Video Game Console Processing and Sales Comparisons

By: Jorin Kramer, Noah Abrigo, Nicholas Taubert
CSC302 Group 5

Raw Data - Graphics Speed (GFLOPS)

System	GFlops	Year	Handheld	Company
Xbox Series X	12000	2020	FALSE	Microsoft
Playstation 5	9200	2020	FALSE	Sony
Xbox One X	6000	2017	FALSE	Microsoft
Playstation 4 Pro	4200	2016	FALSE	Sony
Xbox Series S	4000	2020	FALSE	Microsoft
PlayStation 4	1843	2013	FALSE	Sony
Xbox One S	1400	2016	FALSE	Microsoft
Xbox One	1310	2013	FALSE	Microsoft
Nintendo Switch (Do	500	2017	FALSE	Nintendo
Nintendo Switch	393.2	2017	TRUE	Nintendo
Xbox 360	240	2005	FALSE	Microsoft
PlayStation 3	230.4	2006	FALSE	Sony
Nintendo Wii U	176	2012	FALSE	Nintendo
Xbox	20	2001	FALSE	Microsoft
Nintendo Wii	12	2006	FALSE	Nintendo
Nintendo Gamecube	9.4	2001	FALSE	Nintendo
PlayStation 2	6.2	2000	FALSE	Sony
Nintendo 3DS	4.8	2011	TRUE	Nintendo
Dreamcast	1.4	1998	FALSE	SEGA
Nintendo DS	0.6	2004	TRUE	Nintendo
Nintendo 64	0.1	1996	FALSE	Nintendo

Raw Data - Games Sales

|	Unnamed: 0	img	title	console	genre	publisher	developer	vg_score	critic_score	user_score	total_shipped	total_sales	na_sales	jp_sales	pal_sales	other_sales	release_date	last_update
0	3	/games/boxart/full_3351915AmericaFrontccc.jpg	Warriors	Series	Action	Unknown	Omega Force	NaN	NaN	NaN	42.06	NaN	NaN	NaN	NaN	NaN	NaN	2020-03-24
1	4	/games/boxart/full_5741036AmericaFrontccc.jpg	God of War	Series	Action	Sony Computer Entertainment	SIE Santa Monica Studio|Ready at Dawn|Javagrou...	NaN	NaN	NaN	35.05	NaN	NaN	NaN	NaN	NaN	2005-03-22	2020-03-04
2	5	/games/boxart/full_6662824AmericaFrontccc.png	Devil May Cry	Series	Action	Capcom	Capcom|Ninja Theory	NaN	NaN	NaN	22.00	NaN	NaN	NaN	NaN	NaN	2001-10-16	2020-02-03
3	6	/games/boxart/full_6510540AmericaFrontccc.jpg	Grand Theft Auto V	PS3	Action	Rockstar Games	Rockstar North	NaN	9.4	NaN	NaN	20.32	6.37	0.99	9.85	3.12	2013-09-17	NaN
4	7	/games/boxart/full_6800951AmericaFrontccc.jpg	Frogger	Series	Action	Konami	Konami|Parker Brothers|SCE Cambridge Studio|Bl...	NaN	NaN	NaN	20.00	NaN	NaN	NaN	NaN	NaN	1981-10-23	2020-02-26

Processing Power
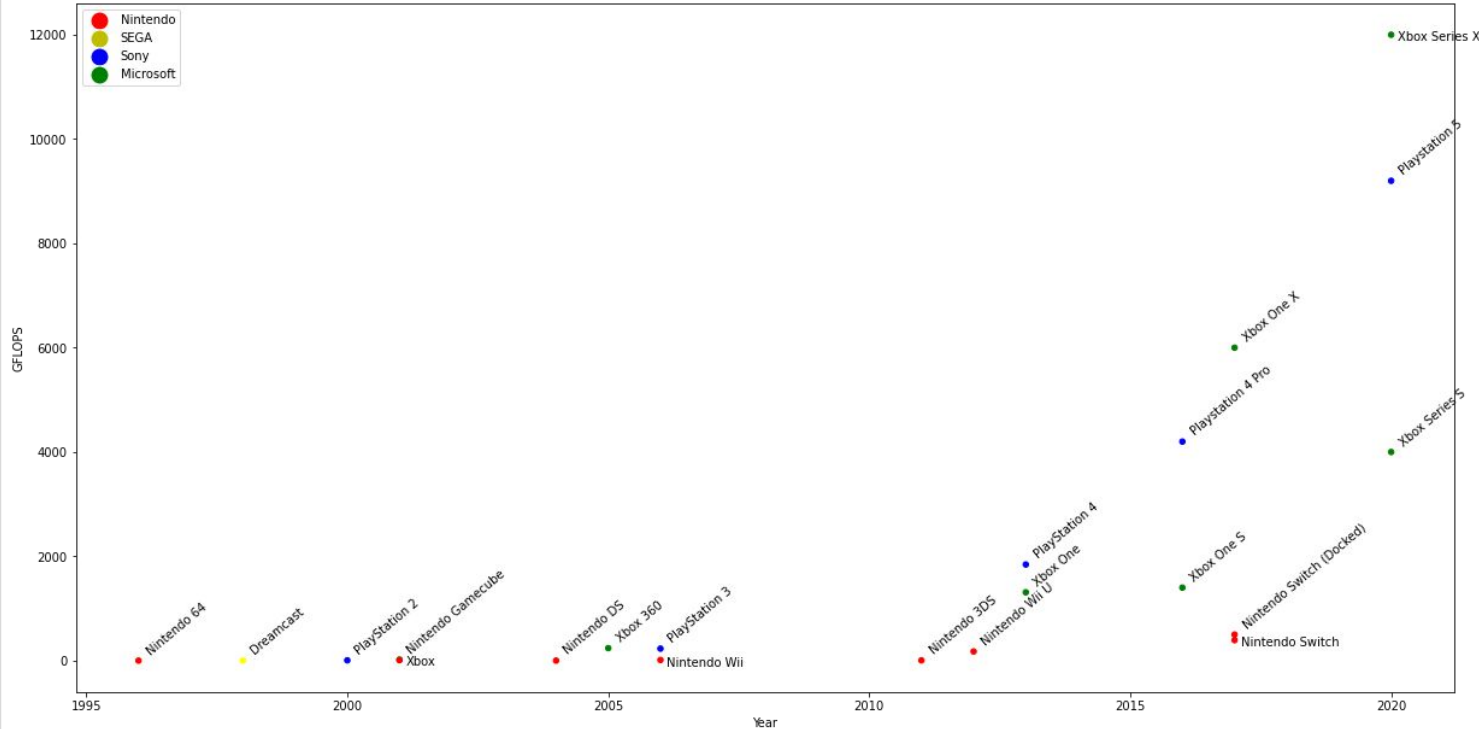
Reason: To visually compare processing power between consoles, keeping age and manufacturer in mind.

Inferences: Nintendo does not improve as much as others

Processing power increase is exponential over time

Consoles tend to release in clusters, or "generations"

Processing Power

Moore's Law (double every 2 years = *sqrt(2) every year
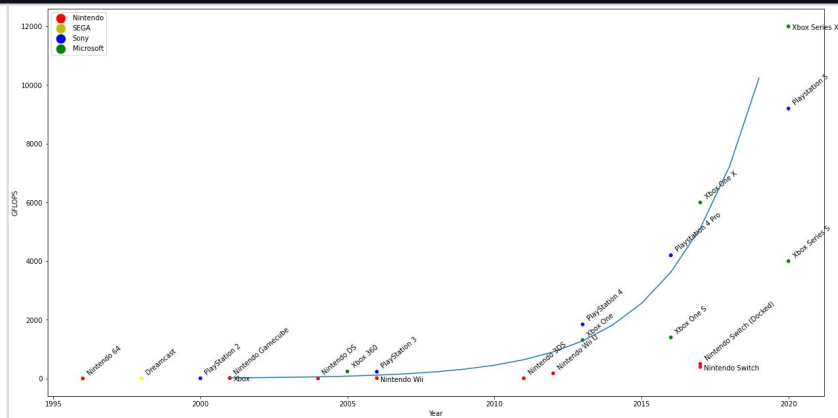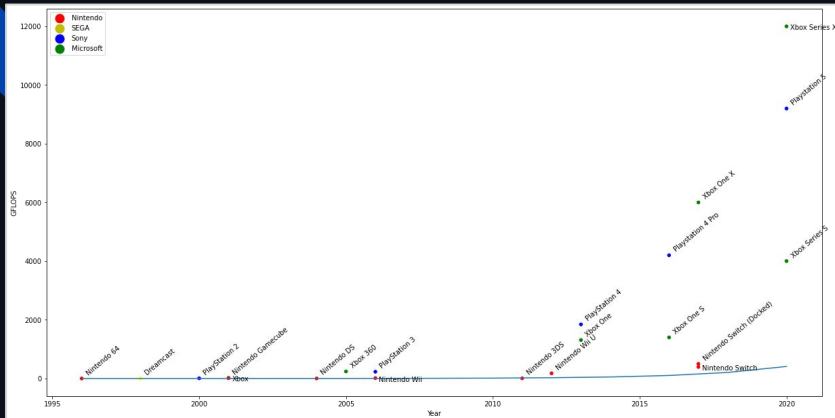
```
x = vals['Year1996']
y = vals['GFLOPS_n64']

plt.plot(x, y)
```

```
x = vals['Year2001']
y = vals['GFLOPS_xbox']

plt.plot(x, y)
```

Starting with N64

Starting with Xbox

Year1996	GFLOPS_n64
1996	0.1
1997	0.1414213562
1998	0.2
1999	0.2828427125
2000	0.4
2001	0.5656854249
2002	0.8
2003	1.13137085
2004	1.6
2005	2.2627417
2006	3.2
2007	4.5254834
2008	6.4
2009	9.050966799
2010	12.8
2011	18.1019336
2012	25.6
2013	36.2038672
2014	51.2
2015	72.40773439
2016	102.4
2017	144.8154688
2018	204.8
2019	289.6309376
2020	409.6

Year2001	GFLOPS_xbox
2001	20
2002	28.28427125
2003	40
2004	56.56854249
2005	80
2006	113.137085
2007	160
2008	226.27417
2009	320
2010	452.54834
2011	640
2012	905.0966799
2013	1280
2014	1810.19336
2015	2560
2016	3620.38672
2017	5120
2018	7240.773439
2019	10240

Processing Power (Code)

```python
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.lines import Line2D
speed = pd.read_csv('/content/drive/Shareddrives/302 Project/Data/speed.csv')

fig, ax = plt.subplots()

colorkey = {'Nintendo':'red', 'Microsoft':'green', 'Sony':'blue', 'SEGA':'yellow'}
speed.plot('Year','GFlops', kind='scatter', c=speed['Company'].map(colorkey))
ax = plt.gca()
fig = plt.gcf()
for i, txt in enumerate(speed['System']):
    if txt == 'Xbox' or txt == 'Nintendo Wii' or txt == 'Nintendo Switch' or txt == 'Xbox Series X':
      rvalue = 0
      yvalue = -5
    else:
      rvalue = 40
      yvalue = 5
    plt.annotate(txt, (speed['Year'][i], speed['GFlops'][i]), xytext=(5, yvalue), ha='left', rotation=rvalue, textcoords='offset points')

ax.set_xlabel("Year")
ax.set_ylabel("GFLOPS")

legend_elements = [Line2D([0], [0], marker='o', color='w', label='Nintendo', markerfacecolor='r', markersize=15),
                   Line2D([0], [0], marker='o', color='w', label='SEGA', markerfacecolor='y', markersize=15),
                   Line2D([0], [0], marker='o', color='w', label='Sony', markerfacecolor='b', markersize=15),
                   Line2D([0], [0], marker='o', color='w', label='Microsoft', markerfacecolor='g', markersize=15)]
ax.legend(handles=legend_elements, loc='upper left')
fig.set_size_inches(20.5, 10.5)
```
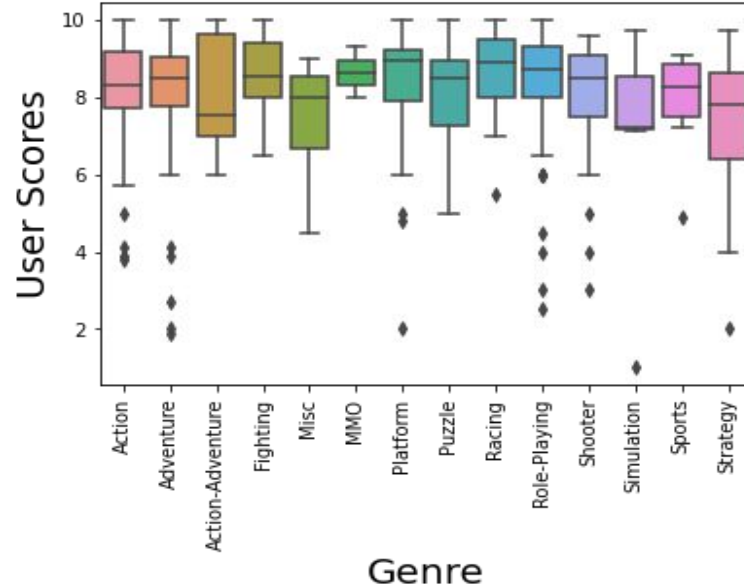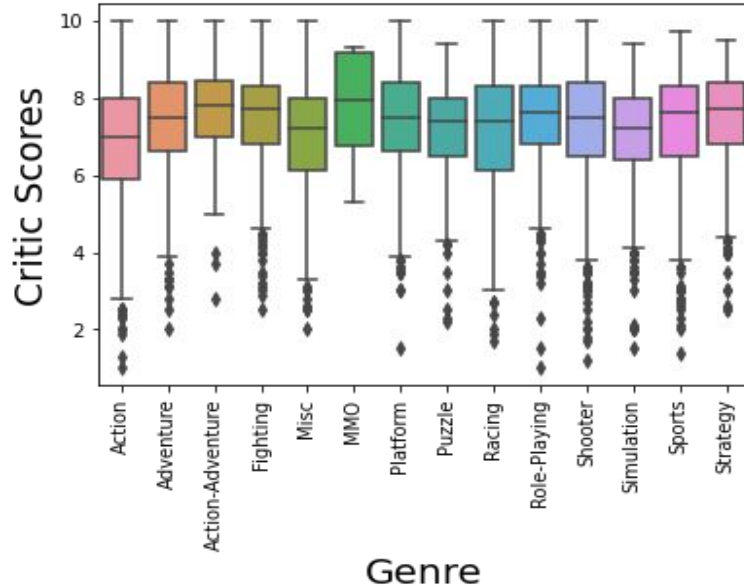
Score by Genre

Sales by Genre

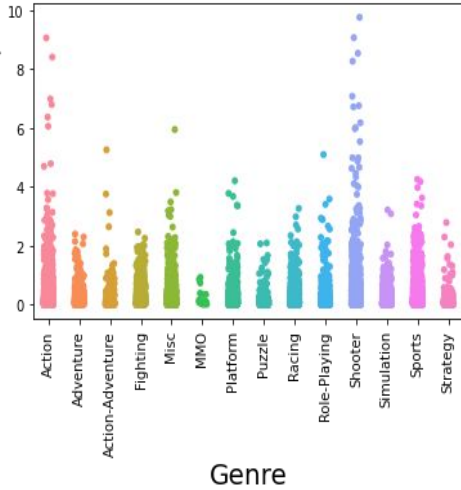**Reason**: Consider the trends of sales of games based on genre and world region

Inferences: Region has huge impact on game sales - different genres perform well in different parts of the world, and game companies should advertise accordingly

Most best selling games come from the same genres

Japan lower than rest (notice y axis)

Total game sales for each console

Reason: To show what consoles have had the most total sales of all time

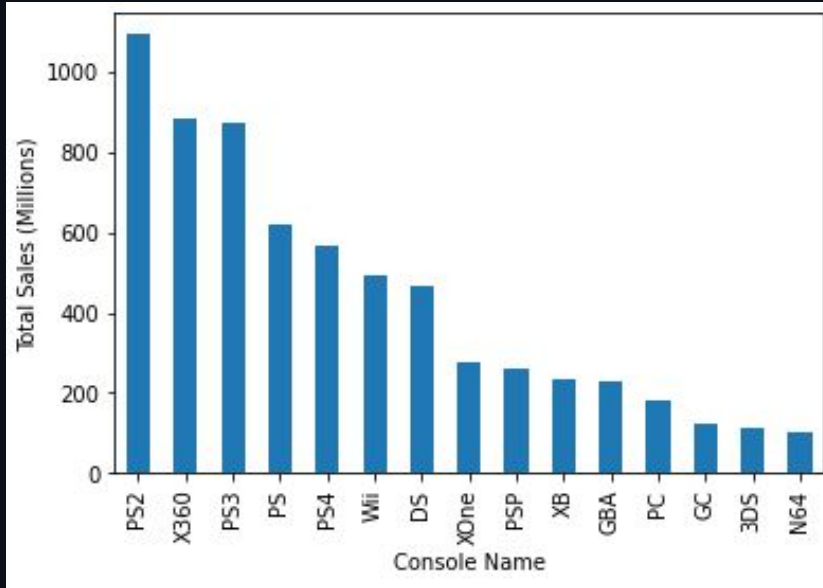**Inference**: This chart seems to show that the total sales in video game consoles seem to have decreased because previous generations are still beating the newer generations


```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('/content/drive/Shareddrives/302 Project/Data/GameSales.csv')
df = df[df['total_sales'].notna()]
df = df.groupby(by="console")["total_sales"].sum().reset_index()
df = df.sort_values(by="total_sales", ascending=False)

plot = df.head(15).plot.bar(x="console", y="total_sales", legend=False)
plt.xlabel("Console Name")
plt.ylabel("Total Sales (Millions)")
```
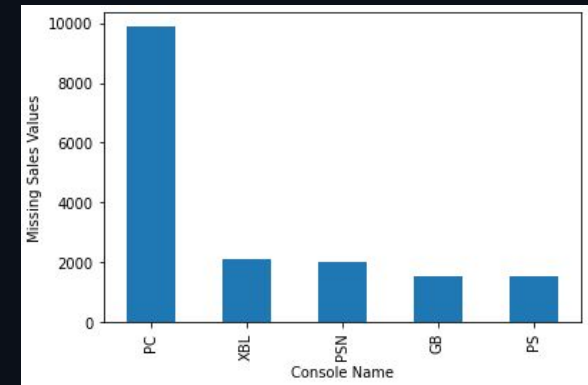
Critic vs. User Scores

Reason: To show the differences in rating by both the players of the games and the reviewers for each game

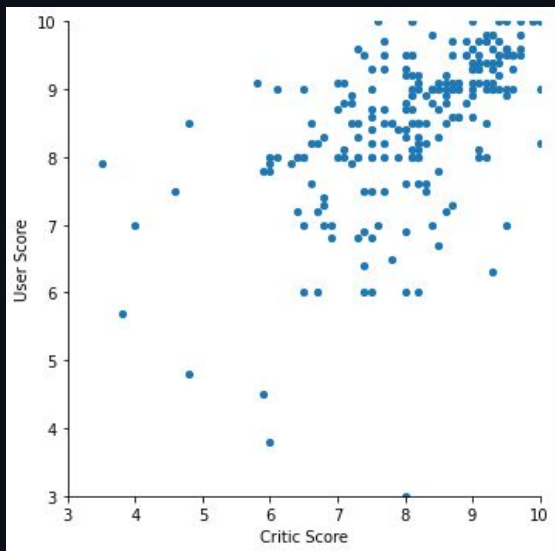**Inference**: Generally all critics will give a lower score than users

```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('/content/drive/Shareddrives/302 Project/Data/GameSales.csv')
df = df.dropna(subset=['critic_score', 'user_score'])
df = df.sort_values(by="total_sales", ascending=False)
```
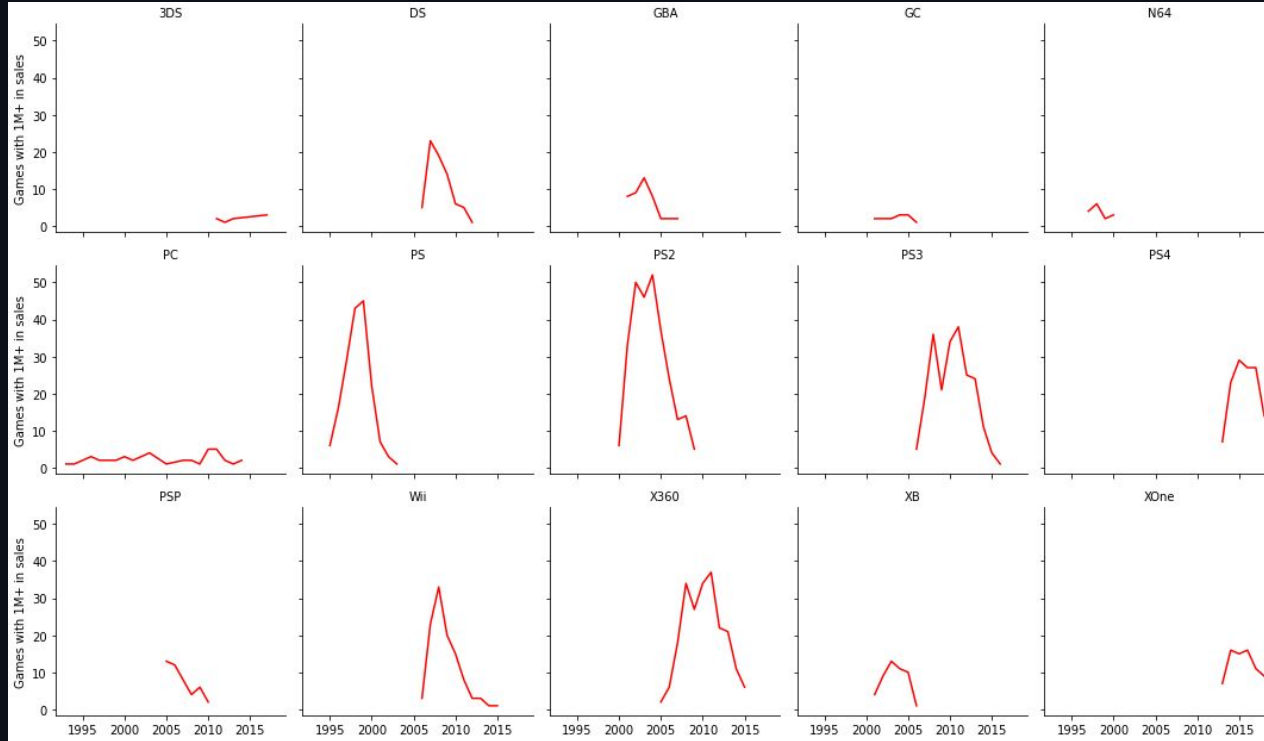
```python
g = sns.FacetGrid(data=df, size=5)
g.map(plt.scatter, 'critic_score', 'user_score', s=18)

plt.xlim(3, 10)
plt.ylim(3, 10)

plt.xlabel("Critic Score")
plt.ylabel("User Score")
```
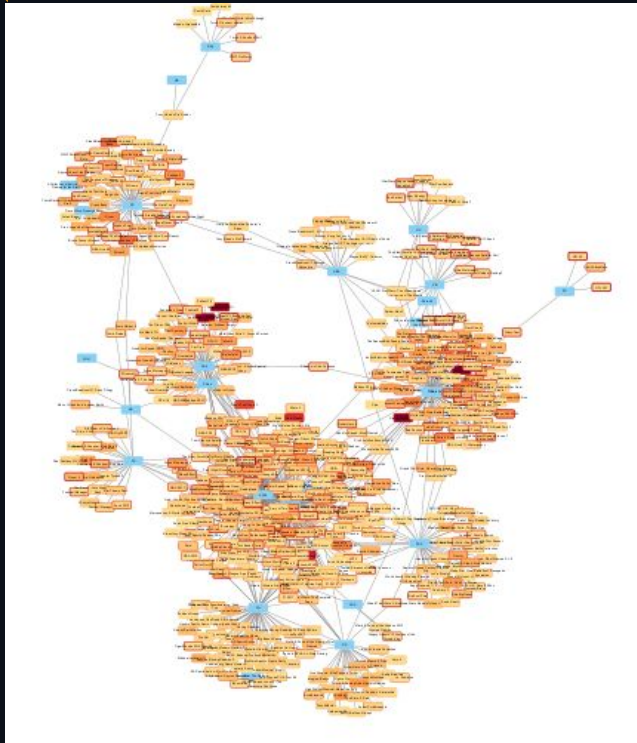
Amount of Games with 1M+ in sales per console

Reason: To show the lifespans of each of the most popular consoles of each generation

Inference: Each console peaked around the middle of their lifespan and as the new generation starts it dropped significantly

Best selling games for each console

Source: Each Console (top 15)
Fill Color: Total Sales
Border: Critic Score
Red is higher
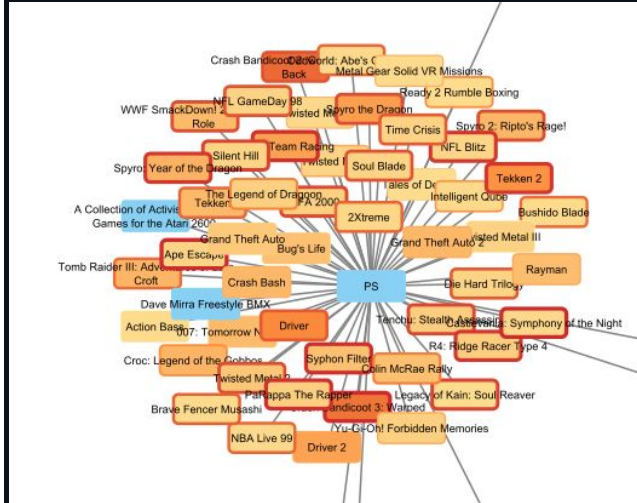
Reason: To show how many popular games each console had

Inference: Last generation PS and XB have the most games as gaming is a large and rising industry and there are still games being made for current gen consoles

Games released per console

The amount of games released on the platform can teach us about how popular the console was during its era. Popularity can be an influence in sales.
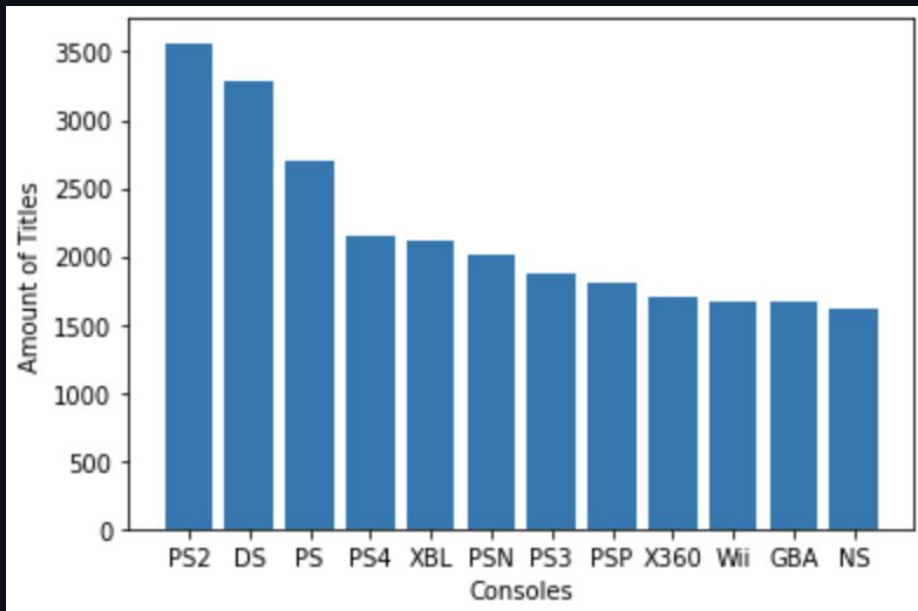
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


df = pd.read_csv("/content/drive/Shareddrives/302 Project/Data/GameSales.csv")
df = df.groupby('console')['title'].count()
df = df.sort_values(ascending=False)
df = df.reset_index()
df = df[1:13]

ax =  plt.bar(df['console'],df['title'])
ax = plt.gca()
fig = plt.gcf()
print(ax.get_yticks())
ax.set_ylabel("Amount of Titles")
ax.set_xlabel("Consoles")
```

Minecraft Sales

Minecraft Sales (Millions)
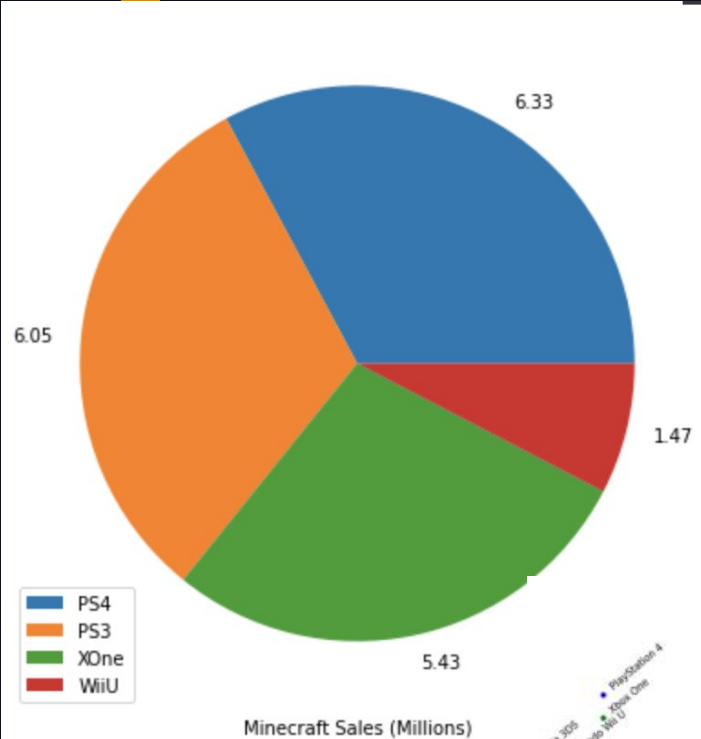
Does better graphics processing lead to more sales?

The PS4 has the greatest processing power among these four consoles showing the sales of Minecraft and also has the most copies sold. The Xbox One lags behind by about a million units.

```
df = pd.read_csv("/content/drive/Shareddrives/302 Project/Data/GameSales.csv")
df2 = df.loc[df['title'] == 'Minecraft']
df2 = df2.loc[(df['console'] == 'XOne') | (df['console'] == 'PS3') | (df['console'] == 'PS4') | (df['console'] == 'WiiU')]
```
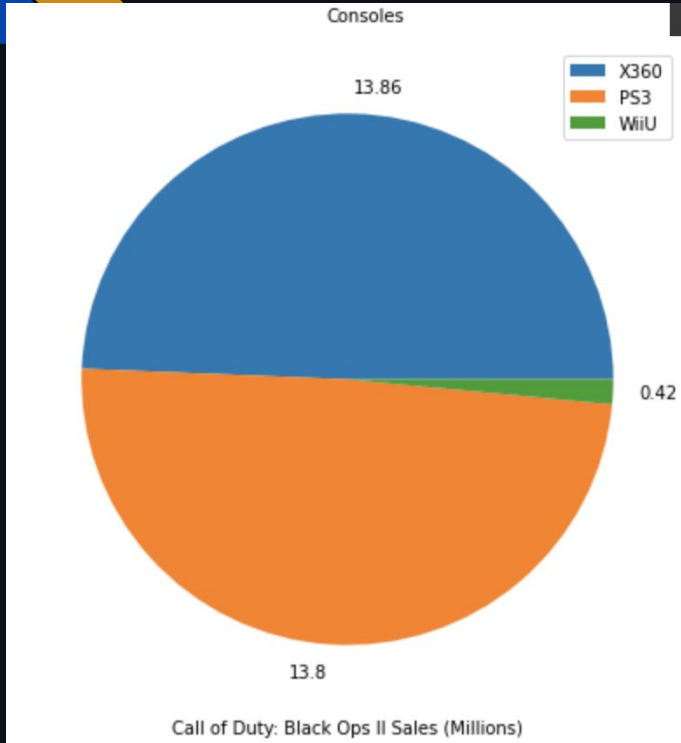
```
plot = df2.plot.pie(y='total_sales', figsize=(7, 7), legend=True, labels=df2['total_sales'])
plot.legend(df2['console'])
plot.set_xlabel("Minecraft Sales (Millions)")
plot.set_ylabel("")
```

Black Ops II sales

Does better graphics processing lead to more sales?

The sale of Call of Duty: Black Ops II is nearly the same between the Xbox 360 and the PS3. These two consoles had nearly the same graphics processing power and they sold nearly the same amount of units of the game.

There may be a correlation between graphics processing power and game sales. Not enough yet to tell.

Consoles
- X360
- PS3
- WiiU

13.86

0.42

13.8

Call of Duty: Black Ops II Sales (Millions)

```python
df = pd.read_csv("/content/drive/Shareddrives/302 Project/Data/GameSales.csv")
df2 = df.loc[df['title'] == 'Call of Duty: Black Ops II']
df2
df2 = df2.loc[(df['console'] == 'X360') | (df['console'] == 'XOne') | (df['console'] == 'PS3') | (df['console'] == 'PS4') | (df['console'] == 'WiiU')]
```

```python
plot = df2.plot.pie(y='total_sales', figsize=(7, 7), legend=True, labels=df2['total_sales'])
plot.legend(df2['console'])
plot.set_xlabel("Call of Duty: Black Ops II Sales (Millions)")
plot.set_ylabel("")
```

Correlation Between Graphics (GFLOPS) and Game Sales Per Console?

Correlation Between Games Sales and Processing Power
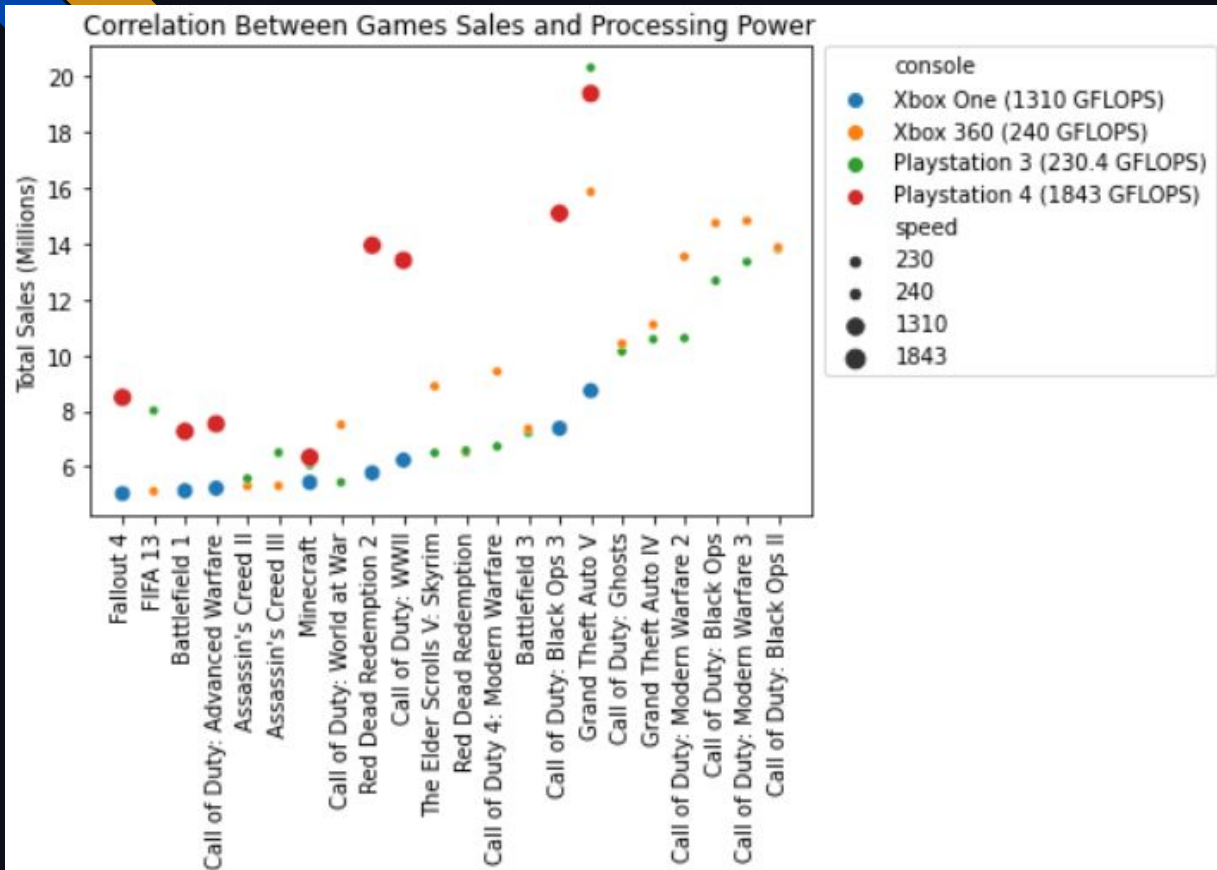
- Games sold for each console
  - Playstation 4 (1843 GFLOP)
  - Xbox One (1310 GFLOPS)
  - Xbox 360 (240 GFLOPS)
  - Playstation (230 GFLOPS)
- Fits the assumption pretty well.
  - Higher GFLOPS consoles have better sales
  - (Respective to their Generation as well)
- External Factors at play
  - Popularity, Console Sales, Game release dates etc.

Some factors may be inconclusive because GTA V which was released on Playstation first, would have more sales etc.
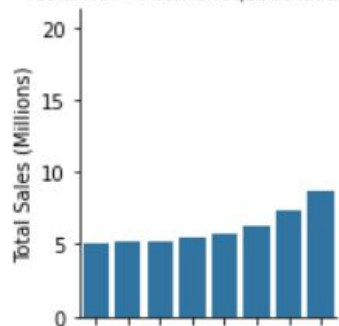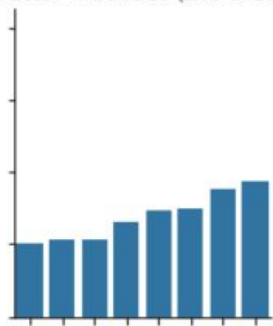
Dataframe Creation

```python
games = pd.read_csv("/content/drive/Shareddrives/302 Project/Data/GameSales.csv")
games = games.dropna(subset=["total_sales","title"]).reset_index()
games = games[games["total_sales"] > 5]
games = games.sort_values(["total_sales"]).reset_index()
size = games.groupby("title").size().reset_index()
games = games.reset_index(drop=True)
for i in range(size["title"].count()):
  if(size[0][i] == 1):
    games = games[games["title"] != size["title"][i]]
games = games.reset_index(drop=True)
games["speed"] = 0
for i in range(games["title"].count()):
    console = games["console"][i]
    if(console == "XOne"):
      games["console"][i] = "Xbox One (1310 GFLOPS)"
      games["speed"][i] = 1310
    if(console == "X360"):
      games["console"][i] = "Xbox 360 (240 GFLOPS)"
      games["speed"][i] = 240
    if(console == "PS3"):
      games["console"][i] = "Playstation 3 (230.4 GFLOPS)"
      games["speed"][i] = 230.4
    if(console == "PS4"):
      games["console"][i] = "Playstation 4 (1843 GFLOPS)"
      games["speed"][i] = 1843
```
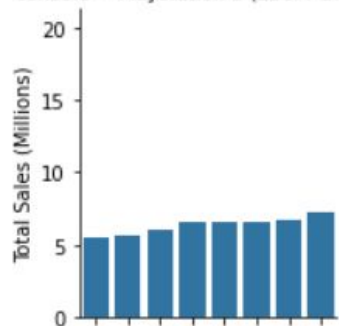
Correlation Between Graphics (GFLOPS) and Game Sales Per Console?

- FacetGrid view of the newly generated dataframe
- Sorted by total sales
- Took titles only released on all 4 consoles
  - This generation and time period was perfect for this examination
  - Looking at two generations of consoles at once.

```python
g = sns.FacetGrid(games, col="console",col_wrap=2)
g.map_dataframe(sns.barplot, x="title", y="total_sales")
plt.xticks(rotation=90)
g.set_xticklabels(rotation=90)
g.set_xlabels("")
g.set_ylabels("Total Sales (Millions)")
g.fig.subplots_adjust(top=0.9)
g.fig.suptitle('Correlation Between Game Sales and Processing Power')
```

Not Accurate?
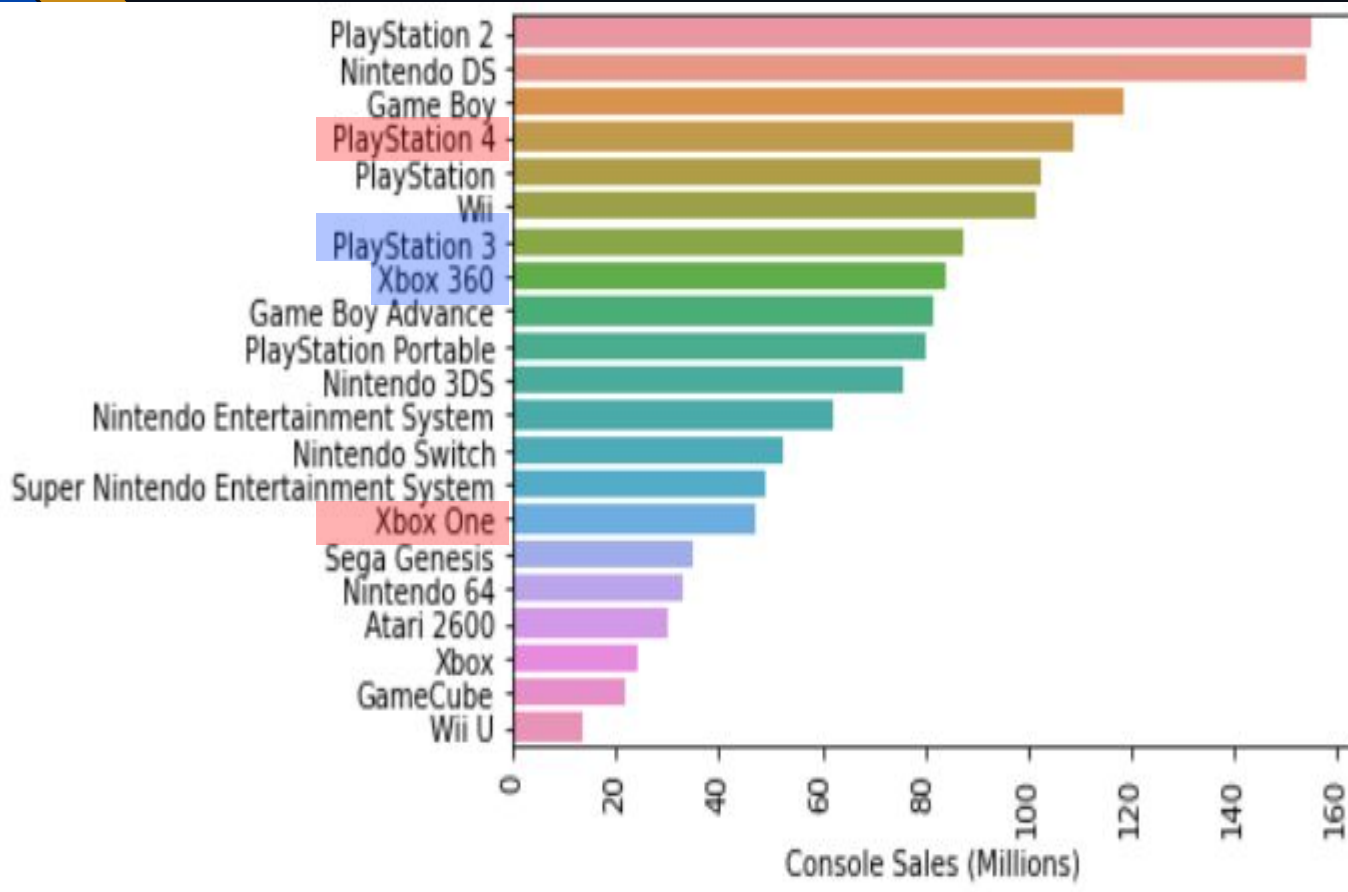
Playstation 4 had more than double the sales of the Xbox One. Hence, why it would have way more games sales for that console. This statement diminishes the strength of my hypothesis

Although, Playstation 3 had more sales than the Xbox 360, but the Xbox sold more copies for almost every game. This statement strengthens my hypothesis

Conclusions

Users tend to rate games higher than critics, but critics have less variation in their scores

Graphics processes tend to increase in quality exponentially over time

Graphics of a console has a relatively strong association with video game sales for that console in comparison to other consoles in its own generation (about each decade).

- If popularity of console sales and game release date not taken into account

More games are being sold more than ever before because of the wide popularity of gaming

Region influences the sales of games depending on genre

Works Cited

Processing Power: https://thegamingsetup.com/guides/console-power-comparison-chart

Game Sales: https://www.kaggle.com/baynebrannen/video-game-sales-2020

Console Sales: https://www.kaggle.com/datasets/jaimepazlopes/game-console-manufactor-and-sales