Board Games

Visualization Techniques produced in Python.

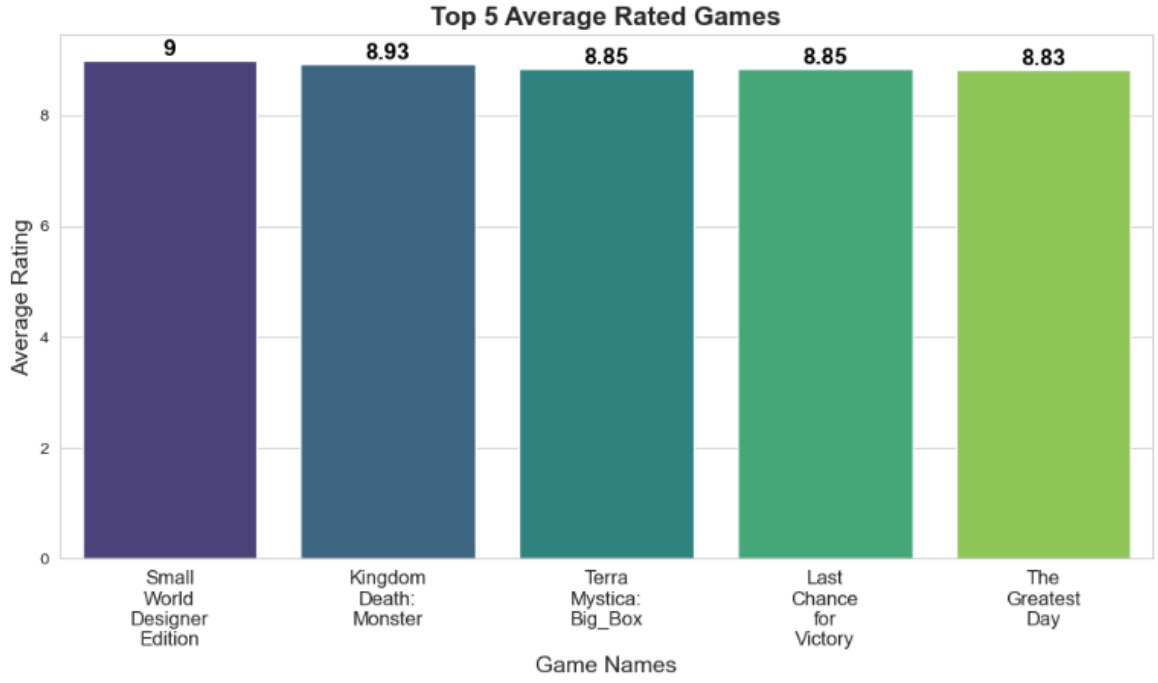
Code available on: https://github.com/norodrigues111/Python-Data-Visualization-Board-Games/blob/main/NataliaDeOliveiraRodrigues_DVisHDip_CA1.ipynb

What are the top 5 “average rated” games?

In [22]:
```
top5_avg_rated_df = df.sort_values(by='average_rating', ascending=False).head()
top5_avg_rated_df['average_rating'] = top5_avg_rated_df['average_rating'].round(2)
top5_avg_rated_df = top5_avg_rated_df[['name','average_rating']]
top5_avg_rated_df
```

Out[22]:

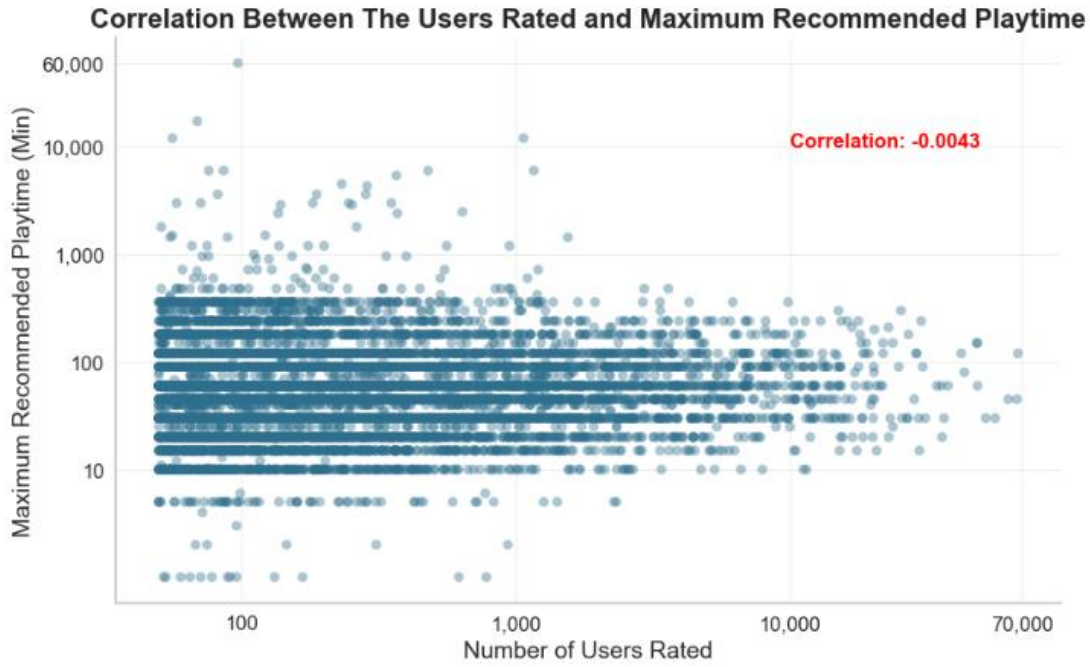
	name	average_rating
8348	Small World Designer Edition	9.00
6392	Kingdom Death: Monster	8.93
9964	Terra Mystica: Big Box	8.85
8526	Last Chance for Victory	8.85
9675	The Greatest Day: Sword, Juno, and Gold Beaches	8.83



Is there a correlation between the “users_rated” and the “max_playtime”?

In [27]:

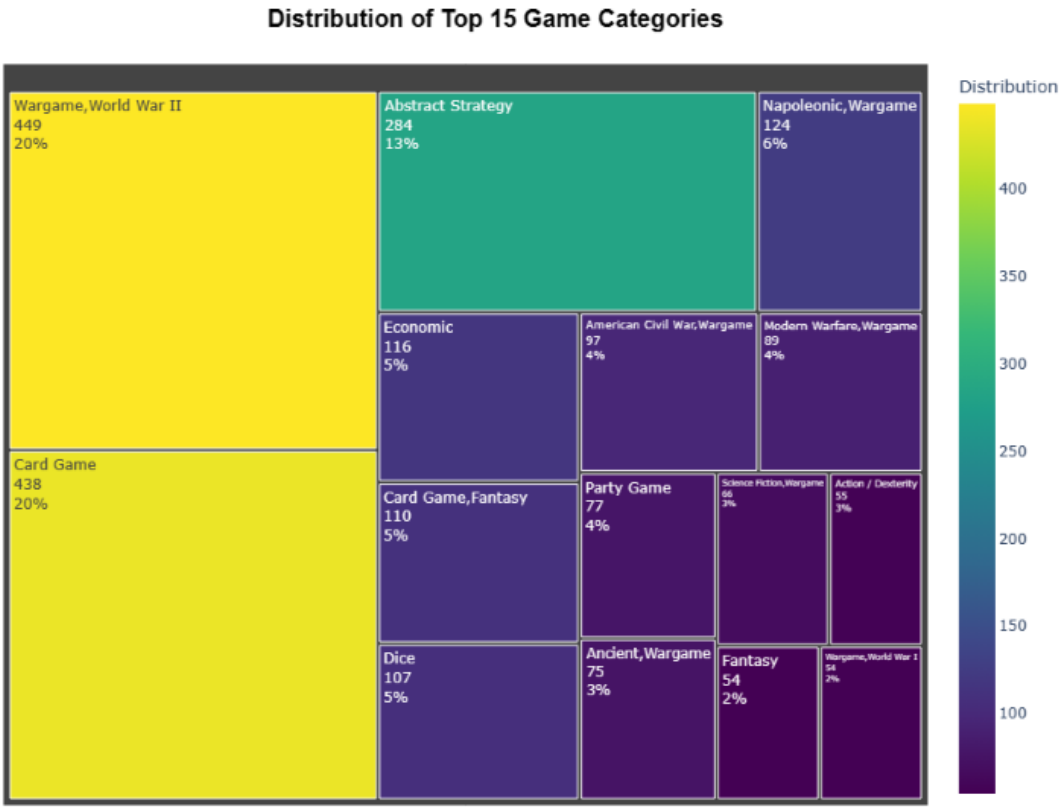
```
plt.figure(figsize=(10,6))

plt.scatter(df['users_rated'],df['max_playtime'], alpha=0.4, edgecolors='none',s=35,color='#2e6f8e',marker='o')
plt.yscale('log')
plt.xscale('log')
plt.xlabel('Number of Users Rated',fontsize=14)
plt.ylabel('Maximum Recommended Playtime (Min)',fontsize=14)
plt.title('Correlation Between The Users Rated and Maximum Recommended Playtime',fontsize=16,fontweight='bold')

text = f'Correlation: {corr_users_playtime:.4f}'
plt.text(10**4, 10**4, text, fontsize=12, color='red',fontweight='bold')

x_ticks = [100, 1000, 10000,70000]
plt.xticks(x_ticks, ['100','1,000','10,000','70,000'])

y_ticks = [10, 100, 1000, 10000,60000]
plt.yticks(y_ticks, ['10','100','1,000','10,000','60,000'])

plt.tick_params(axis='both', which='both', width=0, labelsize=12)

ax = plt.gca()
ax.spines[['top','right']].set_visible(False)
ax.spines[['bottom','left']].set_linewidth(1.5)

plt.grid(True,alpha=0.3)
plt.show()
```



What is the distribution of game categories?

In [33]:

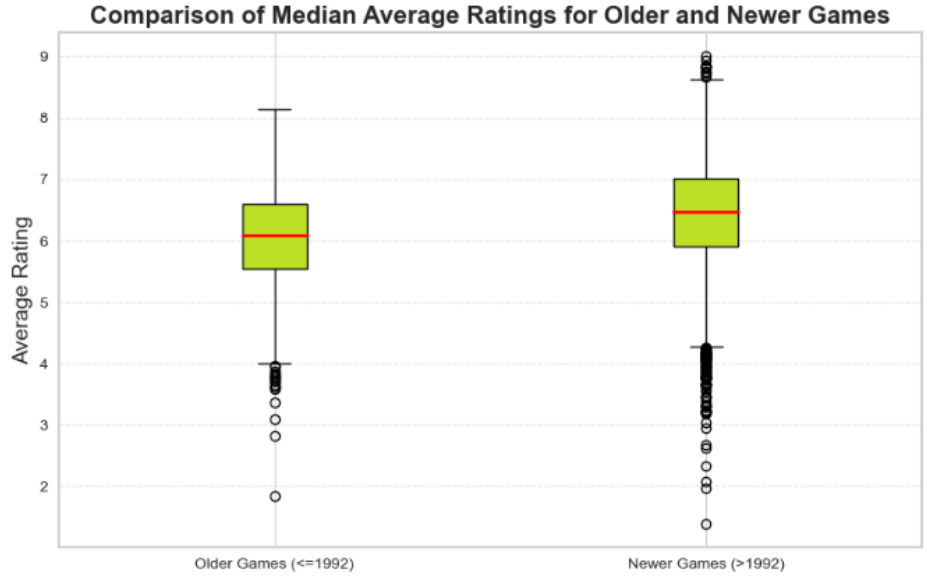
```
fig = px.treemap(category_count, path=['Category'], values='Distribution',
                 color='Distribution',
                 color_continuous_scale=px.colors.sequential.Viridis,
                 width=970, height=750)

fig.update_traces(textinfo='label+value+percent entry', hovertemplate=None)

fig.update_layout(title_text="<b>Distribution of Top 15 Game Categories",
                  title_font=dict(family="Arial", size=20, color="Black"),
                 title_x=0.5)

fig.show()
```



Do older games (1992 and earlier) have a higher MEDIAN “average rating” than newer games (after 1992)?

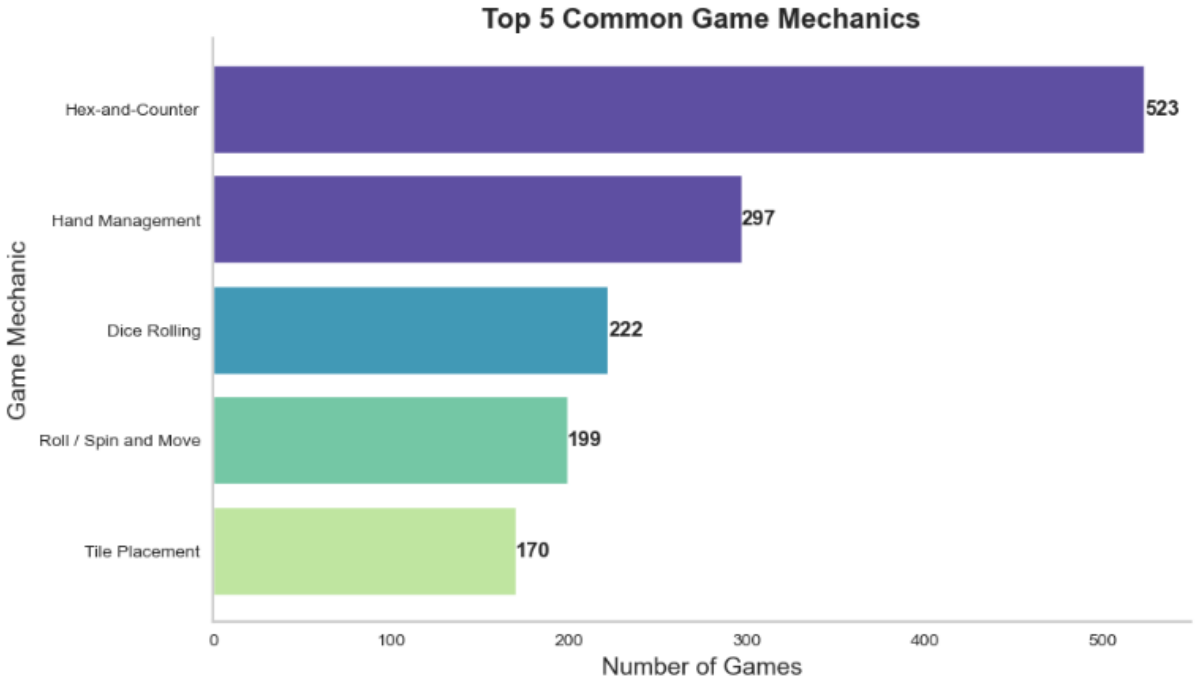
In [36]:

```
plt.figure(figsize=(10, 6))
plt.boxplot([older_games['average_rating'], newer_games['average_rating']],
            labels=["Older Games (<=1992)", "Newer Games (>1992)"],
           patch_artist=True,
                    boxprops={'facecolor': '#bddf26'},
                    medianprops={'color': 'red','linewidth': 2})

plt.ylabel("Average Rating",fontsize=14)
plt.title("Comparison of Median Average Ratings for Older and Newer Games", fontsize=16,fontweight='bold')

ax = plt.gca()
ax.spines[['top','right','bottom','left']].set_linewidth(1.5)

plt.grid(axis="y", linestyle="--", alpha=0.5)

plt.show()

```



What are the 5 most common “mechanics” in the dataset?

In [40]:
```
mechanic_df = mechanic_df.sort_values(by="count", ascending=True)

cmap = plt.get_cmap('Spectral')

plt.figure(figsize=(10, 6))
plt.barh(mechanic_df["mechanic"], mechanic_df["count"], color=cmap(mechanic_df["count"]))
plt.xlabel('Number of Games', fontsize=14)
plt.ylabel('Game Mechanic', fontsize=14)
plt.title("Top 5 Common Game Mechanics", fontsize=16,fontweight='bold')

for i, count in enumerate(mechanic_df["count"]):
    plt.text(count, i, str(count), va='center',fontsize=12,fontweight='bold')

ax = plt.gca()
ax.spines[['top','right']].set_visible(False)
ax.spines[['bottom','left']].set_linewidth(1.5)

plt.grid(False)
plt.show()

```

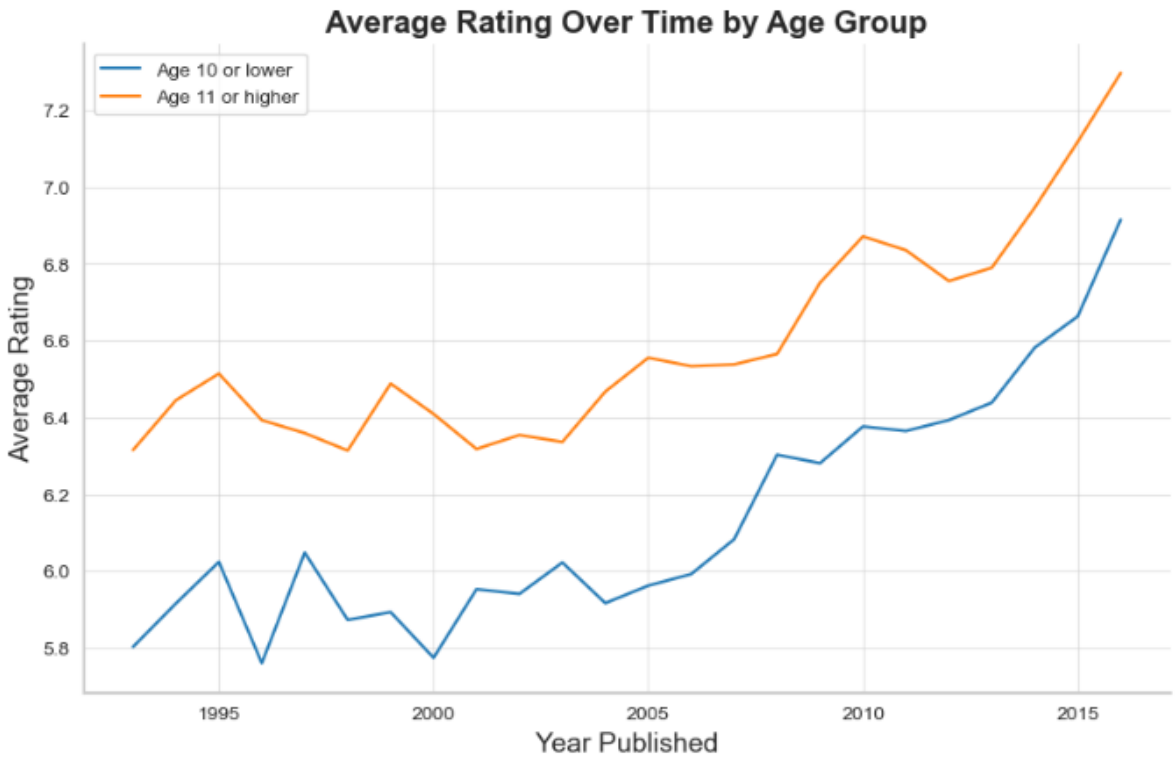


Is there a statistically significant difference in the newer game average ratings between games where the minimum age recommended is equal or lower than 10 and games where the minimum age recommended is equal or higher than 11, and does the year of publication (>1992) have a significant impact on that?

In [41]:
```
equal_or_lower_10 = df[(df['min_age']<=10) & (df['year_published']>1992)]
equal_or_higher_11 = df[(df['min_age']>10) & (df['year_published']>1992)]
equal_or_lower_10.shape,equal_or_higher_11.shape
```

Out[41]: ((5179, 22), (3419, 22))

In [42]:
```
plt.figure(figsize=(10,6))

plt.plot(equal_or_lower_10.groupby('year_published')['average_rating'].mean(), label='Age 10 or lower')
plt.plot(equal_or_higher_11.groupby('year_published')['average_rating'].mean(),label='Age 11 or higher')

plt.title('Average Rating Over Time by Age Group', fontsize=16,fontweight='bold')
plt.xlabel('Year Published',fontsize=14)
plt.ylabel('Average Rating',fontsize=14)

ax = plt.gca()
ax.spines[['top','right']].set_visible(False)
ax.spines[['bottom','left']].set_linewidth(1.5)

plt.legend()
plt.grid(linestyle="-", alpha=0.5)
plt.show()
```



In [43]:
```
t_stat, p_value = ttest_ind(equal_or_lower_10['average_rating'],equal_or_higher_11['average_rating'])

if p_value >= 0.05:
    print(f'Test Result: {p_value}. There is no statistically significance difference in the average rating between the a
else:
    print(f'Test Result: {p_value}. There is statistically significance difference in the average rating between the age
```

```
Test Result: 1.0178865604825051e-152. There is statistically significance difference in the average rating between the age g
roups.
```

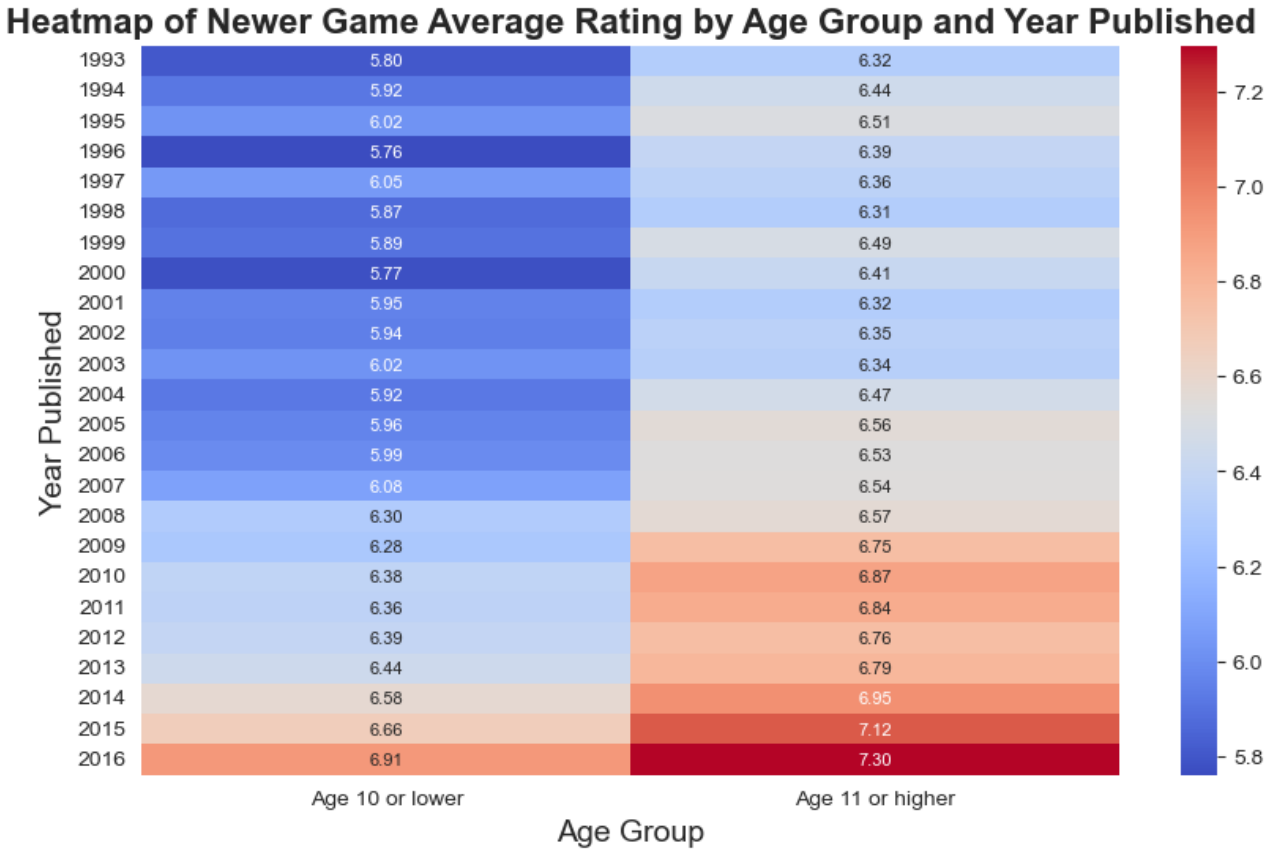
In [44]:
```
age_group_df = pd.concat([equal_or_lower_10,equal_or_higher_11],keys=['Age 10 or lower','Age 11 or higher'],
                         names=['Age Group']).reset_index(level=0)
```

In [45]:
```
heatmap_df = age_group_df.pivot_table(index='year_published',columns='Age Group',values='average_rating',aggfunc='mean')
heatmap_df.head()
```

Out[45]:

Age Group	Age 10 or lower	Age 11 or higher
year_published		
1993	5.802627	6.315130
1994	5.915850	6.444991
1995	6.023819	6.514107
1996	5.760443	6.393016
1997	6.048138	6.359263

In [46]:
```
plt.figure(figsize=(10,6))

sns.heatmap(heatmap_df,cmap='coolwarm',annot=True, annot_kws={'size':8},fmt=".2f")

plt.title('Heatmap of Newer Game Average Rating by Age Group and Year Published',fontsize=16,fontweight='bold')
plt.xlabel('Age Group',fontsize=14)
plt.ylabel('Year Published',fontsize=14)

plt.show()
```


Heatmap of Newer Game Average Rating by Age Group and Year Published
Year Published
1993 5.80 6.32
1994 5.92 6.44
1995 6.02 6.51
1996 5.76 6.39
1997 6.05 6.36
1998 5.87 6.31
1999 5.89 6.49
2000 5.77 6.41
2001 5.95 6.32
2002 5.94 6.35
2003 6.02 6.34
2004 5.92 6.47
2005 5.96 6.56
2006 5.99 6.53
2007 6.08 6.54
2008 6.30 6.57
2009 6.28 6.75
2010 6.38 6.87
2011 6.36 6.84
2012 6.39 6.76
2013 6.44 6.79
2014 6.58 6.95
2015 6.66 7.12
2016 6.91 7.30
7.2
7.0
6.8
6.6
6.4
6.2
6.0
5.8
Age 10 or lower
Age 11 or higher
Age Group