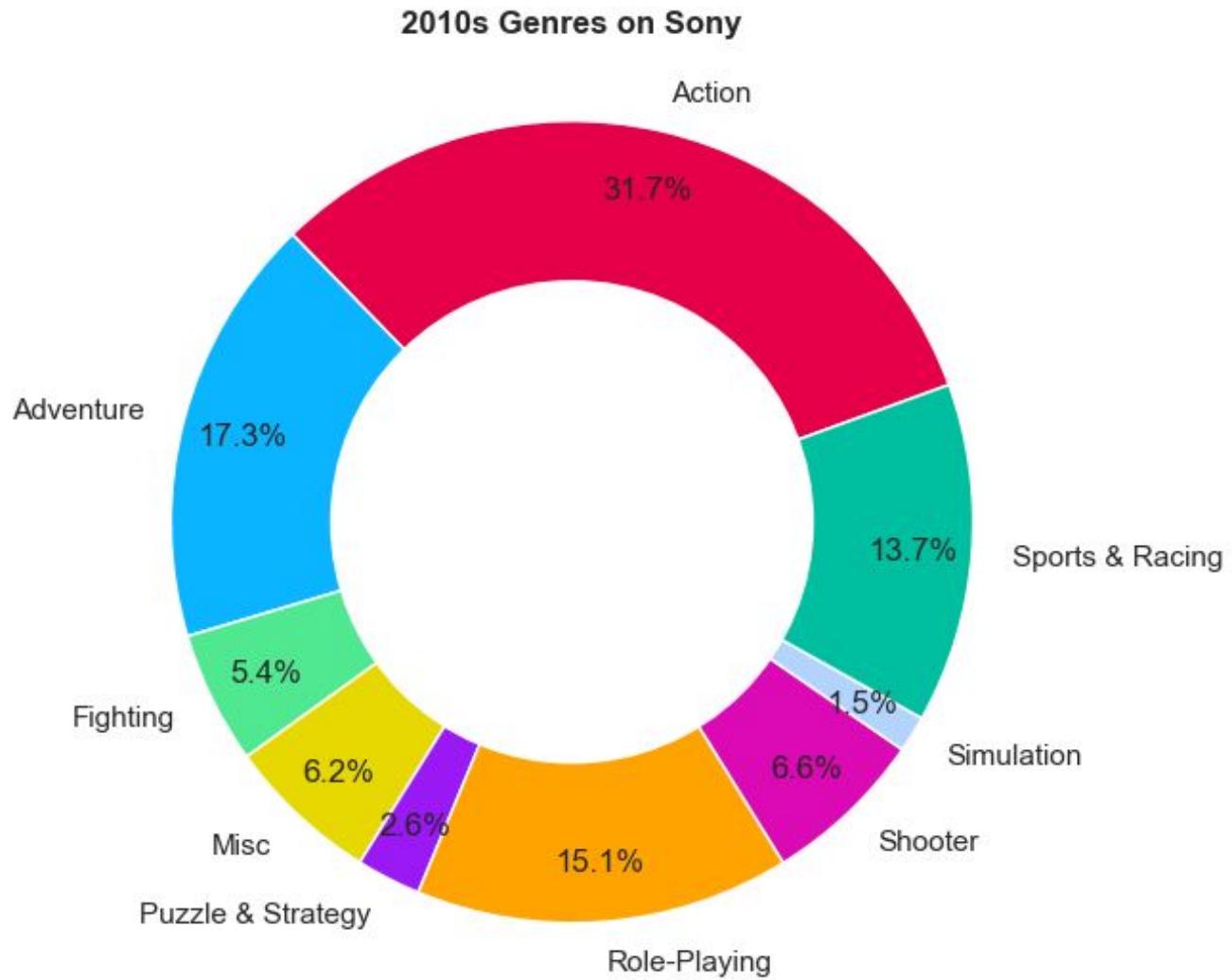


Microsoft

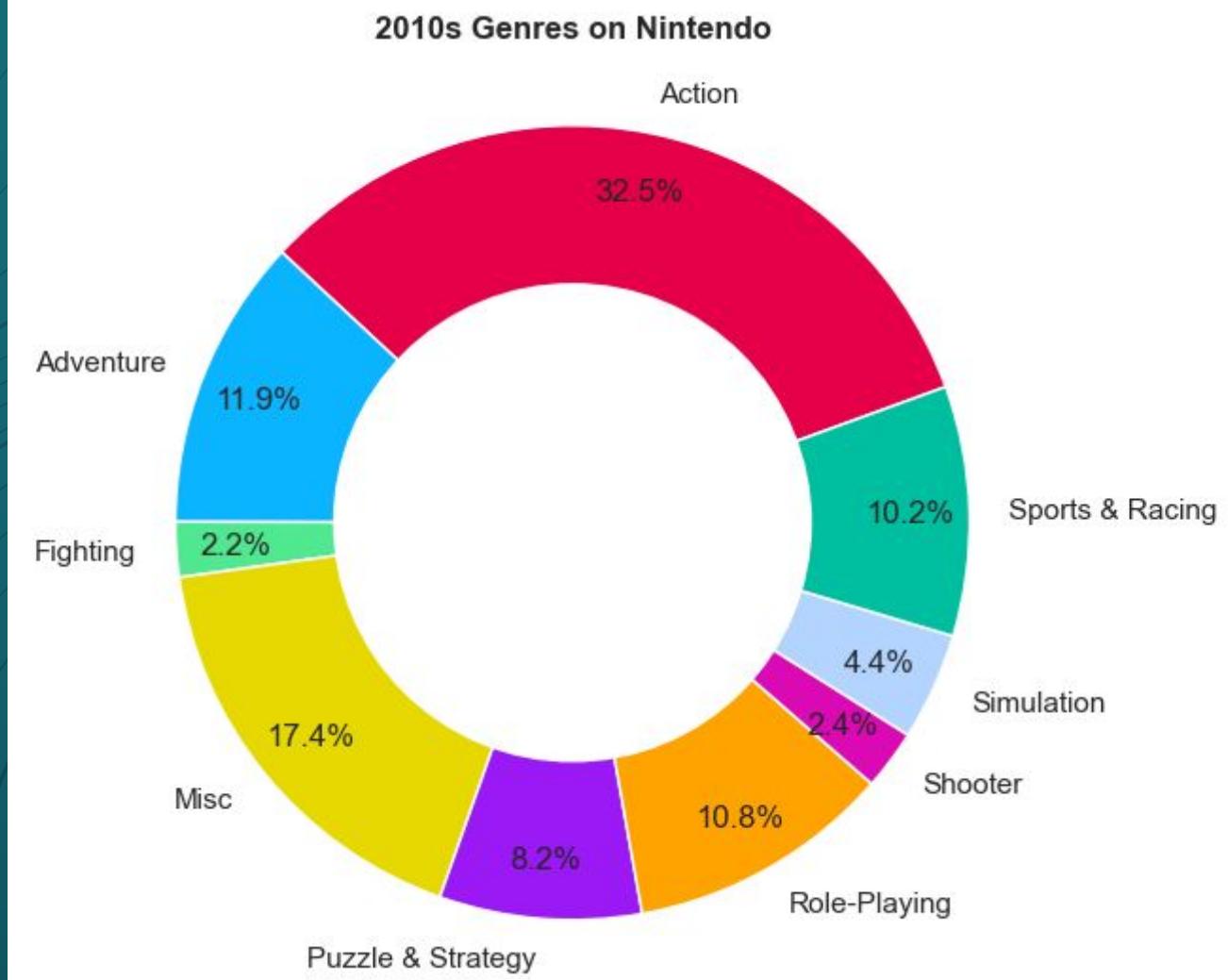




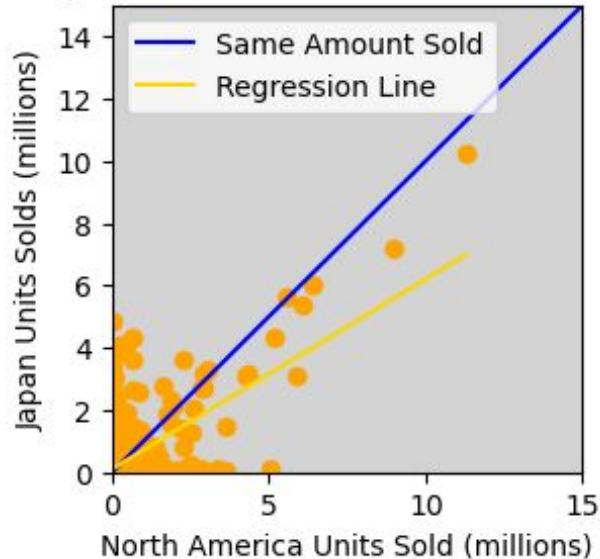
Sony



Nintendo



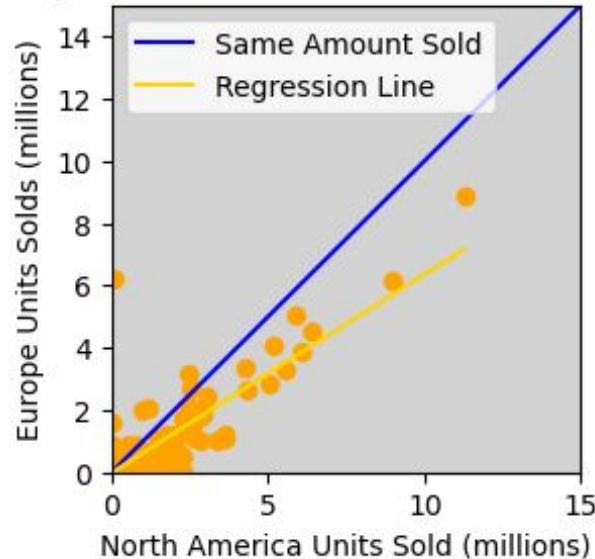
Role-Playing Units Sold in North America vs Japan



Linear Regression: $y = 0.61x + 0.10$

R-squared: 0.402

Role-Playing Units Sold in North America vs Europe



Linear Regression: $y = 0.64x - 0.01$

R-squared: 0.767

Global
TOP
GAMES

1st
Pokemon Red / Blue (Green)
Game Boy 1996 Nintendo

2nd
Pokemon Gold / Silver Game
Boy 1999 Nintendo

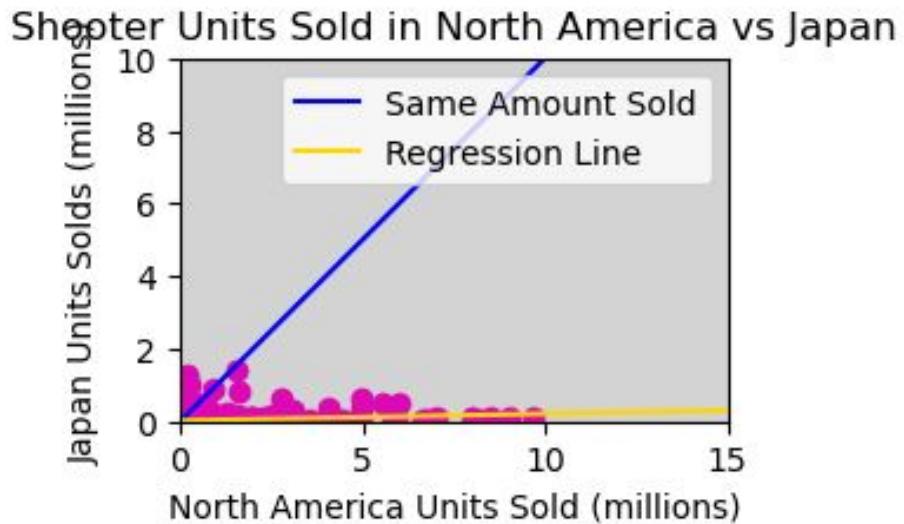
3rd
Pokemon Diamond / Pearl
DS 2006 Nintendo

Standouts

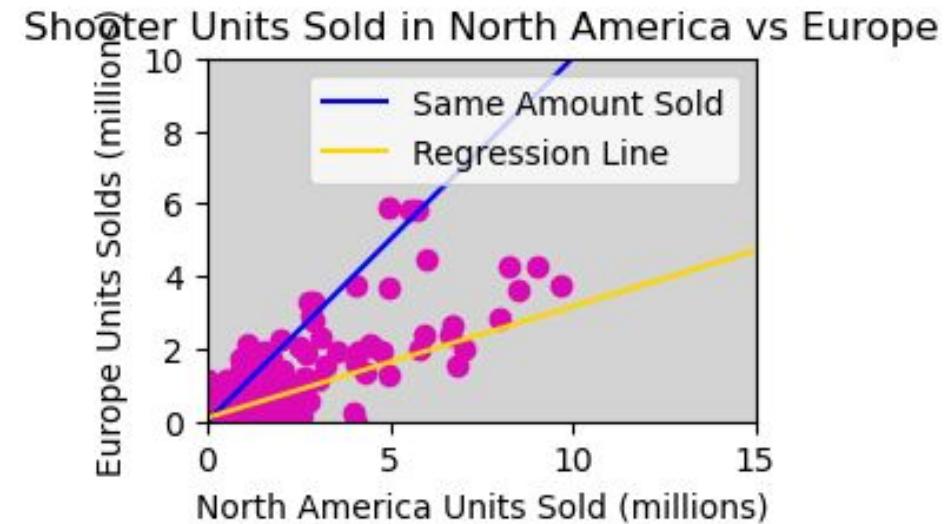
North America
The Elder Scrolls V: Skyrim
Multi 2011 Bethesda Softworks

Japan
Monster Hunter Freedom 3
PSP 2010 Sony

Europe
World of Warcraft
PC 2004 Activision



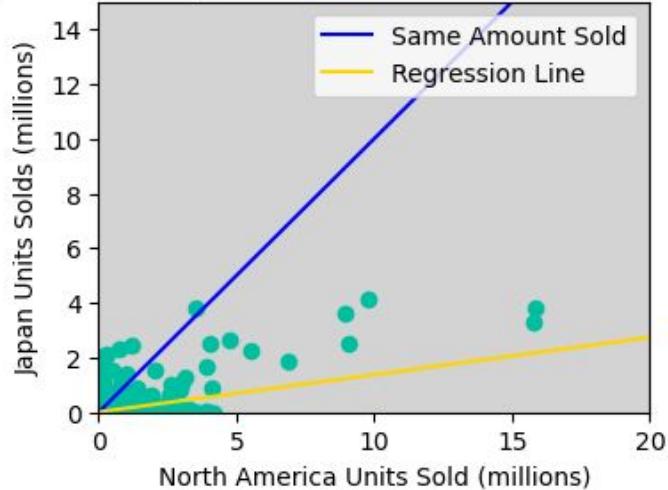
Linear Regression: $y = 0.02 x + 0.02$ R-squared: 0.048



Linear Regression: $y = 0.30 x + 0.10$ R-squared: 0.455

Global TOP GAMES	1st Dunk Hunt (Outlier) Game Boy 1996 Nintendo	2nd Call of Duty: Modern Warfare 3 Multi 2011 Activision	2nd Call of Duty: Black Ops Multi 2010 Activision
Standouts	Japan Splatoon WiiU 2015 Nintendo	Series Call of Duty: Modern Warfare & Black Ops Multi 2007- 2015 Activision	Series Halo Xbox Systems 2001 - 2012 Microsoft Game Studios

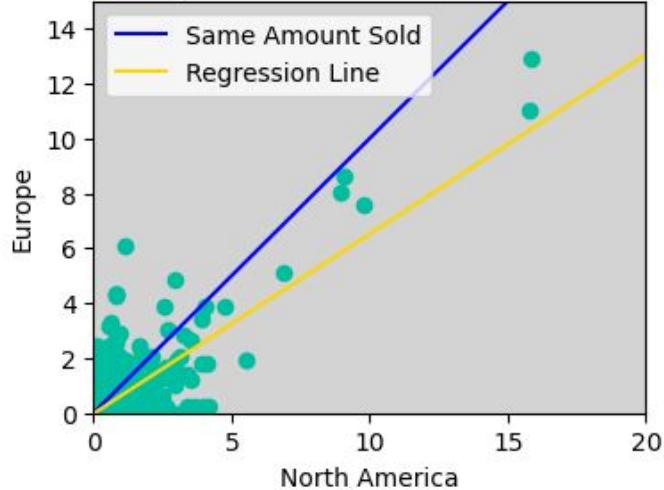
Sports & Racing Units Sold in North America vs Japan



Linear Regression: $y = 0.13x + 0.01$

R-squared: 0.290

Sports & Racing Units Sold in North America vs Europe



Linear Regression: $y = 0.65x - 0.02$

R-squared: 0.754

Global TOP GAMES

1st
Wii Sports (Outlier)
Wii 2006 Nintendo

2nd
Mario Kart Wii
Wii 2008 Nintendo

2nd
Wii Sports Resort
Wii 2009 Nintendo

Standouts

Series
Mario Kart
Nintendo Systems 1992 - 2014
Nintendo

Series
FIFA
Multi 1995 - 2015 Electronic Arts

Series
Gran Turismo
Sony Systems 1997 - 2013
Sony

Outliers

Wii sports - top sports (bundled with console launch)

Duck hunt - top shooter (bundled with console launch)

Mario Kart Wii - 2nd top sports (released with Wii Wheel)

T-Test Comparing North American Shooter Sales (Code)

```
1 pop_x = shooter.NA_Sales
2 pop_y = rpg.NA_Sales
3 print(pop_x.mean())
4 print(pop_x.var())
5 print(len(pop_x))
6 print(pop_y.mean())
7 print(pop_y.var())
8 print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.22750874737578727
0.469441408361786
1429
```

```
1 pop_x = shooter.NA_Sales
2 pop_y = sports.NA_Sales
3 print(pop_x.mean())
4 print(pop_x.var())
5 print(len(pop_x))
6 print(pop_y.mean())
7 print(pop_y.var())
8 print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.29438940092165894
0.9234260171506634
3472
```

```
1 pop_x = shooter.NA_Sales
2 pop_y = shooter.EU_Sales
3 print(pop_x.mean())
4 print(pop_x.var())
5 print(len(pop_x))
6 print(pop_y.mean())
7 print(pop_y.var())
8 print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.2422
0.3093768214571657
1250
```

```
1 st.ttest_ind(pop_x.values, pop_y.values, equal_var=False)
```

```
Ttest_indResult(statistic=5.6097667673362075, pvalue=2.3336277662001025e-08)
```

T-Test Comparing North American Shooter Sales

	NA Shooters	vs EU Role-Playing	vs JP Role-Playing	vs EU Shooters	vs JP Shooters	vs EU Sports & Racing	vs JP Sports & Racing
Mean (million)	0.454	0.130	0.243	0.242	0.030	0.173	0.055
Variance	1.50	0.25	0.43	0.31	0.01	0.52	0.06
Total	1250	1429	1429	1250	1250	3472	3472
P-Value	Null	≈0 (6.08e-18)	≈0 (5.58e-8)	≈0 (2.95e-8)	≈0 (2.03e-32)	≈0 (3.24e-14)	≈0 (6.26e-29)

	Mean (million)	Variance	Total	P-Value
NA Shooters	0.454	1.50	1250	Null
vs NA Role-Playing	0.294	0.92	3472	0.0000316
vs NA Sports & Racing	0.227	0.47	1429	0.0000000794

Results

- Games bundled with console sales outperform single sales.
- Long running franchises and series have a built up following and maintain repeat buyers.
- New unique games with special novelties also are a good way to break into the market.

Prediction

A game needs to be successful in NA to be successful globally, and if you do a combination of RPG and Shooter.



FUTURE and Data Bias

Future searches:

- Multi-platform games
- Correlation of Handhelds and RPGS
- Combine data from long running games series like:
 - FIFA
 - Call of Duty
 - Halo
 - Madden
 - Pokemon
- Mobile games

Data Bias:

- Selection bias (data doesn't include all games)
- Availability Bias (Date after 2015 is not available)
- Streaming and Downloads substituting Physical Units.
- Emergence of Mobile Games in late 2010s