The effect of game genre on country's development

Warat Kosacarn

Abstract

Is there any significant difference between the development of the country where the action and shooting game sales ratio is high and the development of the country where the action and shooting game sales ratio is low? The following datasets were used in this analysis: 1.Game sales 2.World development indicators. The heatmap and line plot were used as the main method to visualize and draw a conclusion about the data. The findings have shown that there is no significant difference in terms of country’s development.

Motivation

In the developed country, everyone has an access to smartphone and computer. Nowadays, games are everywhere. Then, the question arises “How do action and shooting games affect the country development?” “Do some countries tend to have a better development if the people in that country do not play as much action and shooting games?”. The goal of this analysis is to make everyone become aware of the effect of games on society development. This study might be especially beneficial to parents and teachers who are closely related to children. There is a good chance that the games that you let your kids play might affect them in some way. And the children are the future of the nation. In other words, the future of the nation might depend on the kind of games that you let them play.

Datasets

The following datasets were used in this analysis.

1. Game sales: https://www.kaggle.com/gregorut/videogamesales

2. World Development Indicators : https://www.kaggle.com/worldbank/world-development-indicators

The World Development Indicators dataset obtained from the World Bank containing 1344 annual indicators of economic development from 247 countries around the world. It contains about 5.6 million records.

The game sales dataset contains about 11000 records: fields included are Ranking of overall sales, The games name, Platforms of the game release, Year of the game’s release, Genre of the game, Publisher of the game , Sales in North America (in millions), Sales in Europe (in millions), Sales in Japan (in millions), Sales in the rest of the world (in millions), Total worldwide sales

Data Preparation and Cleaning

There are some null values in the game sales dataset. The dropna() method was used to clean the data. As for the world development indicators data, there are the ‘Year’ and ‘Value’ for each indicator. This could be confusing when they are combined into one dataframe. For concreteness, the column ‘Value’ was renamed into something easier to understand. For example, the ‘Value’ column that shows GDP of Japan was renamed as ‘GDP of Japan’ .

Data Preparation and Cleaning

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
12863	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1960	7079.426405
37965	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1961	7727.986365
65586	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1962	8338.393925
94086	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1963	8953.400085
122869	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1964	9895.220852

	Year	GDP of JPN
12863	1960	7079.426405
37965	1961	7727.986365
65586	1962	8338.393925
94086	1963	8953.400085
122869	1964	9895.220852

Research Question

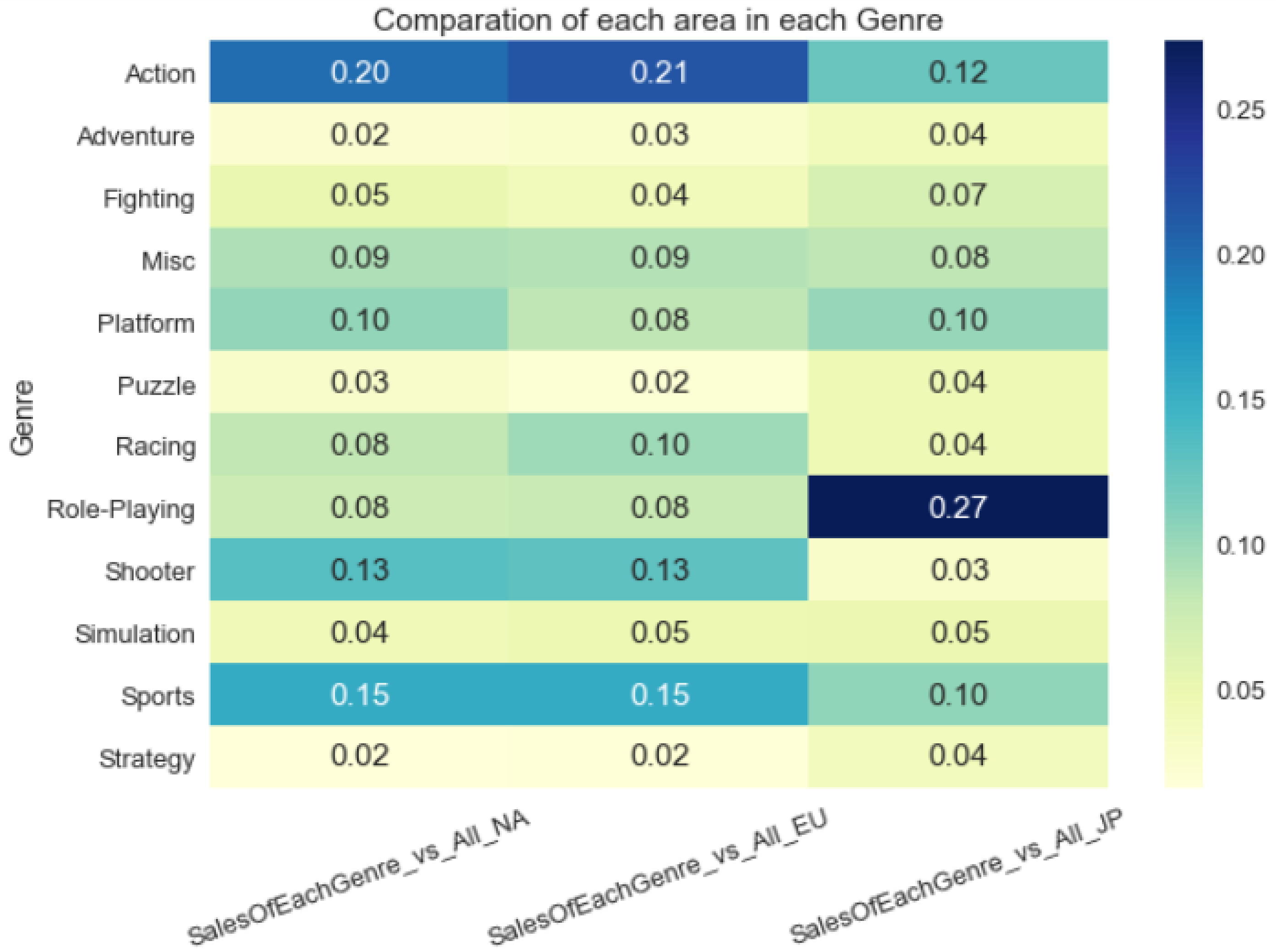
Is there any significant difference between the development of the country where the sales of action and shooting games ratio is high and the development of the country where the sales of action and shooting games ratio is low?

Methods

I mainly used heatmap and line plot for this analysis.

The reasons that these methods were used are : I was looking for the kinds of games that have the high sales ratio in each region. Heatmap plot is the appropriate visualization because it gives a clear picture which genres tend to be popular in a specific region. The line plot is excellent for showing the trend of each country's development indicator. The primary goal of this analysis is to investigate the affect of some of game genres on some aspects of the country's developments.

Findings



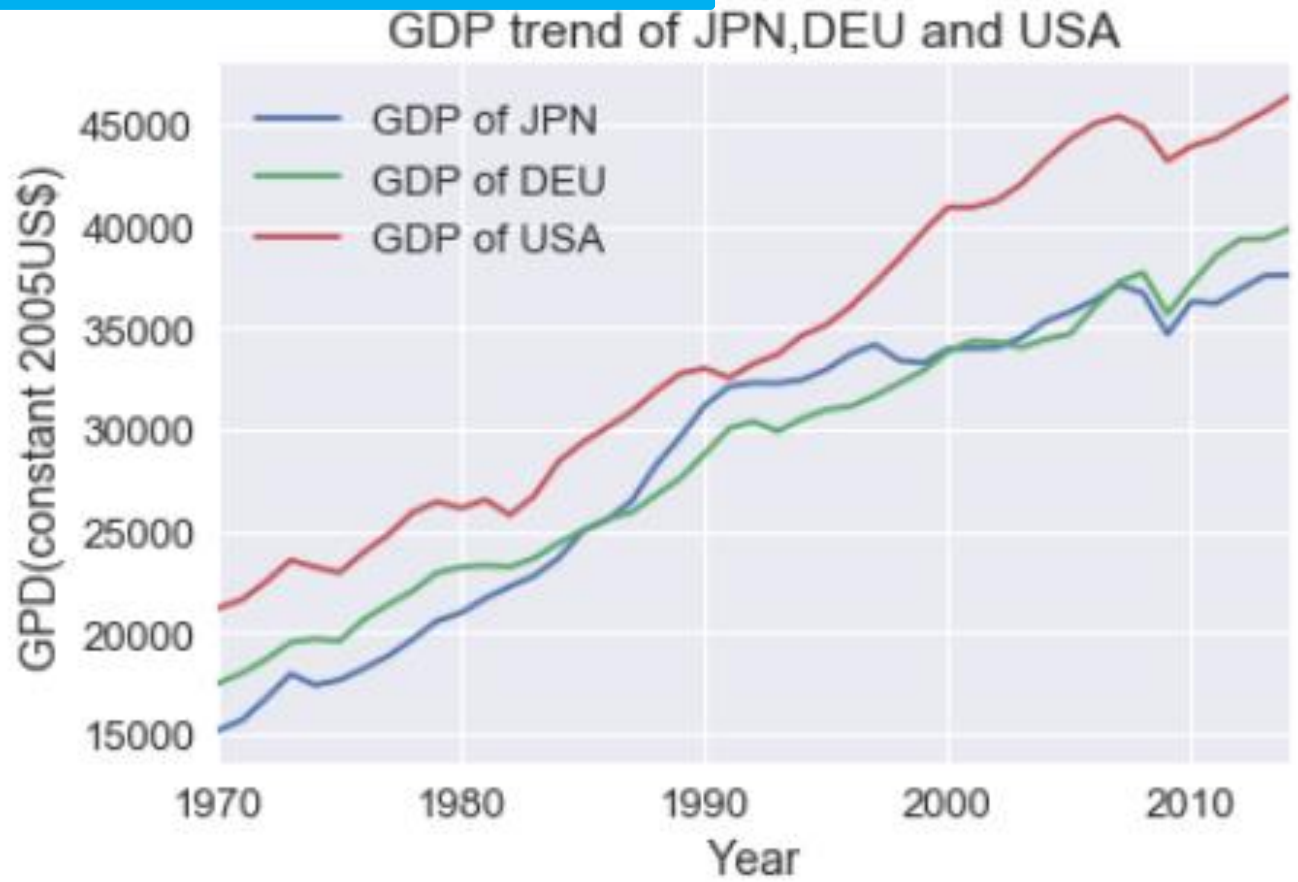
The number in each cell shows the ratio of each genre sale to the sale of all genres in a particular region. It was clearly shown that the action and shooting games are particularly popular in North America (NA) and Europe (EU).

Summary from Heatmap plot

In the previous slide, it was shown that the amount of action and shooting game sales are particularly high in NA and EU but not so high in Japan. The next step is to investigate the trend of development indicators of each representative country from each region. I have selected United States(USA) for NA, Germany (DEU) for EU

For the upcoming slides, the main focus is on the trend not the actual number of each country's indicator.

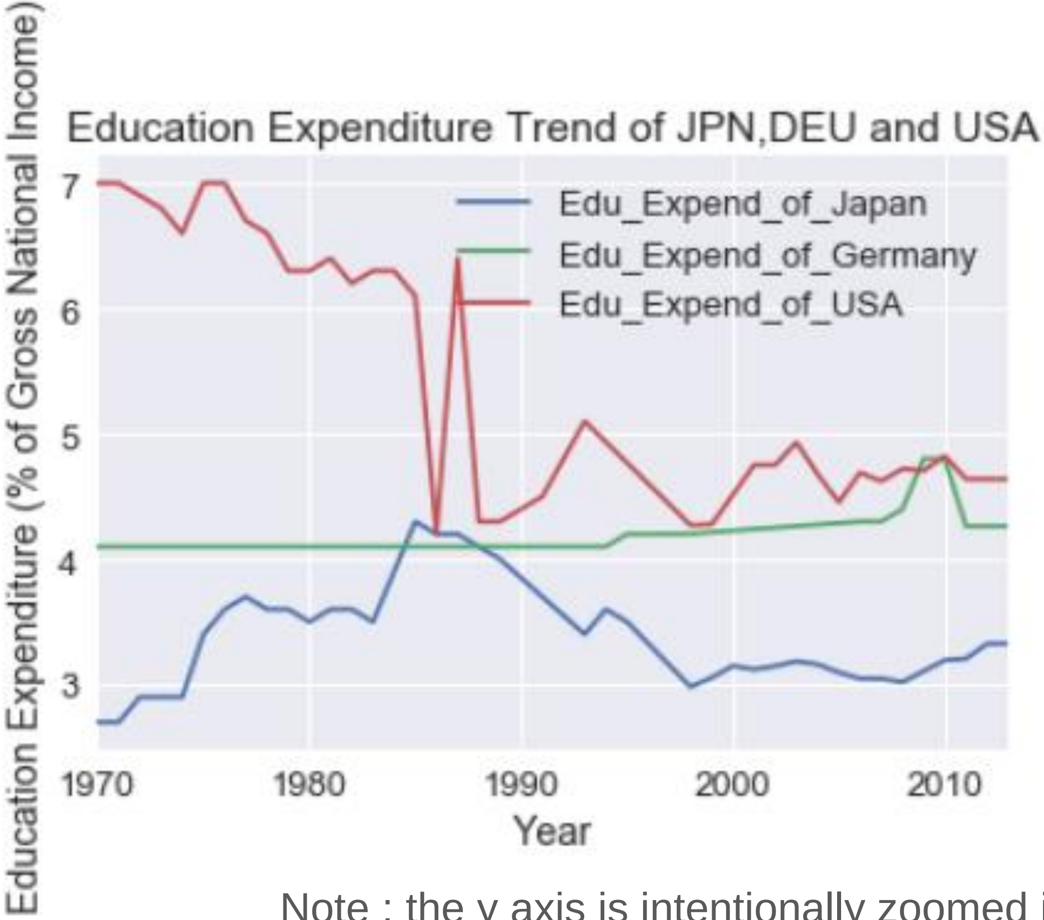
Findings



All of the representative countries have shown the **similar uptrend of GDP**. It is safe to say that action and shooting games have no effect on the country's GDP.

Note : the y axis is intentionally zoomed in so that the difference can be seen clearly.

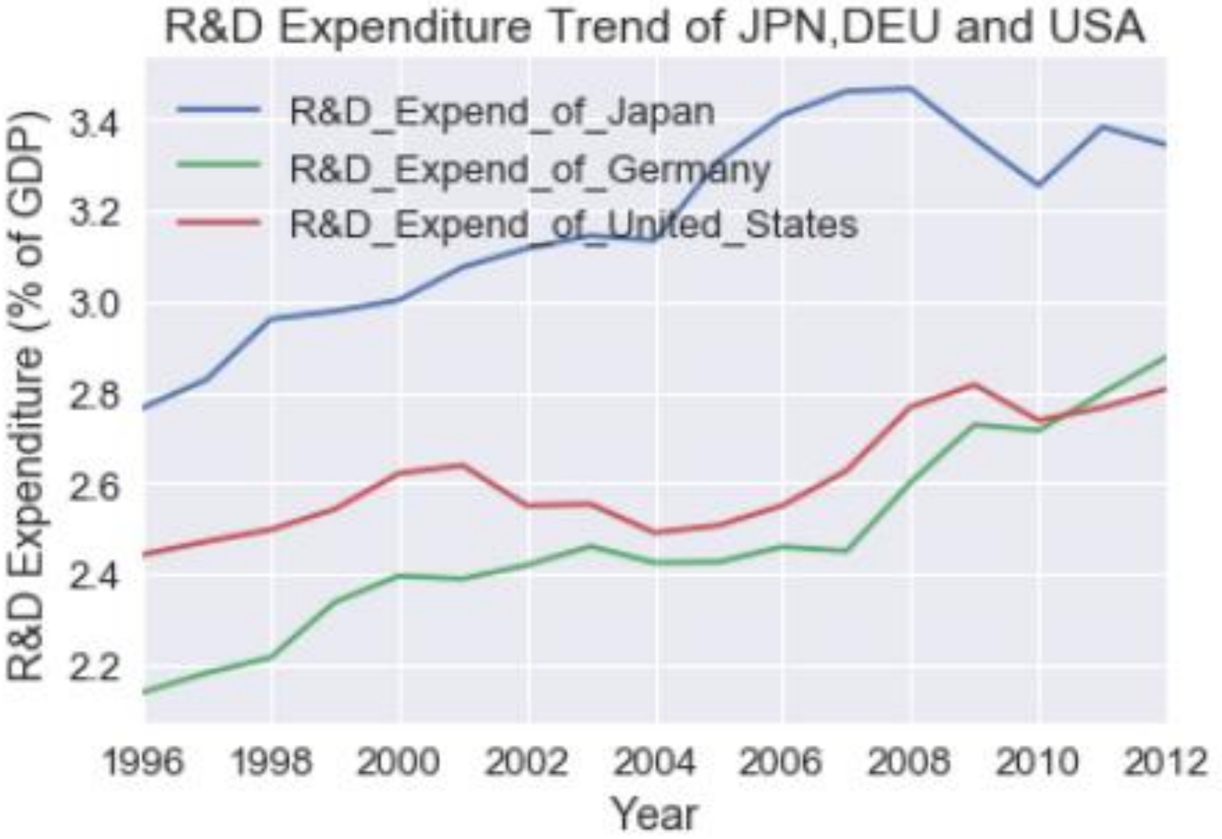
Findings



USA line has shown a critical declination in terms of education expenditure. However, it is not safe to say that it was due to action and shooting games because Germany line did not show any significant drop. Japan line has also shown the rise and drop. Therefore, there is no clear evidence that the action and shooting games negatively affect the country development from the aspect of education expenditure.

Note : the y axis is intentionally zoomed in so that the difference can be seen clearly.

Findings



All of the representative countries have shown the **similar uptrend of R&D expenditure**. It is safe to say that action and shooting games have no effect on the country’s Research and Development Expenditure.

Note : the y axis is intentionally zoomed in so that the difference can be seen clearly.

Limitations

I do not have access to violence and crime dataset which could have been useful in terms of determine whether the action and shooting games have any influence on those area. So I have picked three indicators out of a thousand of them from the world development indicators and visualize the results. Another limitation of this analysis is that I have selected only one country from the North America and Europe region. I did this so that it will be easier to see the effect of game on the country that belongs to a specific region.

Conclusions

In conclusion, it was shown by analyzing three of the country's development indicators that the action and shooting games has no negative effect on country's development. *The GDP and R&D expenditure trends of the country where the action and shooting game sales ratio are high are similar to the trend of the country where the action and shooting game sales ratio are low.* There is also no clear evidence that the action and shooting games negatively affect the country development in the aspect of education expenditure.

Acknowledgements

Thank you my friends for giving a feedback on my work. Thank you Kaggle for great dataset. This analysis would not have been possible without Kaggle.

References

I have used the following websites to guide my analysis, data manipulation and visualization.

https://seaborn.pydata.org/tutorial/distributions.html

https://pandas.pydata.org/pandas-docs/stable/dsintro.html

Jupyter Notebook

The effect of action and shooting games on country development.

I was playing game then this thought came to me "How is the game affecting me?" "Do games have any major impact on the person?". Each country is made up of people. People are the reflection of country. If there is any kind of game that affects an individual negatively. Then, it makes sense that these kinds of games might be affecting the country development as well.

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
# import game sales dataset into pandas DataFrame
gamesales = pd.read_csv('vgsales.csv')
gamesales.head()
```

Out[3]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

In [4]:

```
# Check for null value
gamesales.isnull().any()
```

Out[4]:

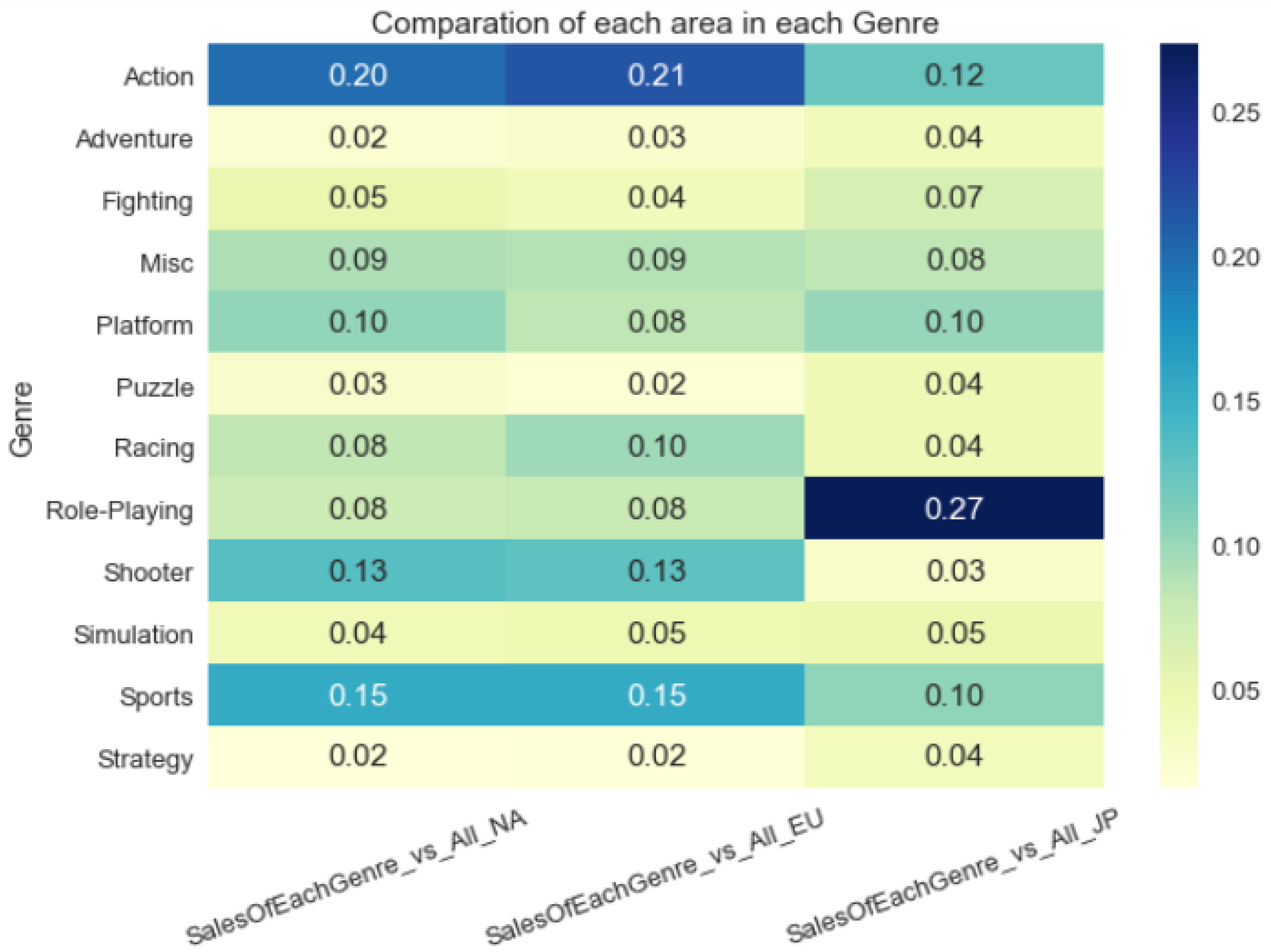
```
Rank            False
Name            False
Platform        False
Year             True
Genre           False
Publisher        True
NA_Sales        False
EU_Sales        False
JP_Sales        False
Other_Sales     False
Global_Sales    False
dtype: bool
```

In [5]:
```
# Get rid of null value
gamesales=gamesales.dropna()
```

In [5]:
```
gamesales.isnull().any()
```

Out[5]:
```
Rank            False
Name            False
Platform        False
Year            False
Genre           False
Publisher       False
NA_Sales        False
EU_Sales        False
JP_Sales        False
Other_Sales     False
Global_Sales    False
dtype: bool
```

In [7]:
```
# Next I want to group up the game genre and sum the sale of three major areas (NA,EU,JP)
GenreGroupVsSale= gamesales.groupby(['Genre']).sum().loc[:,'NA_Sales':'Global_Sales']
# Create new column to show the ratio of sales of each game genre to the total game sales in each region.
GenreGroupVsSale['SalesOfEachGenre_vs_All_NA'] = GenreGroupVsSale['NA_Sales']/GenreGroupVsSale.NA_Sales.sum()
GenreGroupVsSale['SalesOfEachGenre_vs_All_EU'] =GenreGroupVsSale['EU_Sales']/GenreGroupVsSale.EU_Sales.sum()
GenreGroupVsSale['SalesOfEachGenre_vs_All_JP'] = GenreGroupVsSale['JP_Sales']/GenreGroupVsSale.JP_Sales.sum()
# Use the seaborn heatmap plot to visualize the data
plt.figure(figsize=(10, 15))
sns.set(font_scale=1.3)
plt.subplot(211)
sns.heatmap(GenreGroupVsSale.loc[:, 'SalesOfEachGenre_vs_All_NA':'SalesOfEachGenre_vs_All_JP'],cmap="YlGnBu", annot=True,
fmt = '.2f')
plt.title("Comparation of each area in each Genre")
plt.xticks(rotation=20)
plt.show()
```


Comparation of each area in each Genre
Genre
Action
Adventure
Fighting
Misc
Platform
Puzzle
Racing
Role-Playing
Shooter
Simulation
Sports
Strategy
0.20 0.21 0.12
0.02 0.03 0.04
0.05 0.04 0.07
0.09 0.09 0.08
0.10 0.08 0.10
0.03 0.02 0.04
0.08 0.10 0.04
0.08 0.08 0.27
0.13 0.13 0.03
0.04 0.05 0.05
0.15 0.15 0.10
0.02 0.02 0.04
SalesOfEachGenre_vs_All_NA
SalesOfEachGenre_vs_All_EU
SalesOfEachGenre_vs_All_JP
0.25
0.20
0.15
0.10
0.05

In [7]:
```
# import World Development Indicator dataset into pandas DataFrame
worldDevelopmentIndicator = pd.read_csv('Indicators.csv')
```

In [8]:
```
# select GDP for Japan
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('GDP per capita')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('JPN')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('constant 2005 US')

GDP_of_Japan = worldDevelopmentIndicator[mask1 & mask2 & mask3]
GDP_of_Japan.head()
```

Out[8]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
12863	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1960	7079.426405
37965	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1961	7727.986365
65586	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1962	8338.393925
94086	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1963	8953.400085
122869	Japan	JPN	GDP per capita (constant 2005 US$)	NY.GDP.PCAP.KD	1964	9895.220852

In [35]:
```
#slice out only the year and value so that we can use it to create new dataframe later
GDP_of_Japan_year_and_value=GDP_of_Japan.loc[:,'Year':'Value']
GDP_of_Japan_year_and_value=GDP_of_Japan_year_and_value.rename(columns={'Value':'GDP of JPN'})
GDP_of_Japan_year_and_value.head()
```

Out[35]:

	Year	GDP of JPN
12863	1960	7079.426405
37965	1961	7727.986365
65586	1962	8338.393925
94086	1963	8953.400085
122869	1964	9895.220852

In [20]:

```
#Repeat the same procedure for Germany
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('GDP per capita')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('DEU')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('constant 2005 US')

GDP_of_Germany = worldDevelopmentIndicator[mask1 & mask2 & mask3]
```

In [28]:

```
GDP_of_Germany_year_and_value=GDP_of_Germany.loc[:,'Year':'Value']
GDP_of_Germany_year_and_value=GDP_of_Germany_year_and_value.rename(columns={'Value':'GDP of DEU'})
GDP_of_Germany_year_and_value.head()
```

Out[28]:

	Year	GDP of DEU
324685	1970	17473.039645
389408	1971	17987.384810
458378	1972	18671.350796
528175	1973	19501.855695
597968	1974	19667.769739

In [29]:
```
#Repeat the same procedure for United States
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('GDP per capita')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('USA')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('constant 2005 US')

GDP_of_United_States = worldDevelopmentIndicator[mask1 & mask2 & mask3]
```

In [30]:
```
GDP_of_United_States_year_and_value=GDP_of_United_States.loc[:,'Year':'Value']
GDP_of_United_States_year_and_value=GDP_of_United_States_year_and_value.rename(columns={'Value':'GDP of USA'})
GDP_of_United_States_year_and_value.head()
```

Out[30]:

	Year	GDP of USA
22282	1960	15482.707760
48759	1961	15578.409657
77142	1962	16276.426685
105760	1963	16749.789436
134798	1964	17476.822248

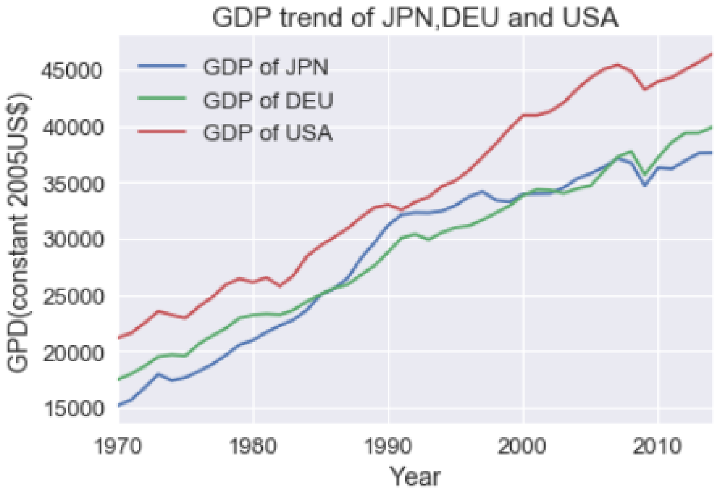
In [31]:
```
a=GDP_of_Japan_year_and_value.merge(GDP_of_Germany_year_and_value, on='Year', how='inner')
GDP_summary=a.merge(GDP_of_United_States_year_and_value,on='Year', how='inner')
```

In [34]:
```
GDP_summary.head()
```

Out[34]:

	Year	GDP of JPN	GDP of DEU	GDP of USA
0	1970	15161.772216	17473.039645	21183.217795
1	1971	15671.171047	17987.384810	21606.394645
2	1972	16753.343720	18671.350796	22501.419194
3	1973	17949.864596	19501.855695	23545.391149
4	1974	17394.688923	19667.769739	23210.587448

In [37]:
```
GDP_summary.plot(x='Year',title='GDP trend of JPN,DEU and USA')
plt.ylabel('GPD(constant 2005US$)')
plt.show()
#note that the y axis is intentionally zoomed in so that the difference can be seen clearly.
```



In [14]:
```python
#education expenditure (% of Gross National Income)
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('education expenditure')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('JPN')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('of GNI')

edu_expend_of_Japan = worldDevelopmentIndicator[mask1 & mask2 & mask3]
edu_expend_of_Japan.head()
```

Out[14]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
331661	Japan	JPN	Adjusted savings: education expenditure (% of ...	NY.ADJ.AEDU.GN.ZS	1970	2.7
397380	Japan	JPN	Adjusted savings: education expenditure (% of ...	NY.ADJ.AEDU.GN.ZS	1971	2.7
466393	Japan	JPN	Adjusted savings: education expenditure (% of ...	NY.ADJ.AEDU.GN.ZS	1972	2.9
536093	Japan	JPN	Adjusted savings: education expenditure (% of ...	NY.ADJ.AEDU.GN.ZS	1973	2.9
606004	Japan	JPN	Adjusted savings: education expenditure (% of ...	NY.ADJ.AEDU.GN.ZS	1974	2.9

In [39]:
```python
edu_expend_of_Japan_year_and_value=edu_expend_of_Japan.loc[:,'Year':'Value']
edu_expend_of_Japan_year_and_value=edu_expend_of_Japan_year_and_value.rename(columns={'Value':'Edu_Expend_of_Japan'})
```

In [40]:
```
#education expenditure (% of Gross National Income)
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('education expenditure')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('DEU')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('of GNI')

edu_expend_of_Germany = worldDevelopmentIndicator[mask1 & mask2 & mask3]
```

In [42]:
```
edu_expend_of_Germany_year_and_value=edu_expend_of_Germany.loc[:,'Year':'Value']
edu_expend_of_Germany_year_and_value=edu_expend_of_Germany_year_and_value.rename(columns={'Value':'Edu_Expend_of_Germany'})
```

In [43]:
```
#education expenditure (% of Gross National Income)
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('education expenditure')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('USA')
mask3 = worldDevelopmentIndicator['IndicatorName'].str.contains('of GNI')

edu_expend_of_United_States = worldDevelopmentIndicator[mask1 & mask2 & mask3]
```

In [45]:
```
edu_expend_of_United_States_year_and_value=edu_expend_of_United_States.loc[:,'Year':'Value']
edu_expend_of_United_States_year_and_value=edu_expend_of_United_States_year_and_value.rename(columns=
{'Value':'Edu_Expend_of_USA'})
```

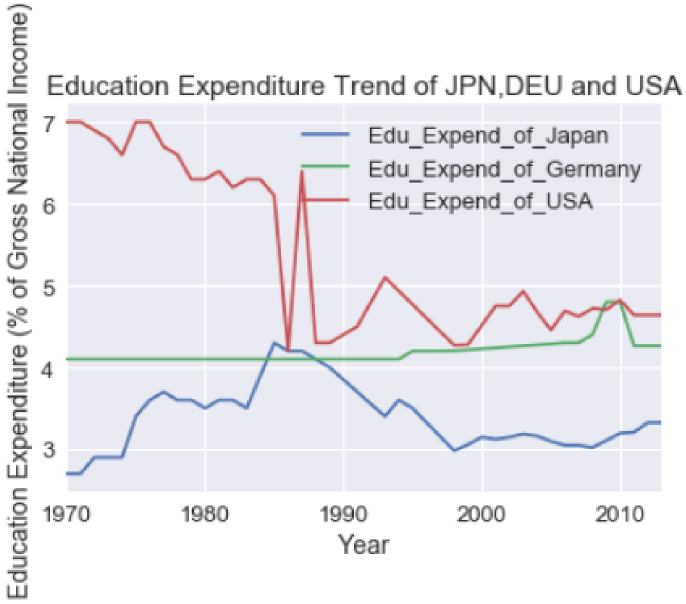
In [46]:
```
b=edu_expend_of_Japan_year_and_value.merge(edu_expend_of_Germany_year_and_value, on='Year', how='inner')
Edu_expend_summary=b.merge(edu_expend_of_United_States_year_and_value,on='Year', how='inner')
Edu_expend_summary.head()
```

Out[46]:

	Year	Edu_Expend_of_Japan	Edu_Expend_of_Germany	Edu_Expend_of_USA
0	1970	2.7	4.1	7.0
1	1971	2.7	4.1	7.0
2	1972	2.9	4.1	6.9
3	1973	2.9	4.1	6.8
4	1974	2.9	4.1	6.6

In [47]:

```
Edu_expend_summary.plot(x='Year',title='Education Expenditure Trend of JPN,DEU and USA')
plt.ylabel('Education Expenditure (% of Gross National Income)')
plt.show()
#note that the y axis is intentionally zoomed in so that the difference can be seen clearly.
```



In [49]:

```
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('development')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('JPN')

R_and_D_expend_of_Japan = worldDevelopmentIndicator[mask1 & mask2]
R_and_D_expend_of_Japan.head()
```

Out[49]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
2691170	Japan	JPN	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS	1996	2.76501
2826528	Japan	JPN	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS	1997	2.82761
2963410	Japan	JPN	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS	1998	2.96020
3103053	Japan	JPN	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS	1999	2.97734
3251921	Japan	JPN	Research and development expenditure (% of GDP)	GB.XPD.RSDV.GD.ZS	2000	3.00169

In [51]:

```
R_and_D_expend_of_Japan_year_and_value=R_and_D_expend_of_Japan.loc[:,'Year':'Value']
R_and_D_expend_of_Japan_year_and_value=R_and_D_expend_of_Japan_year_and_value.rename(columns={'Value':'R&D_Expend_of_Japan'})
```

In [52]:

```
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('development')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('DEU')

R_and_D_expend_of_Germany = worldDevelopmentIndicator[mask1 & mask2]
```

In [54]:

```
R_and_D_expend_of_Germany_year_and_value=R_and_D_expend_of_Germany.loc[:,'Year':'Value']
R_and_D_expend_of_Germany_year_and_value=R_and_D_expend_of_Germany_year_and_value.rename(columns=
{'Value':'R&D_Expend_of_Germany'})
```

In [55]:

```
mask1 = worldDevelopmentIndicator['IndicatorName'].str.contains('development')
mask2 = worldDevelopmentIndicator['CountryCode'].str.contains('USA')

R_and_D_expend_of_United_States = worldDevelopmentIndicator[mask1 & mask2]
```

In [57]:

```
R_and_D_expend_of_United_States_year_and_value=R_and_D_expend_of_United_States.loc[:,'Year':'Value']
R_and_D_expend_of_United_States_year_and_value=R_and_D_expend_of_United_States_year_and_value.rename(columns=
{'Value':'R&D_Expend_of_United_States'})
```

In [58]:
```
c=R_and_D_expend_of_Japan_year_and_value.merge(R_and_D_expend_of_Germany_year_and_value, on='Year', how='inner')
R_and_D_expend_summary=c.merge(R_and_D_expend_of_United_States_year_and_value,on='Year', how='inner')
R_and_D_expend_summary.head()
```

Out[58]:

	Year	R&D_Expend_of_Japan	R&D_Expend_of_Germany	R&D_Expend_of_United_States
0	1996	2.76501	2.13894	2.44182
1	1997	2.82761	2.18150	2.47091
2	1998	2.96020	2.21557	2.49675
3	1999	2.97734	2.33730	2.54174
4	2000	3.00169	2.39503	2.62050

In [60]:
```
R_and_D_expend_summary.plot(x='Year',title='R&D Expenditure Trend of JPN,DEU and USA')
plt.ylabel('R&D Expenditure (% of GDP)')
plt.show()
#note that the y axis is intentionally zoomed in so that the difference can be seen clearly.
```