

Section 1. Project

Analysis of Global Game Data

Codestates_AI_01_최중훈

2021. 01. 22

INDEX

Q_1. 지역에 따라 선호하는 게임 장르

Q_2. 연도별 게임 트렌드

Q_3. 매출 top 30 게임

Q_4. 게임 제작 회사와 국가가 매출에 끼치는 영향

Q_5. 게임 출시 횟수와 매출액

Data Description

총 16,598개의 출시 게임 중 결측치를 제외한 16,277개의 출시 게임 데이터 사용

기간 : 1980 - 2020 (2010년 이후 데이터의 비중이 낮음)

Feature

Name	Platform	Genre	Year	Publisher
Sales in North America	Sales in Europe	Sales in Japan	Sales in Others	

Q_1. 지역에 따라 선호하는 게임 장르

Process

Pre-processing

Genre
vs
Region

Two Samples
X2-test

Visualization

Q_1. 지역에 따라 선호하는 게임 장르

Two samples X2-test 시행 : Genre Vs. Sales in Region

- 귀무가설 : 각 지역 매출액과 장르 간 연관성이 없을 것이다.
- 대안가설 : 각 지역 매출액과 장르 간 연관성이 있을 것이다.

테스트 결과 :

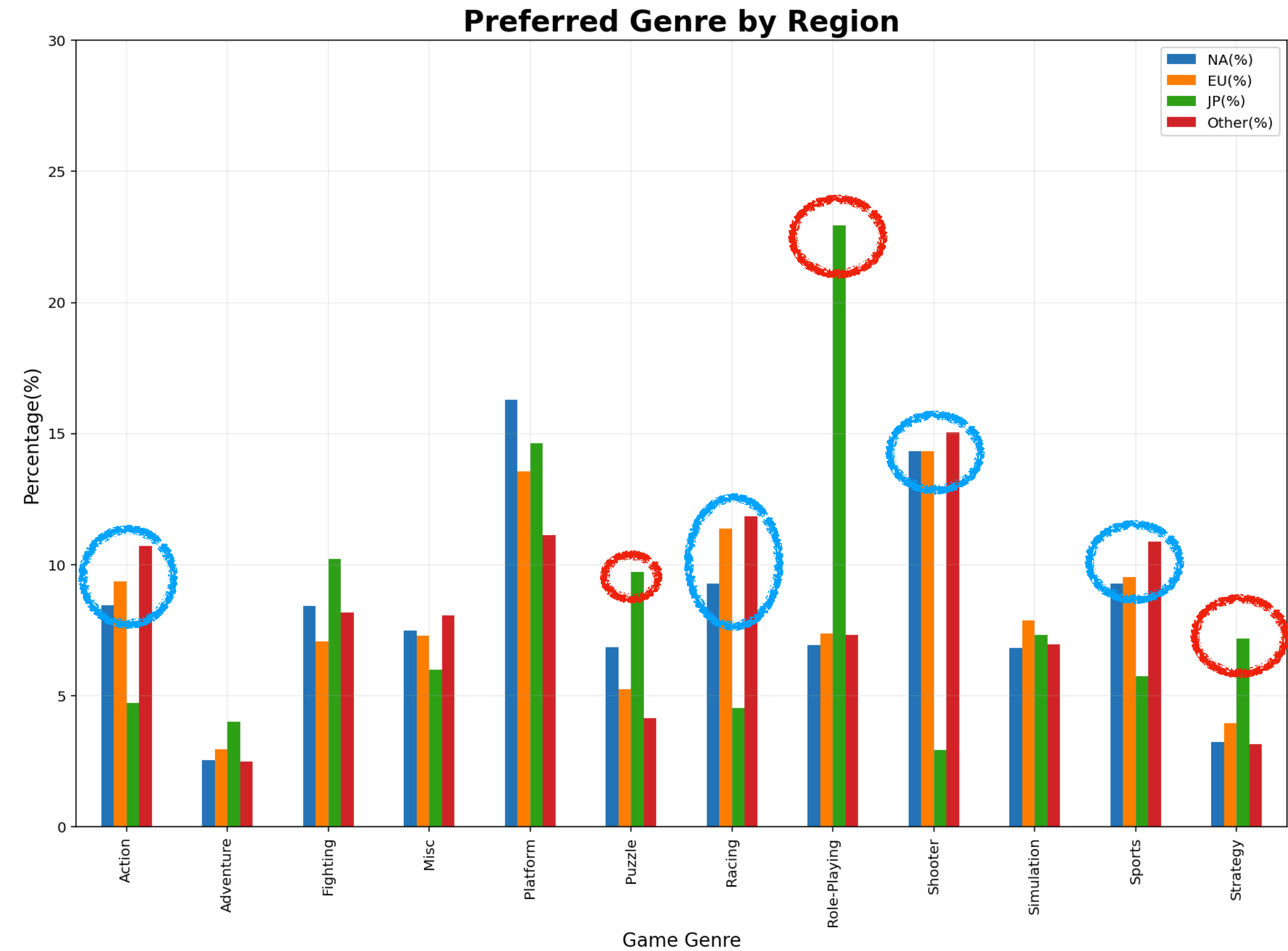
North America	P-value : 7.549444125974192e-63
Europe	P-value : 6.8167487877648e-38
Japan	P-value : 5.08736657649697e-172
Other Countries	P-value : 1.0672942341351104e-19



북미, 유럽, 일본, 기타 국가 모두 X2-test 결과 **p-value 값이 0**이므로
신뢰수준 95%에서 귀무가설 "지역 매출액과 장르 간 연관성이 없을 것이다"라는 귀무가설을 기각할 근거가 충분하다.

Q_1. 지역에 따라 선호하는 게임 장르

Visualization



- 지역에 상관없이 선호하는 장르 : Platform, Simulation
- 일본에서 유독 선호하는 장르 : RPG, Puzzle, Strategy
- 일본을 제외한 지역에서 선호하는 장르 : Shooter, Racing, Sports, Action

▶ “지역별 선호하는 장르가 다를까?” 라는 물음에
그럴 가능성이 높다는 결론을 내릴 수 있음

Process

Pre-processing

Year
vs
Platform, Genre

Two Samples
X2-test

Visualization

Q_2. 연도별 게임 트렌드

Two samples X2-test 시행 : Year Vs. Platform, Genre

- 연도별 게임 출시 트렌드 분석
- 귀무가설 : 게임 출시연도와 플랫폼 & 장르 간 연관성이 없을 것이다.
- 대안가설 : 게임 출시연도와 플랫폼 & 장르 간 연관성이 있을 것이다.

테스트 결과 :

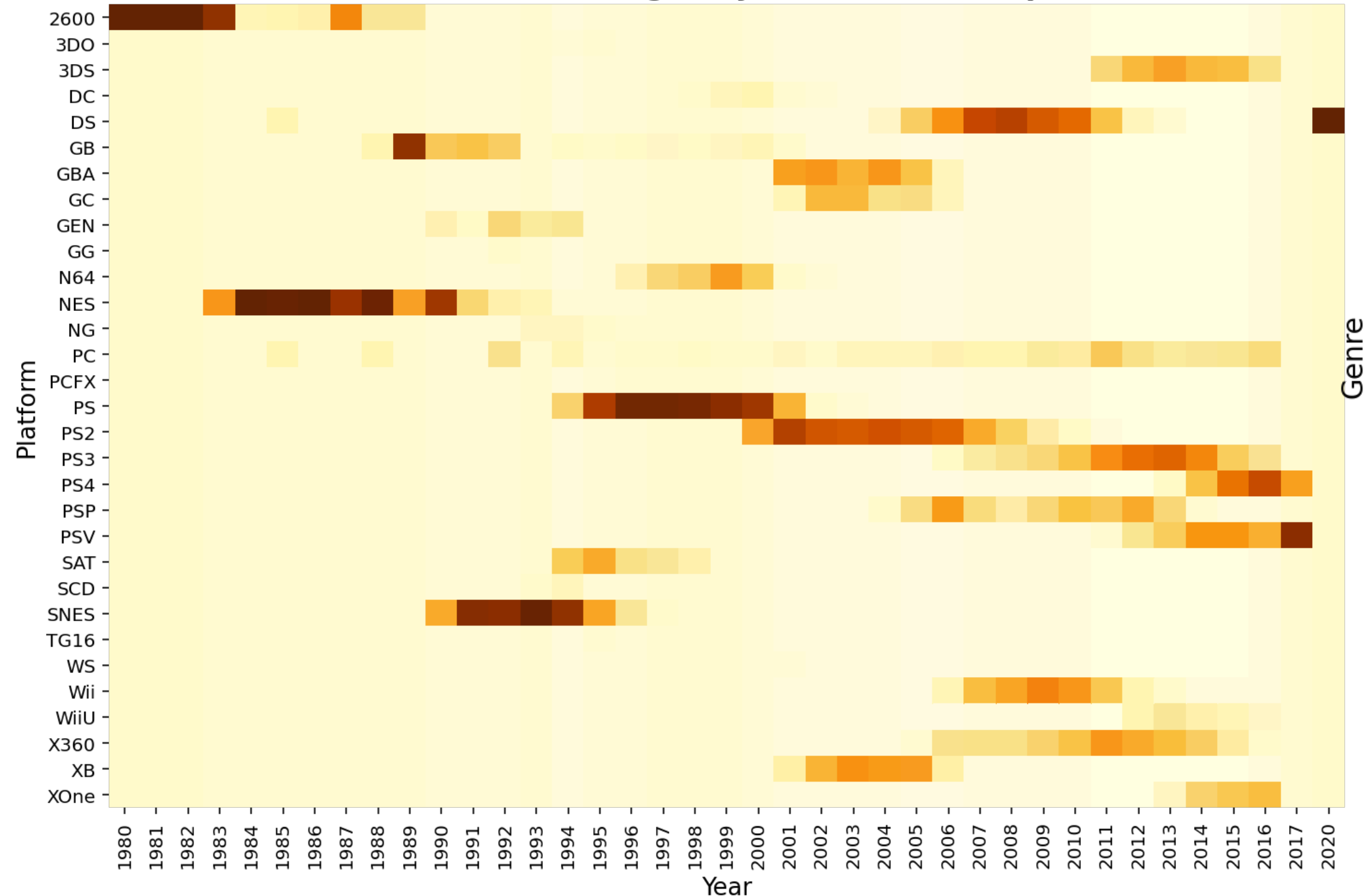
Platform by Year	P-value : 0
Genre by Year=	P-value :0

▶ 출시연도 Vs. 플랫폼, 장르 둘다 X2-test 결과 **p-value** 값이 **0**이므로
신뢰수준 95%에서 귀무가설 "게임 출시연도와 플랫폼&장르 간 연관성이 없을 것이다."라는 귀무가설을 기각할 근거가 충분하다.

Q_2. 연도별 게임 트렌드

Visualization

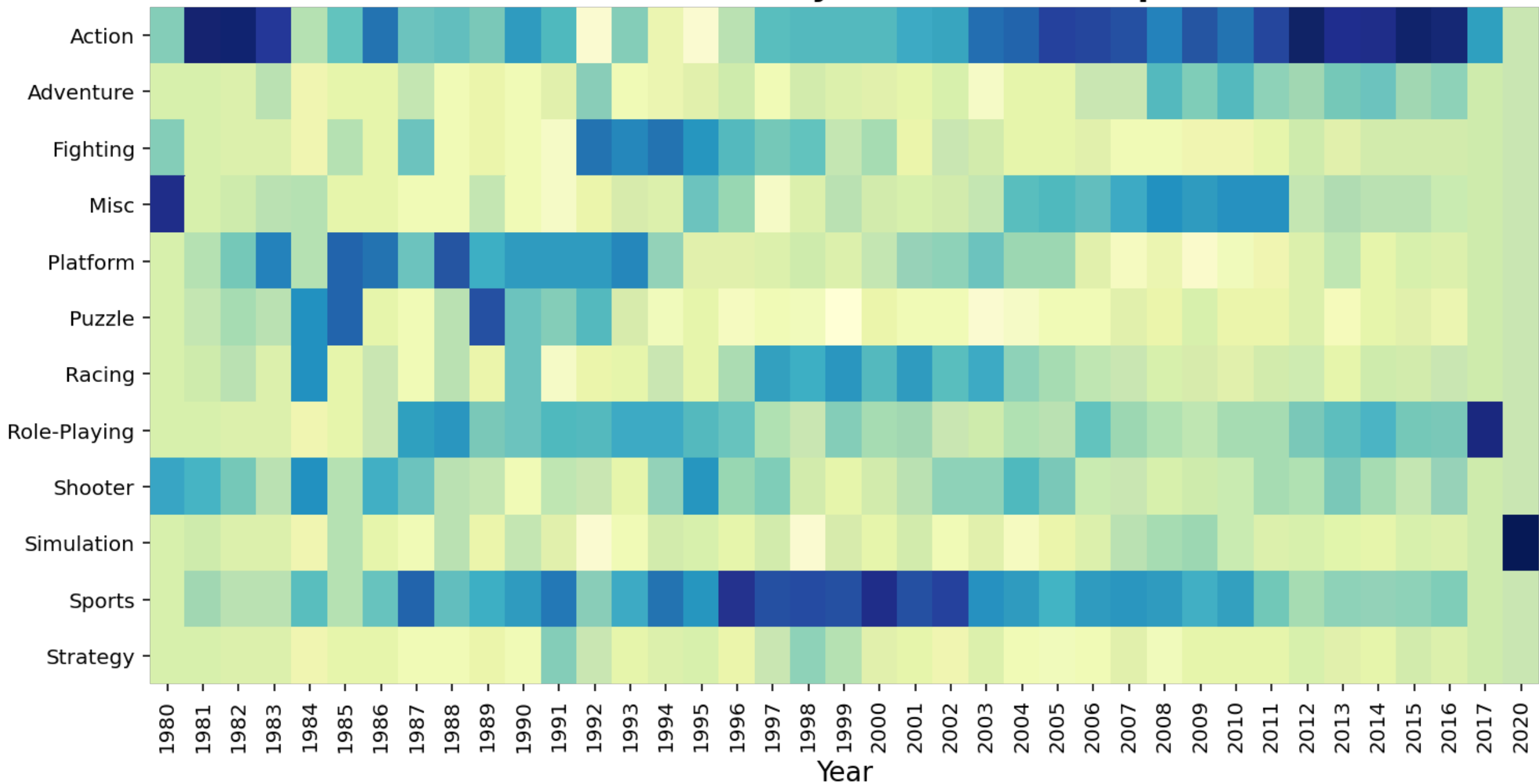
Platform Changes by Year on heatmap



플랫폼

- 1. 초창기 : 아리타2600 독점
- 2. 1980s 닌텐도, 1990s PS, 2000s Xbox 출시
- 3. 현재 3강 브랜드에 PC까지 크게 4개의 플랫폼이 형성됨
- 4. 2010s 들어 3DS, PSP, PSV 등 휴대성을 필두로 한 휴대용 게임기의 점유율이 높아짐

Genre Trends by Year on heatmap



장르

- 1. Action 장르의 꾸준한 점유율
- 2. 2000s 이후 Sports 장르의 약세
- 3. 최근 Adventure, RPG, Shooter 등 고성능 게임의 출시 경향이 높아짐

▶ 연도별 게임 플랫폼의 발달에 따른 게임 출시 트렌드가 존재한다.

Q_3. 매출 top 30 게임

Process

Pre-processing

: 매출 top30 선별

Whole Sales
vs
Regional Sales

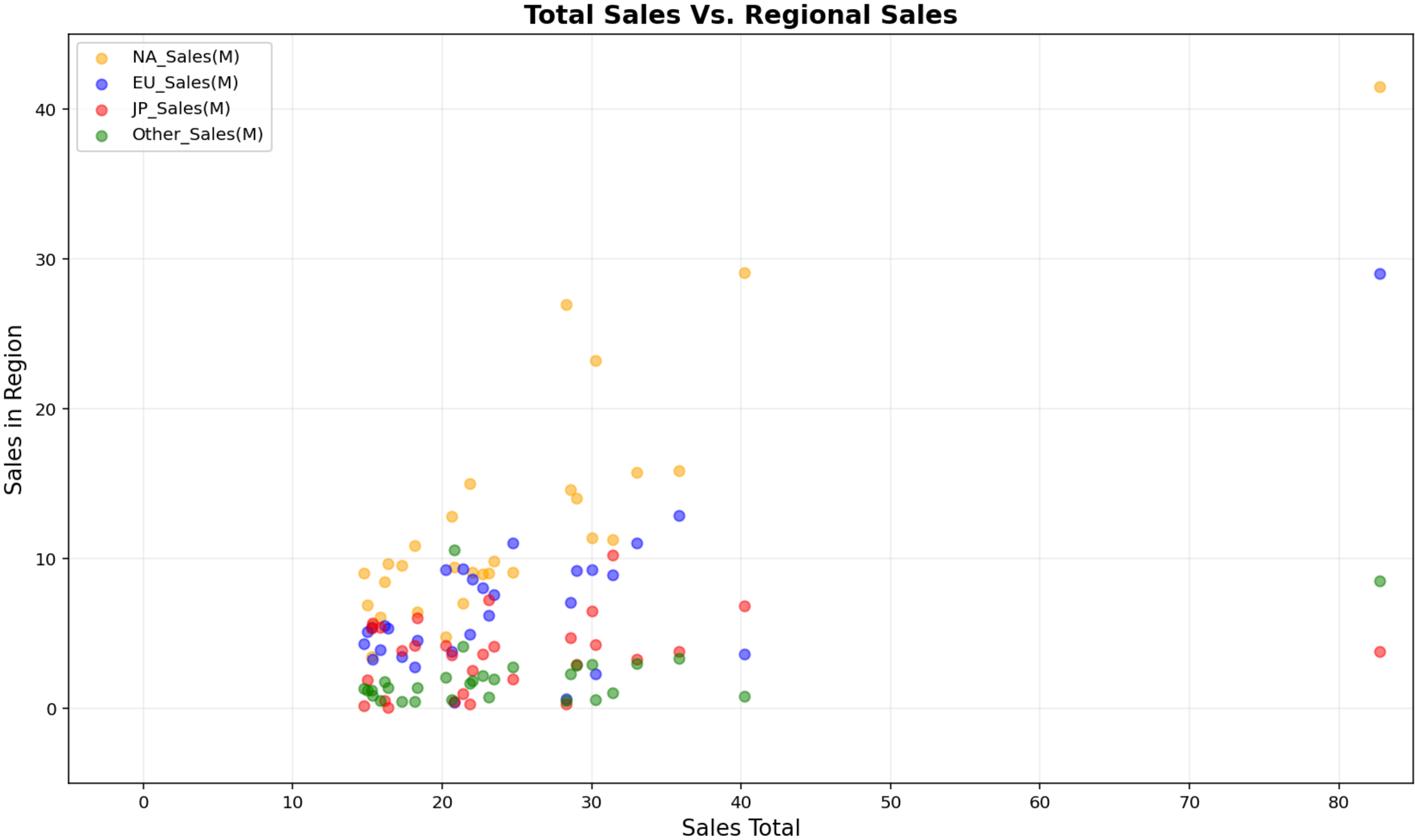
Choose
Most-related
Region

Visualization

Q_3. 매출 top 30 게임

전체 매출액과 지역별 매출액의 상관관계 분석

- 어느 지역을 타겟으로 해야 할 지 명확히 확인할 수 있을 것이라는 판단



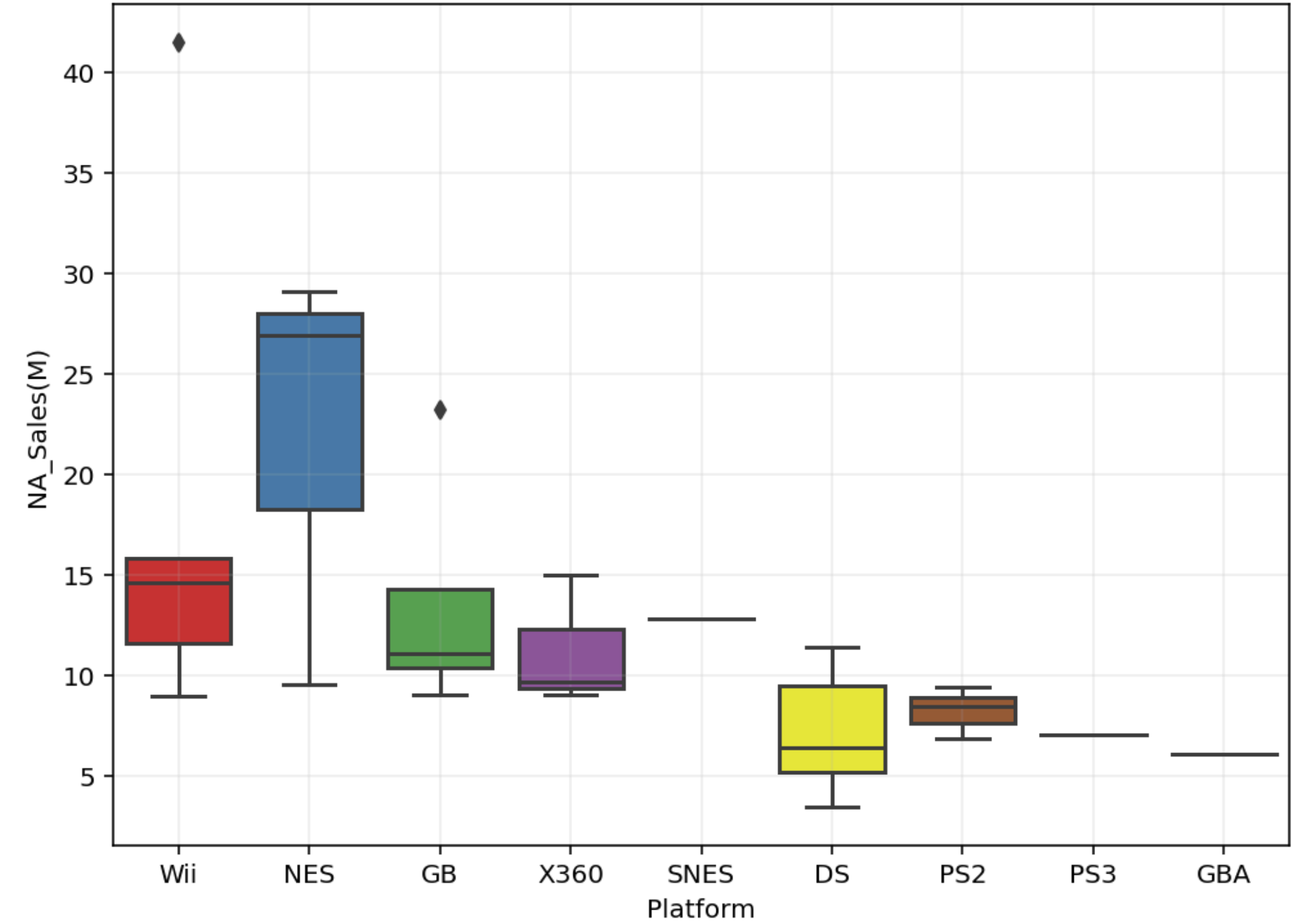
[상관관계 지수 확인]

North America	0.868
Europe	0.799
Japan	0.187
Other Countries	0.497

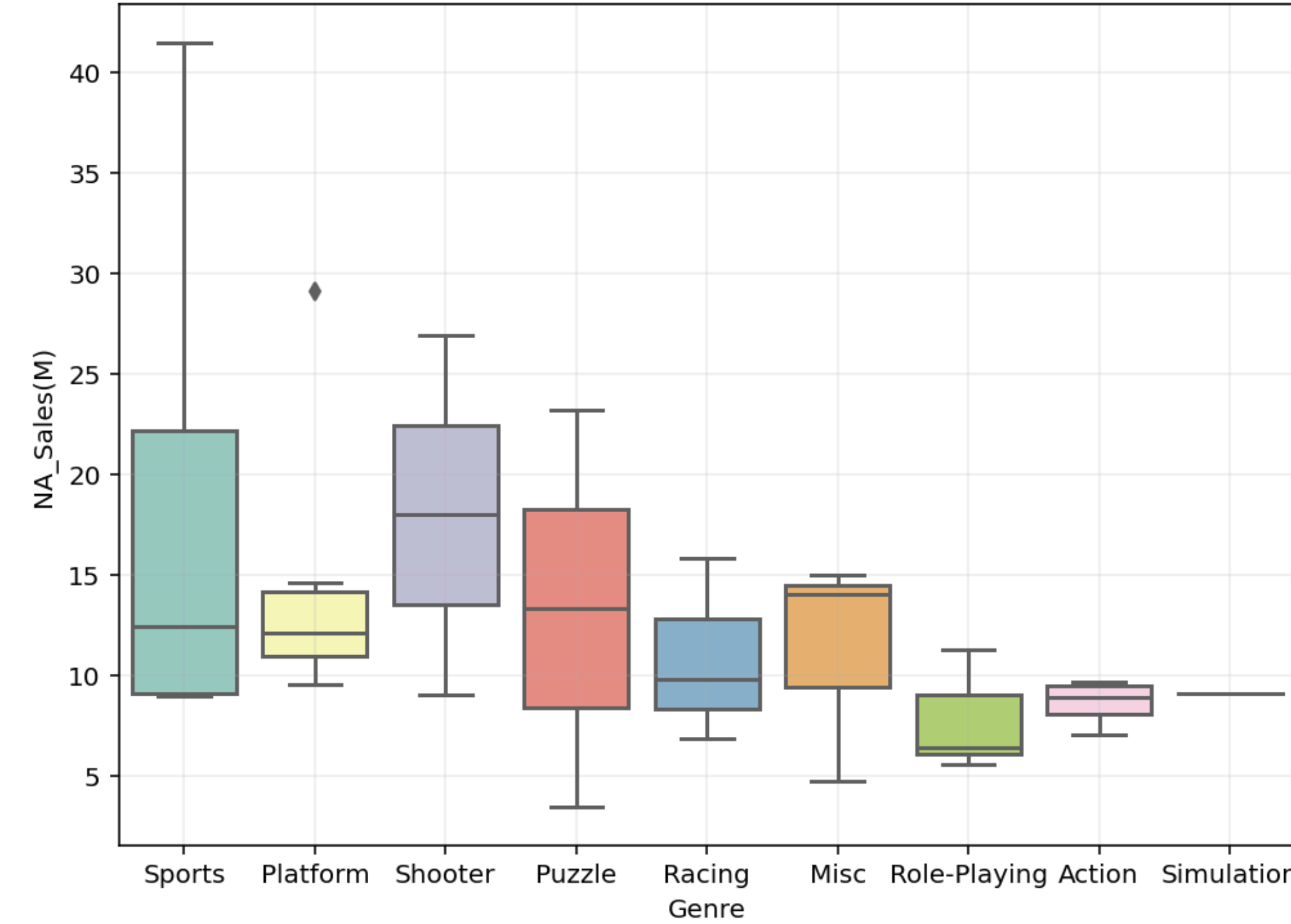
Q_3. 매출 top 30 게임

북미지역에서 선호하는 플랫폼 및 장르 추가 분석 실시

- 1. 북미 지역 매출 top30 추출
- 2 “top30의 플랫폼 및 장르” 시각화

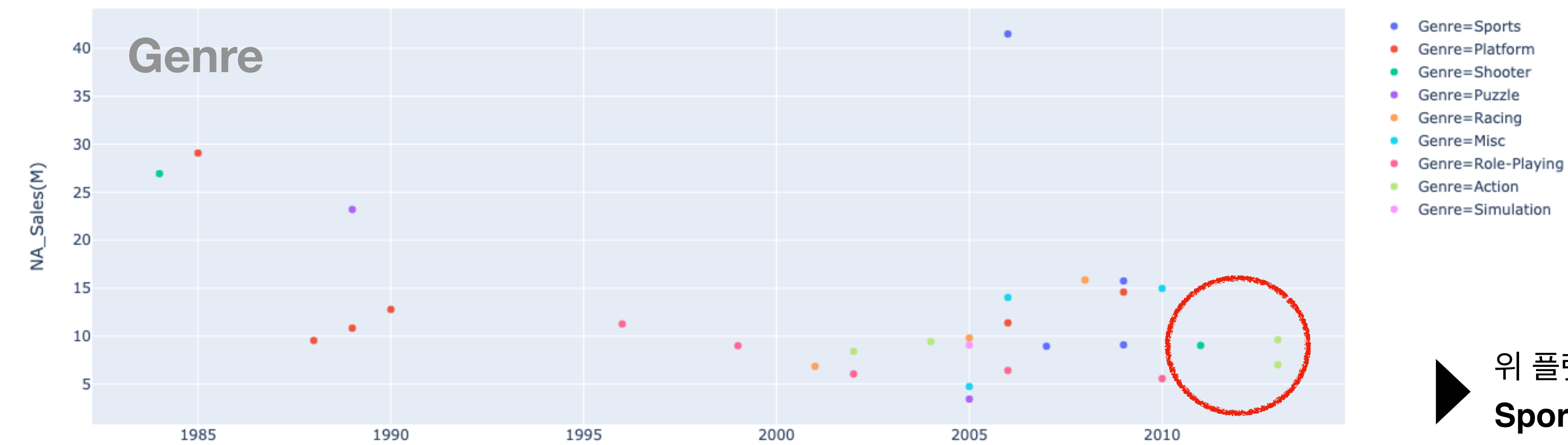
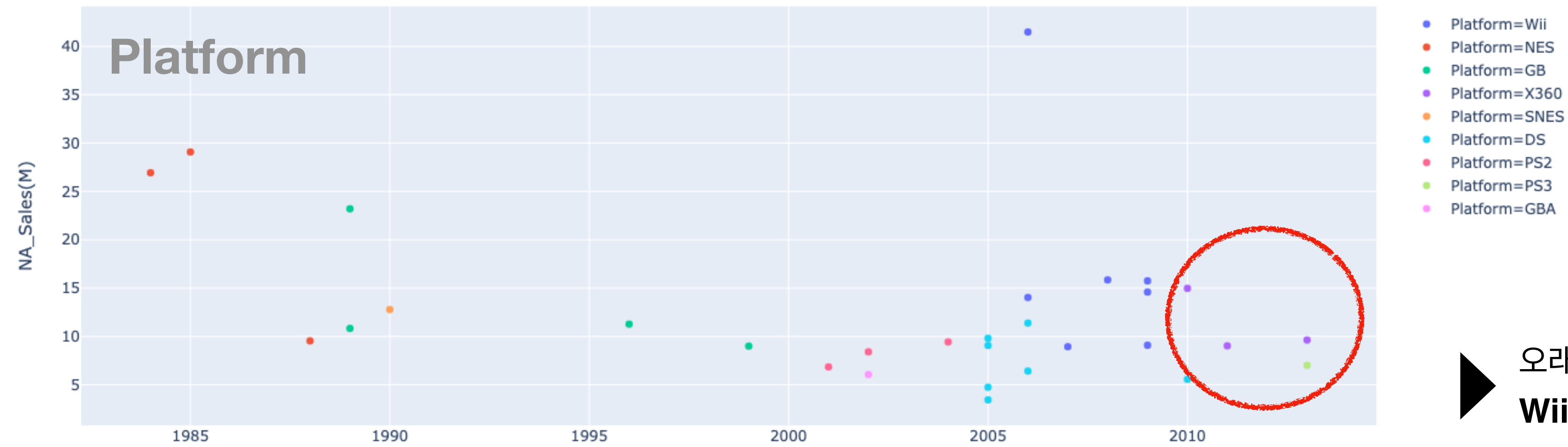


NES, GB 플랫폼이 높은 수치를 가지지만,
플랫폼 부문은 출시 연도를 고려해야만 함



북미 지역은 sports, shooter, racing, action 등
액티브한 게임을 선호하는 경향이 있음을 알 수 있음

Q_3. 매출 top 30 게임



Q_4. 게임 제작 회사와 국가가 매출에 끼치는 영향

Process

Pre-processing

: 매출 상위 30개
회사의 국가 조사
: 국가 Feature 추가

Pre-processing

: 전체 데이터에서
매출 상위 30개에 해당하는
회사가 제작한 게임 추출

K-means Clustering

Accuracy 측정

Q_4. 게임 제작 회사와 국가가 매출에 끼치는 영향

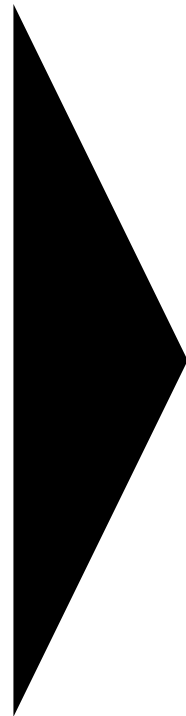
1. 매출 top30 회사의 국가 조사 및 전체 데이터에 적용

2. 매출 top30에 해당하지 않는 회사는 삭제

Company	Country
Nintendo	Japan
Microsoft Game Studio	USA
Take-Two Interactive	USA
Sony Computer Entertainment	Japan
Activision	USA

4. Accuracy 측정

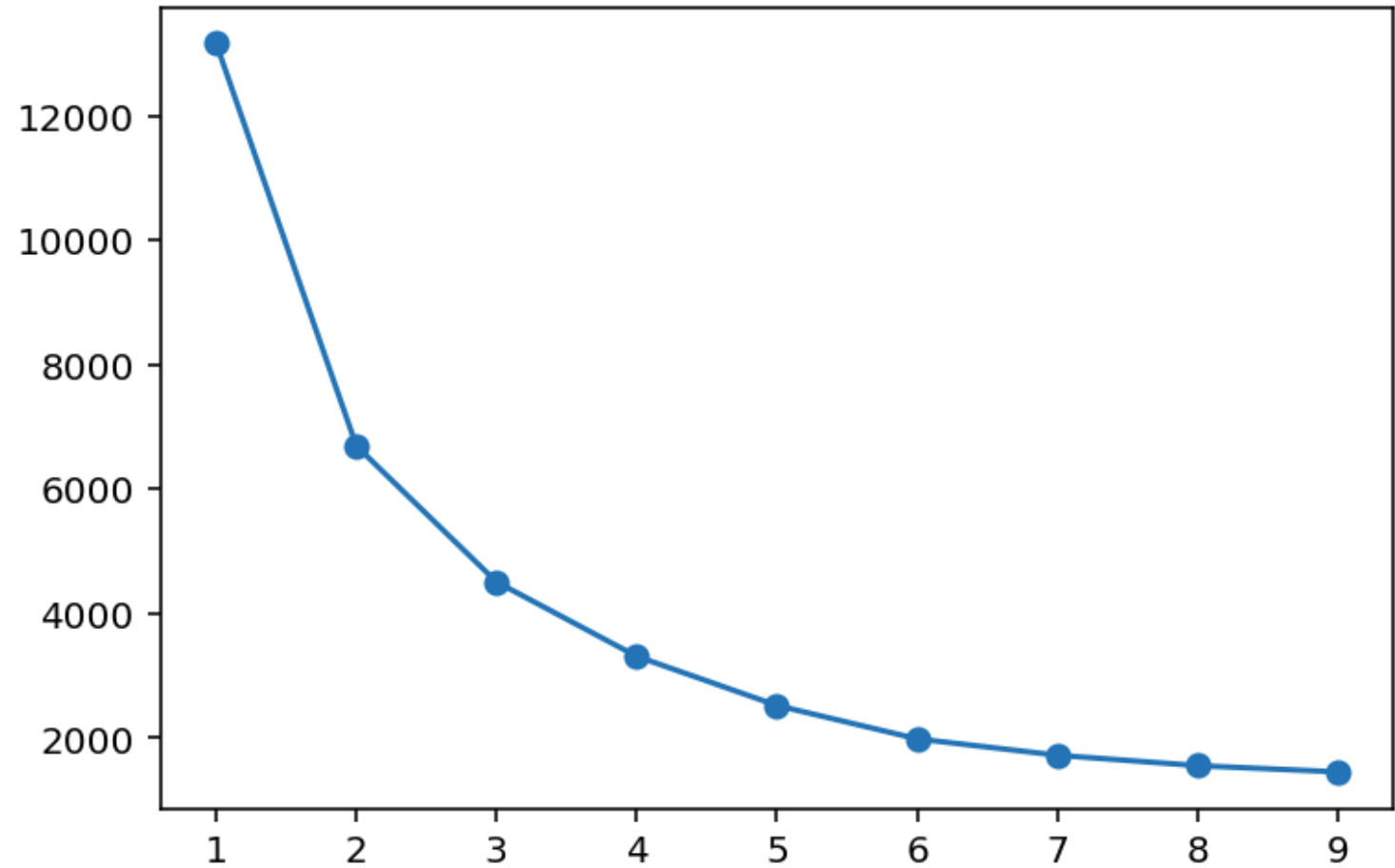
Origin == Cluster	Accuracy
USA == 0 Japan == 1	53.71%
USA == 1 Japan == 0	46.28%



K-means Clustering을 실시해서 얻은 군집과 실제 국가는 거의 일치한다고 볼 수 없다.
고로 게임 제작회사의 국가가 매출에 결과적으로 미치는 영향은 크게 없음을 알 수 있다.

3. 이를 바탕으로 K-means Clustering

Elbow method 결과 6개 이상의 클러스터를 설정할 수 있지만,
분석 목적이 USA, Japan에 해당하는 2개의 클러스터가 필요하므로 무시



Process

Pre-processing

: 게임 제작사 총 576개 확인

: 출시횟수 feature 생성

Normalization

T - test

전체 매출액
Vs.
출시 횟수

Conclusion

Q_5. 게임 출시 횟수와 매출액

Two samples T-test 시행 : 전체 매출액 Vs. 게임 출시 횟수

- groupby를 사용해서 총 576개 게임 제작사의 전체 매출액과 게임 출시 횟수 데이터셋 생성
- scale 조정을 위해 MinMax Normalization 후 T-test 시행

	Publisher	whole_sales	count
0	10TACLE Studios	0.006410	0.001503
1	1C Company	0.005609	0.001503
2	20th Century Fox Video Games	0.090385	0.003005
3	2D Boy	0.007212	0.000000
4	3DO	0.065371	0.026296
...
571	id Software	0.002404	0.000000
572	imageepoch Inc.	0.002404	0.000751
573	inXile Entertainment	0.019231	0.000000
574	mixi, Inc	0.204327	0.000000
575	responDESIGN	0.014423	0.000751

576 rows × 3 columns

- 귀무가설 : 전체 매출액과 출시횟수의 평균값이 같을 것이다.
- 대안가설 : 전체 매출액과 출시횟수의 평균값이 같지 않을 것이다.

T-stat	P-value	상관계수
5.290	0.0	0.223

신뢰수준 95%에서 귀무가설을 기각할 근거가 충분하다.

고로 게임 회사의 전체 매출액과 출시횟수에
연관성이 크게 없을 것이라고 생각할 수 있다.

Q_5. 게임 출시 횟수와 매출액

그렇다면 출시횟수에 관계없이 unit sales가 높은 회사는 어디일까?

1. unit sales (전체 매출액 / 출시 횟수) 가 높은 top30 회사의 데이터셋 생성

(30, 4)

	Publisher	whole_sales	Count	unit_sales(sales/count)
0	Palcom	4.17	1	4.17
1	UEP Systems	2.26	1	2.26
2	Valve	1.73	1	1.73
3	Hello Games	1.60	1	1.60
4	Westwood Studios	1.55	1	1.55

2. 전체 데이터에서 해당 회사가 제작한 게임 총 50개를 찾음 (str.contains 사용)

(50, 11)

	Name	Platform	Genre	Publisher	Year_2	NA_Sales(M)	EU_Sales(M)	JP_Sales(M)	Other_Sales(M)	whole_sales
0	Sesame Street: Elmo's Letter Adventure	N64	Misc	NewKidCo	99	0.14	0.04	0.00	0.00	0.18
1	Teenage Mutant Ninja Turtles	NES	Action	Palcom	89	3.38	0.44	0.31	0.04	4.17
2	The Lord of the Rings: The Fellowship of the Ring	PS2	Action	Black Label Games	02	0.59	0.46	0.00	0.15	1.20
3	Sesame Street: Elmo's Number Journey	PS	Misc	NewKidCo	98	0.13	0.09	0.00	0.02	0.24
4	Zoop	PS	Puzzle	MediaQuest	95	0.03	0.02	0.00	0.00	0.05

Q_5. 게임 출시 횟수와 매출액

그렇다면 출시횟수에 관계없이 unit sales가 높은 회사는 어디일까?

3. unit sales top30 제작사가 만든 게임들만 추출하여 새로운 데이터셋 생성

(39, 12)

	Publisher	Name	Platform	Genre	Year	NA_Sales(M)	EU_Sales(M)	JP_Sales(M)	Other_Sales(M)	sales_total	Count	unit_sales
0	Palcom	Teenage Mutant Ninja Turtles	NES	Action	89	3.38	0.44	0.31	0.04	4.17	1	4.17
1	UEP Systems	Cool Boarders 2	PS	Sports	97	1.52	0.46	0.20	0.08	2.26	1	2.26
2	Valve	Portal 2	PS3	Shooter	11	0.83	0.63	0.02	0.25	1.73	1	1.73
3	Hello Games	No Man's Sky	PS4	Action	16	0.58	0.74	0.02	0.26	1.60	1	1.60
4	Westwood Studios	Command & Conquer: Tiberian Sun	PC	Strategy	99	1.55	0.00	0.00	0.00	1.55	1	1.55

▶ 위 회사들(unit sales top30)을 모티브로 거대 게임 제작사와 같이 많은 게임을 출시하지 않고, 어떻게 효율적으로 높은 매출을 올릴 수 있을 지 다음 분기 전략을 구상해 볼 수 있을 것이다.

감사합니다.