In [4]:

```
import pandas as pd
import numpy as np
from scipy import stats
import seaborn as sns
sns.set(color_codes=True)
import matplotlib.pyplot as plt
from scipy.stats import ttest_1samp, ttest_ind, mannwhitneyu, levene, shapiro
from statsmodels.stats.power import ttest_power
```

In [2]:

```
dataframe = pd.read_csv("games-1.csv")
```

In [141]:

```
dataframe.head()
```

Out[141]:

	id	type	name	yearpublished	minplayers	maxplayers	playingtime	minplaytime	ma
0	1	boardgame	Twilight Struggle	2005	2	2	180	180	
1	2	boardgame	Terra Mystica	2012	2	5	150	60	
2	3	boardgame	Caverna: The Cave Farmers	2013	1	7	210	30	
3	4	boardgame	Through the Ages: A Story of Civilization	2006	2	4	240	240	
4	5	boardgame	Puerto Rico	2002	2	5	150	90	

In [142]:

```
dataframe=dataframe[dataframe.isnull().any(axis=1)]
dataframe
```

Out[142]:

id	type	name	yearpublished	minplayers	maxplayers	playingtime	minplaytime	maxplaytime	ı

In [143]:
```
dataframe[dataframe.isnull().any(axis=1)].count()
```

Out[143]:
```
id                      0
type                    0
name                    0
yearpublished           0
minplayers              0
maxplayers              0
playingtime             0
minplaytime             0
maxplaytime             0
minage                  0
users_rated             0
average_rating          0
bayes_average_rating    0
total_owners            0
total_traders           0
total_wanters           0
total_wishers           0
total_comments          0
total_weights           0
average_weight          0
dtype: int64
```

In [147]:

```
dataframe=dataframe[np.abs(dataframe.playingtime-dataframe.playingtime.mean())<=(
dataframe
```

Out[147]:

	id	type	name	yearpublished	minplayers	maxplayers	playingti
0	1	boardgame	Twilight Struggle	2005	2	2	
1	2	boardgame	Terra Mystica	2012	2	5	
2	3	boardgame	Caverna: The Cave Farmers	2013	1	7	
3	4	boardgame	Through the Ages: A Story of Civilization	2006	2	4	
4	5	boardgame	Puerto Rico	2002	2	5	
5	6	boardgame	Agricola	2007	1	5	
6	7	boardgame	Android: Netrunner	2012	2	2	
7	8	boardgame	Mage Knight Board Game	2011	1	4	
8	9	boardgame	The Castles of Burgundy	2011	2	4	
9	10	boardgame	Eclipse	2011	2	6	
10	11	boardgame	Power Grid	2004	2	6	
11	12	boardgame	Star Wars: Imperial Assault	2014	2	5	
12	13	boardgame	War of the Ring (second edition)	2012	2	4	
13	14	boardgame	Robinson Crusoe: Adventures on the Cursed Island	2012	1	4	
14	15	boardgame	Le Havre	2008	1	5	
15	16	boardgame	Brass	2007	3	4	
16	17	boardgame	Tzolk'in: The Mayan Calendar	2012	2	4	
17	18	boardgame	Dead of Winter: A Crossroads Game	2014	2	5	
18	19	boardgame	7 Wonders	2010	2	7	
19	20	boardgame	Caylus	2005	2	5	
20	21	boardgame	Keyflower	2012	2	6	
21	22	boardgame	Dominion: Intrigue	2009	2	4	
22	23	boardgame	Dominant Species	2010	2	6	
23	24	boardgame	Race for the Galaxy	2007	2	4	
24	25	boardgame	Star Wars: X-Wing Miniatures Game	2012	2	4	
25	26	boardgame	El Grande	1995	2	5	

	id	type	name	yearpublished	minplayers	maxplayers	playingt
26	27	boardgame	Eldritch Horror	2013	1	8	
27	28	boardgame	Lords of Waterdeep	2012	2	5	
28	29	boardgame	Battlestar Galactica	2008	3	6	
29	30	boardgame	Twilight Imperium (Third Edition)	2005	3	6	
...	...	...	...	...	...	...	
2635	2636	boardgameexpansion	Mad Scientist University Course Packet: PSI PHI	2005	3	7	
2636	2637	boardgameexpansion	Mad Scientist University Course Packet: Indepe...	2005	3	7	
2637	2638	boardgameexpansion	Mad Scientist University Course Packet: Home E...	2005	3	7	
2638	2639	boardgameexpansion	Paris Paris: Zusatzkarten Expansion	2003	2	4	
2639	2640	boardgameexpansion	Knatsch: Katalog Edition 1	2002	2	6	
2640	2641	boardgameexpansion	Knatsch: Katalog Edition 2	2002	2	6	
2641	2642	boardgameexpansion	Knatsch: Katalog Edition 3	2002	2	6	
2642	2643	boardgameexpansion	Starfight: Star Captains	2005	2	2	
2643	2644	boardgameexpansion	Starfight: Expansion Pack I ships	2005	2	2	
2644	2645	boardgameexpansion	Starfight: Expansion Pack III ships	2005	2	2	
2645	2646	boardgameexpansion	Starfight: Expansion Pack II Starmaps	2005	2	2	
2646	2647	boardgameexpansion	Starfight: Expansion Pack V Ships	2005	2	2	
2647	2648	boardgameexpansion	Starfight: Expansion Pack IV Death Cube	2005	2	2	
2648	2649	boardgameexpansion	Starfight: Drakus Empire Captains	2004	2	2	
2649	2650	boardgameexpansion	La die Kirche ins Dorf Erweiterung	2005	2	6	
2650	2651	boardgameexpansion	From the Casebook of Nick Velvet	1986	1	6	

	id	type	name	yearpublished	minplayers	maxplayers	playingti
2651	2652	boardgameexpansion	Age of Steam Expansion: 1830's Pennsylvania / ...	2006	3	6	
2652	2653	boardgameexpansion	Killer Bunnies Bunny Blanks #2	2006	2	8	
2653	2654	boardgameexpansion	Killer Bunnies Bunny Blanks #1	2006	2	8	
2654	2655	boardgameexpansion	Starfight: Mohrg Empire Captains	2004	2	2	
2655	2656	boardgameexpansion	Dungeon Twister: Minotaur	2006	2	2	
2656	2657	boardgameexpansion	China: Grenzstreitigkeiten	2006	3	5	
2657	2658	boardgameexpansion	Havoc Expansion: The Character Cards	2005	2	6	
2658	2659	boardgameexpansion	Runebound: Avatars of Kelnov	2006	2	6	
2659	2660	boardgameexpansion	Runebound: Cult of the Rune	2006	2	6	
2660	2661	boardgameexpansion	Runebound: Drakes and Dragonspawn	2006	2	6	
2661	2662	boardgameexpansion	Runebound: Walkers of the Wild	2006	2	6	
2662	2663	boardgameexpansion	Buckeyes!	2002	2	4	
2663	2664	boardgameexpansion	Star Fleet Battles: Commanders SSD Book #1	1983	2	12	
2664	2665	boardgameexpansion	Star Fleet Battles: Commanders SSD Book #2	1983	2	12	

2658 rows × 20 columns

In [148]:
```
dataframe.groupby('type')['playingtime'].mean()
```

Out[148]:
```
type
boardgame              87.641077
boardgameexpansion    101.390452
Name: playingtime, dtype: float64
```

```
In [3]: dataframe.groupby('type')['maxplayers'].mean()
```

```
Out[3]: type
        boardgame               4.983244
        boardgameexpansion      5.608696
        Name: maxplayers, dtype: float64
```

```
In [149]: dataframe.groupby('type')['playingtime'].count()
```

```
Out[149]: type
          boardgame               1485
          boardgameexpansion      1173
          Name: playingtime, dtype: int64
```

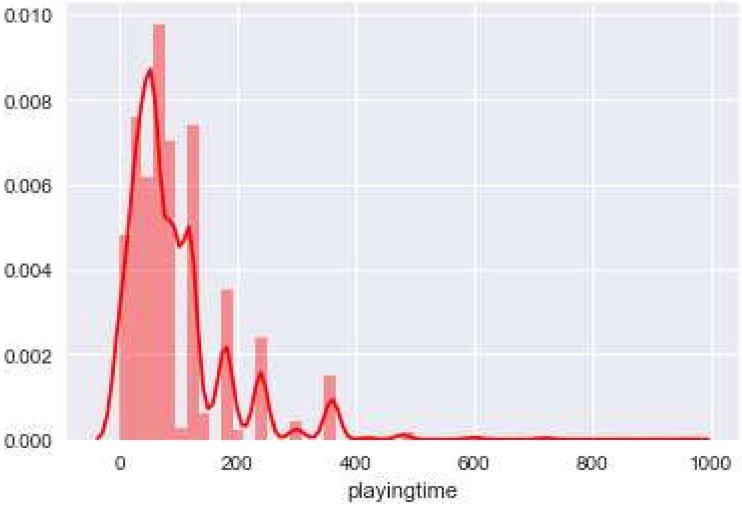
```
In [150]: dataframe.groupby('type')['playingtime'].sum()
```

```
Out[150]: type
          boardgame               130147
          boardgameexpansion      118931
          Name: playingtime, dtype: int64
```

```
In [155]: sns.distplot(dataframe['playingtime'],color='red')
```

```
D:\Anaconda\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The
'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```

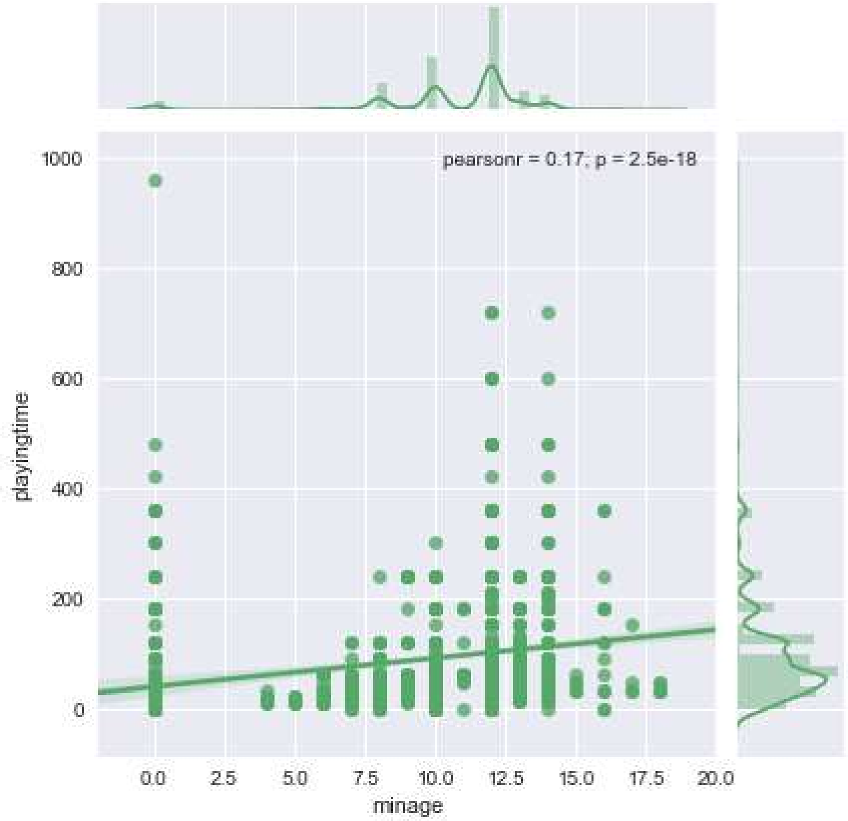
```
Out[155]: <matplotlib.axes._subplots.AxesSubplot at 0x2399ae9fef0>
```



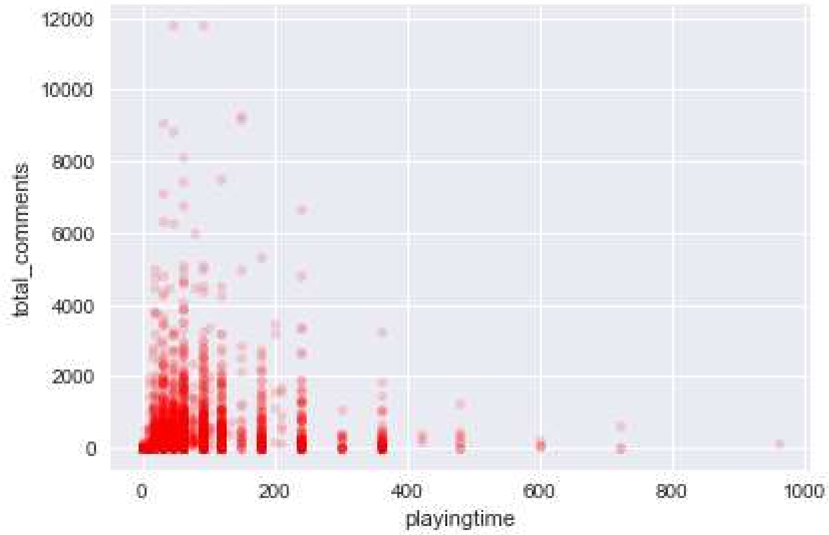
In the above ghraph, I observed that there is skewness towords right. It shown the distribution of playingtime.

In [156]:
```
sns.jointplot(data=dataframe, x='minage', y='playingtime', kind='reg', color='g')
plt.show()
```

```
D:\Anaconda\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The
'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
D:\Anaconda\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The
'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```



In [157]:
```
dataframe.plot.scatter(x='playingtime', y='total_comments', alpha=.15,color='red'
plt.tight_layout()
```

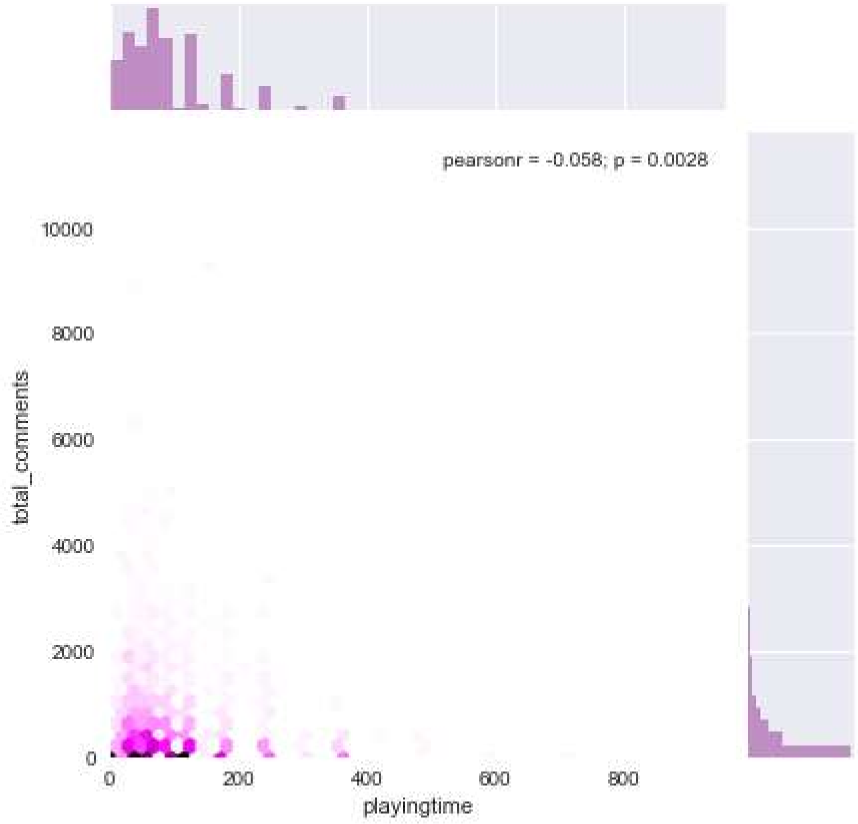


In the above scatter plot i observed that there is no relation between playingtime and total_comments.It means as one increass or decreses the other parameter is not correspondingly increasing or decreasing

In [160]:

```
sns.jointplot(dataframe['playingtime'], dataframe['total_comments'], kind="hex",c
plt.show()
```

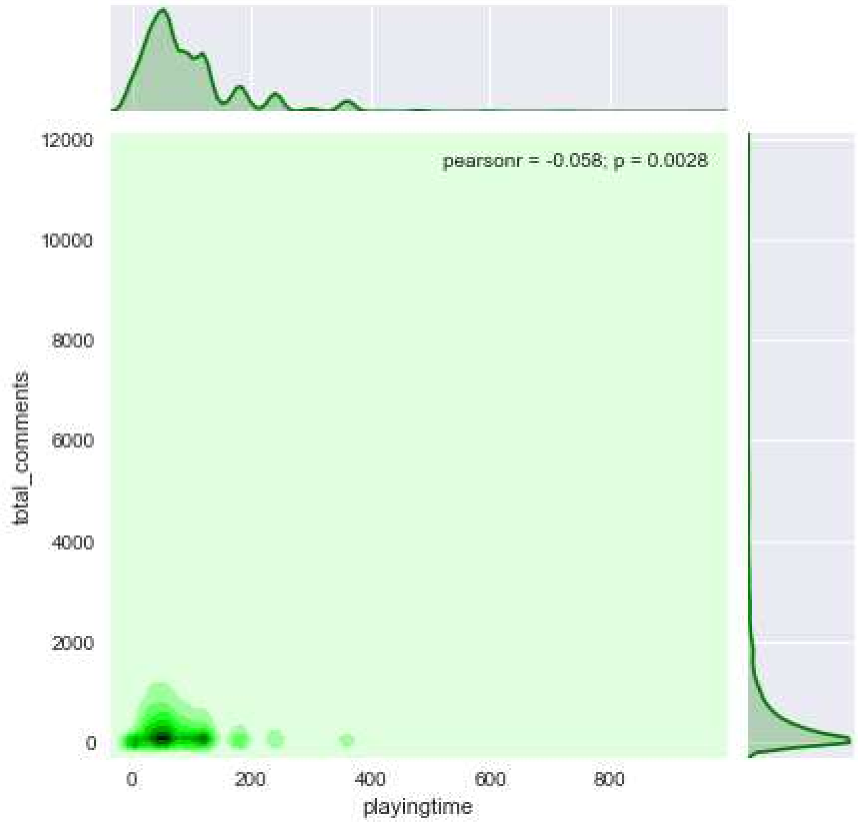
```
D:\Anaconda\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The
'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
D:\Anaconda\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The
'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```



In the above graph the darker hexagon indicate that maximun no. of points lying in that region and at other places the data points are minimun

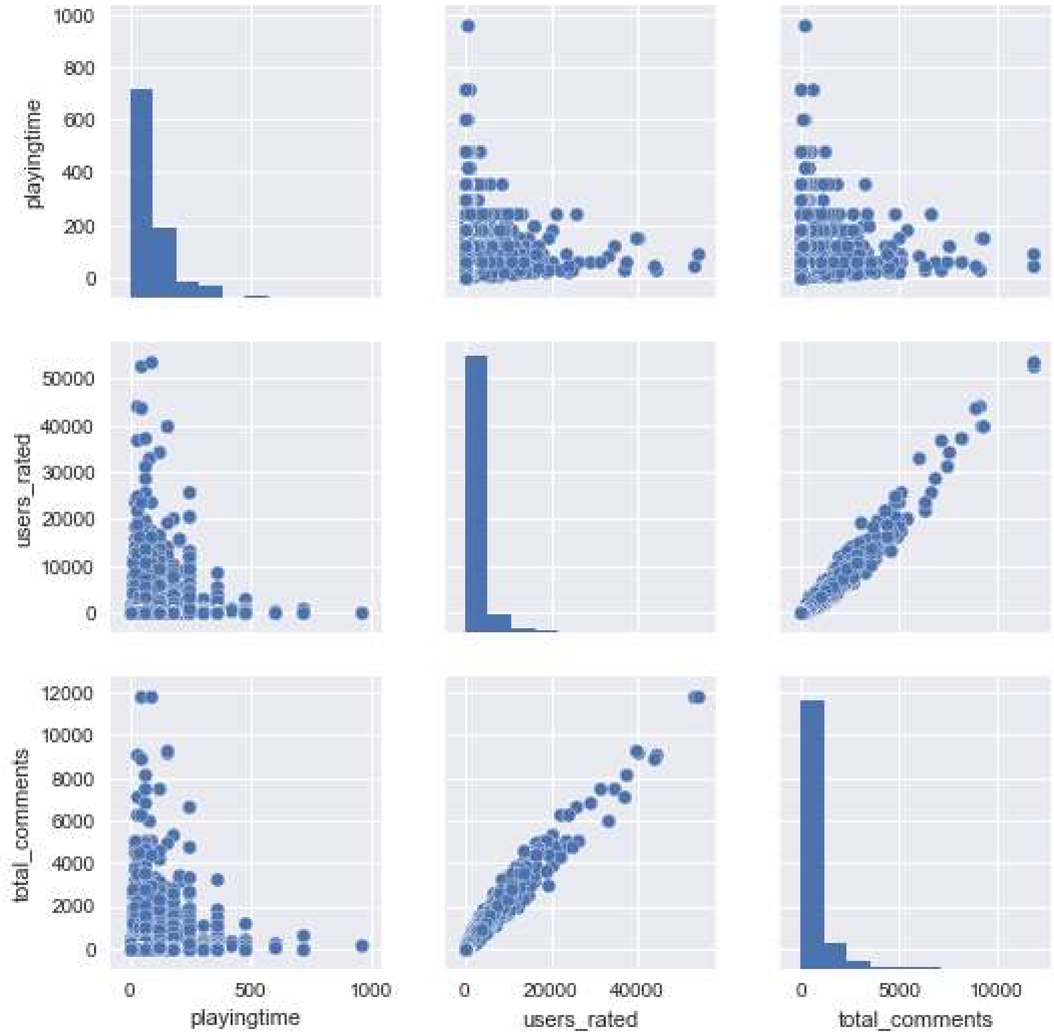
In [161]:

```
sns.jointplot(dataframe['playingtime'], dataframe['total_comments'], kind="kde",c
plt.show()
```



The above graph represents the maximum number of points occur in this green range only

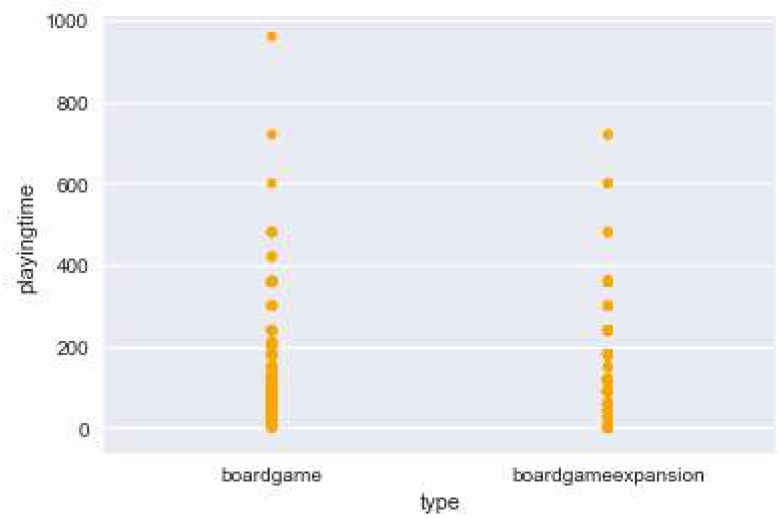
```
In [163]: sns.pairplot(dataframe[['playingtime', 'users_rated', 'total_comments']])
          plt.show()
```



Above graph indicates the relationship between playingtime, users_rated and total_comments. There is a direct relationship between users_rated and total_comments.

```
In [164]: sns.stripplot(dataframe['type'], dataframe['playingtime'], color='orange')
```

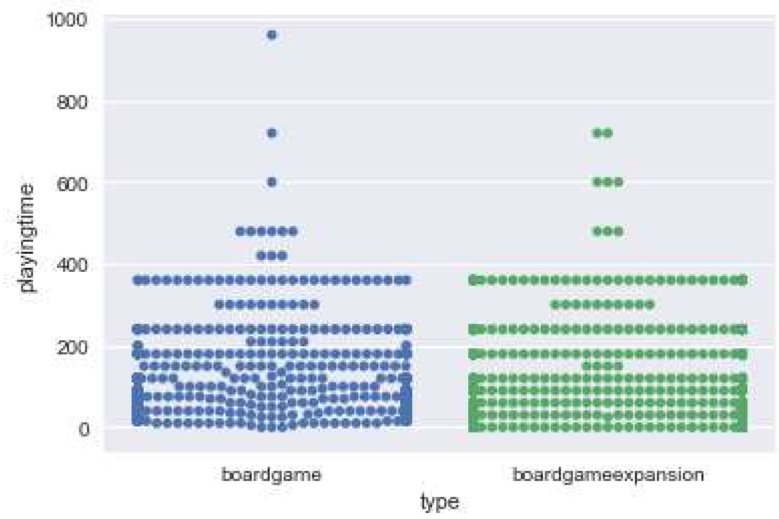
```
Out[164]: <matplotlib.axes._subplots.AxesSubplot at 0x239980caf98>
```



A different approach would be to use the function swarmplot(), which positions each scatterplot point on the categorical axis with an algorithm that avoids overlapping points:

```
In [165]: sns.swarmplot(dataframe['type'], dataframe['playingtime'])
```

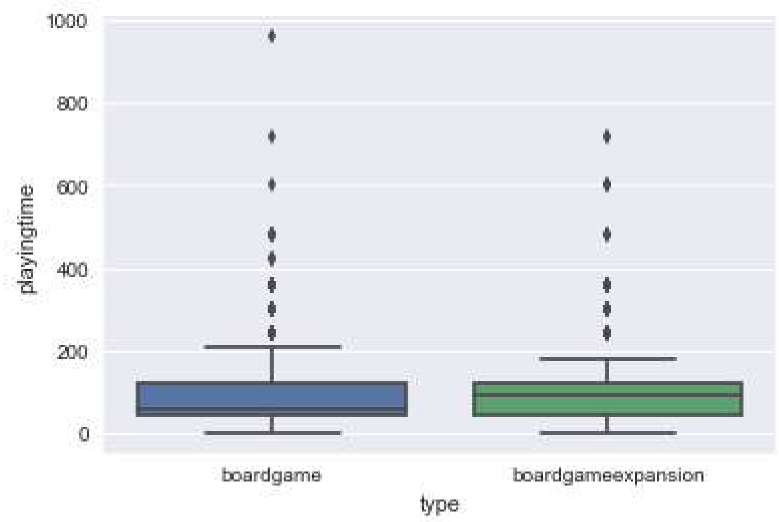
```
Out[165]: <matplotlib.axes._subplots.AxesSubplot at 0x239a2655a58>
```



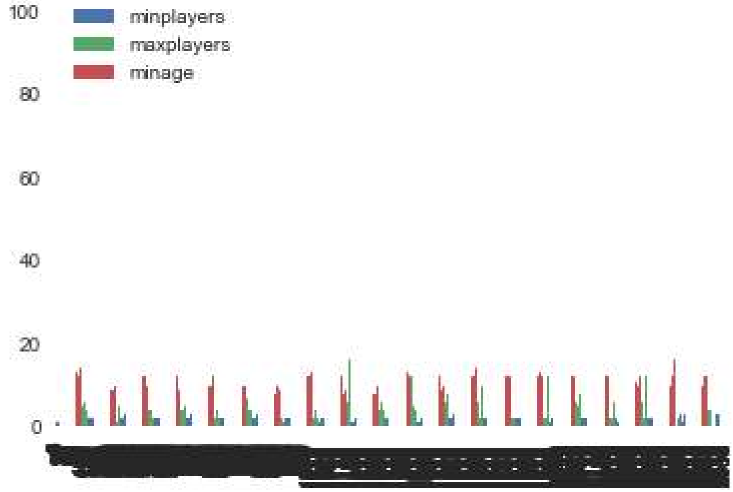
Here we can see the distribution of playingtime for the each game type by using swarmplot. There is no overlapping of the points

In [167]:
```
sns.boxplot(dataframe['type'], dataframe['playingtime'])
```

Out[167]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x23997654da0>
```

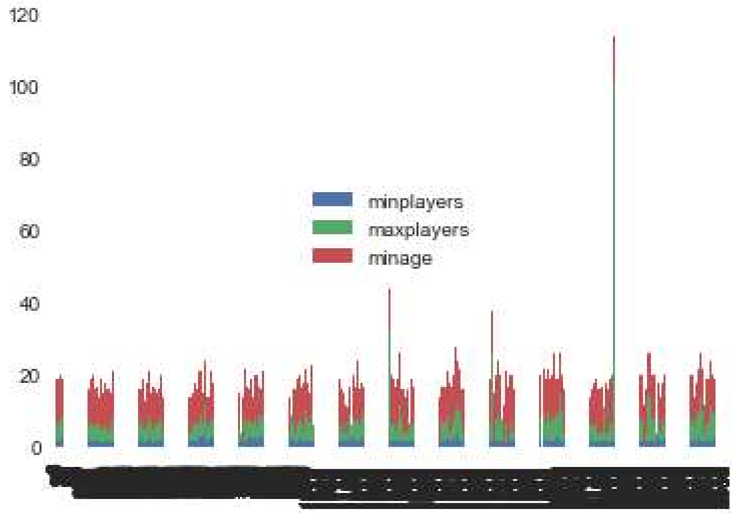


In [168]:
```
df2 = pd.DataFrame(dataframe, columns=['minplayers', 'maxplayers', 'minage'])
df2.plot.bar();
```



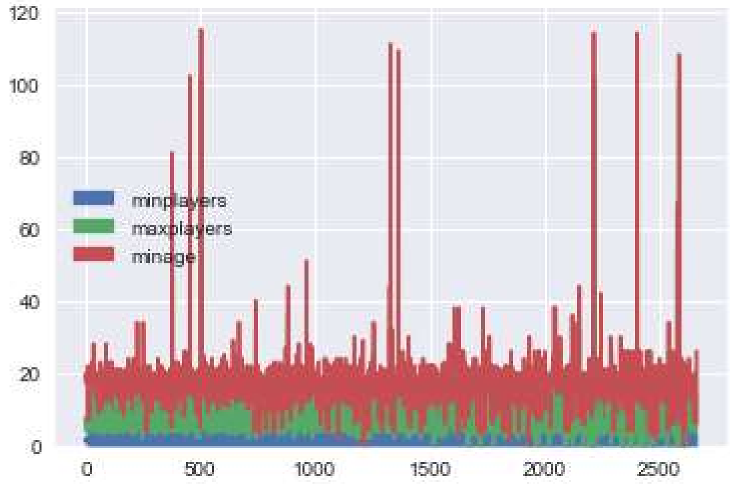
In [111]:

```
df2.plot.bar(stacked=True);
```



In [112]:

```
df2.plot.area();
```

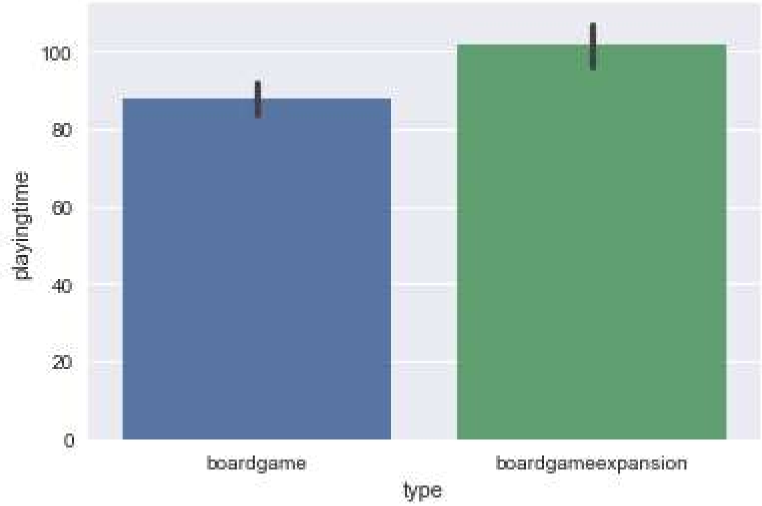


In [154]:

```
sns.barplot(dataframe['type'], dataframe['playingtime'])
```

Out[154]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x239a3d0fe80>
```



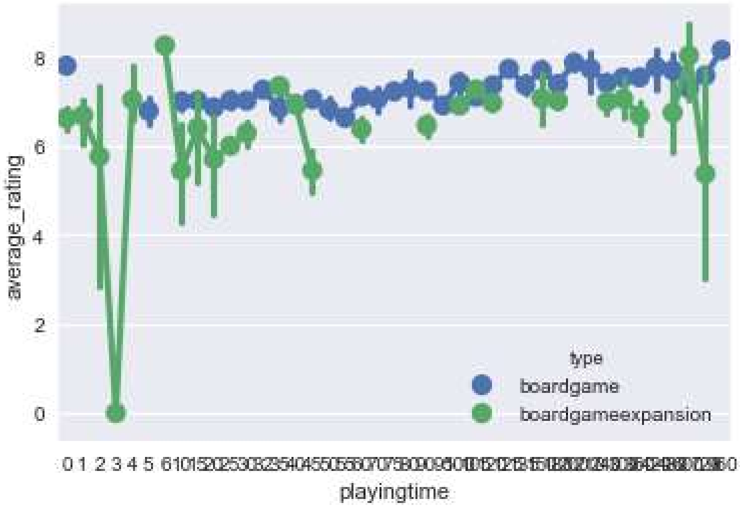
Here we can observe that playingtime of boardgameexpansion is higher than boardgame

In [169]:

```
sns.pointplot(dataframe['playingtime'], dataframe['average_rating'], hue=datafram
```

Out[169]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2399a42e630>
```



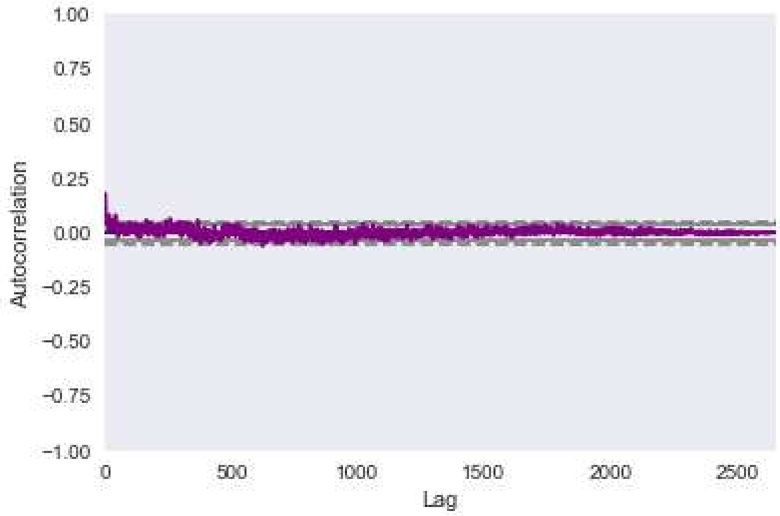
In [170]:

```
from pandas.tools.plotting import autocorrelation_plot
autocorrelation_plot(dataframe.playingtime,color='purple')
```

```
D:\Anaconda\lib\site-packages\ipykernel_launcher.py:2: FutureWarning: 'pandas.t
ools.plotting.autocorrelation_plot' is deprecated, import 'pandas.plotting.auto
correlation_plot' instead.
```

Out[170]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x239947bed68>
```



Two sample t-Test

In [ ]:

```
#two-sample t-test
#null hypothesis:  the Amount of Time Gamers spend for each Game Type is equal
#alternate hypothesis: the Amount of Time Gamers spend for each Game Type is not
#this test assumes the two groups have the same variance...
#(can be checked with tests for equal variance - Levene)
```

In [55]:

```
# Seperating the data into 2 groups
group1 = dataframe[dataframe['type'] == 'boardgame']['playingtime']
group2 = dataframe[dataframe['type'] == 'boardgameexpansion']['playingtime']
```

In [56]:

```
t_statistic, p_value=ttest_ind(group1, group2)
print(t_statistic, p_value)
```

```
-4.03345605586225 5.651199025520449e-05
```

In [50]:

```
# p_value < 0.05 => alternative hypothesis:
# the Amount of Time Gamers spend for each Game Type is not equal
print ("two-sample t-test p-value=", p_value)
```

```
two-sample t-test p-value= 5.651199025520449e-05
```

Here, I can conclude that the nill hypothesis is rejected. It means that Amount of Time Gamers spend for each Game Type is not equal.

In [62]:

```
# two-sample wilcoxon test
# a.k.a Mann Whitney U - Used when samples are not normally distributed
u, p_value = mannwhitneyu(group1, group2)
print ("two-sample wilcoxon-test p-value=", p_value)
```

```
two-sample wilcoxon-test p-value= 0.00036885766918812494
```

In [58]:

```
# Calculating Power of Test
# Compute the difference in Means between 2 sample means and divide by pooled Sta
# number of Observations/tuples
# Set the alpha value to 0.05 and alternative values 'two-sided' , 'larger' , 'sm
(np.mean(group1) - np.mean(group2)) / np.sqrt(((1485-1)*np.var(group1)+(1173-1)*n
```

Out[58]:

```
-0.1130929016695113
```

In [59]:

```
print(ttest_power(-0.11, nobs=22, alpha=0.10, alternative='two-sided'))
```

```
0.1420216683112745
```

The power of any test of statistical significance is defined as the probability that it will reject a false null hypothesis. Statistical power is inversely related to beta or the probability of making a Type II error. In short, power = 1 – β.

Result ¶

After Analysis, I conclude that the Amount of Time Gamers spend for each Game Type is not equal. The average playing time for Boardgame type is lesser than the Boardgameexpansion and the maximun number of players for boardgame type is lesser than the Boardgameexpansion. The average amount of time spend by gamers for each Game Type is

type

boardgame 87.641077

boardgameexpansion 101.390452

If we increase the maximum number of players for boardgame type then the amount of Time Gamers spend for boardgame type will increase.