In [43]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [38]:

```
matches = pd.read_csv("E:\\Datasets\\DataSet\\Football 1872-2019.csv")
```

In [39]:

```
matches.head()
```

Out[39]:

	date	home_team	away_team	home_score	away_score	tournament	city	country
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland
1	1873-03-08	England	Scotland	4	2	Friendly	London	England
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland
3	1875-03-06	England	Scotland	2	2	Friendly	London	England
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland

In [77]:

```
matches.shape
```

Out[77]:

```
(40945, 9)
```

In [40]:

```
matches.isnull().sum()
```

Out[40]:

```
date          0
home_team     0
away_team     0
home_score    0
away_score    0
tournament    0
city          0
country       0
neutral       0
dtype: int64
```

In [192]:

```
matches.describe()
```

Out[192]:

	home_score	away_score	year
count	40945.000000	40945.000000	40945.000000
mean	1.745463	1.188521	1990.024399
std	1.748619	1.405116	24.196213
min	0.000000	0.000000	1872.000000
25%	1.000000	0.000000	1977.000000
50%	1.000000	1.000000	1997.000000
75%	2.000000	2.000000	2008.000000
max	31.000000	21.000000	2019.000000

In [41]:

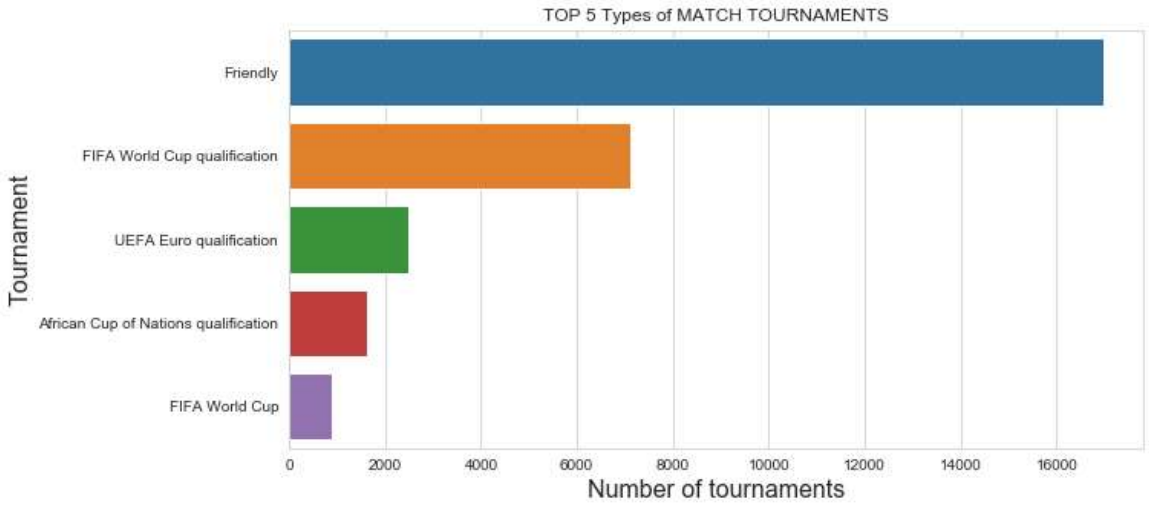
```
matches = matches.astype({'date':'datetime64[ns]'})
```

In [82]:

```
tournament = matches['tournament'].value_counts()
tournament = tournament[:5]
plt.figure(figsize = (10,5))
barplot = sns.barplot(y=tournament.index, x=tournament.values)
barplot.set_ylabel('Tournament', size=16)
barplot.set_xlabel('Number of tournaments', size=16)
barplot.set_title("TOP 5 Types of MATCH TOURNAMENTS")
```

Out[82]:

```
Text(0.5, 1.0, 'TOP 5 Types of MATCH TOURNAMENTS')
```



In [255]:

```
tournament1 = matches['city'].value_counts()
tournament1 = tournament1[:5]

plt.figure(figsize = (10,5))
barplot = sns.barplot(y=tournament1.index, x=tournament1.values,color='pink')
barplot.set_ylabel('Cities', size=16)
barplot.set_xlabel('Number of Cities', size=16)
barplot.set_title("TOP 5 ""Match hosting cities""")
```

Out[255]:

```
Text(0.5, 1.0, 'TOP 5 Match hosting cities')
```

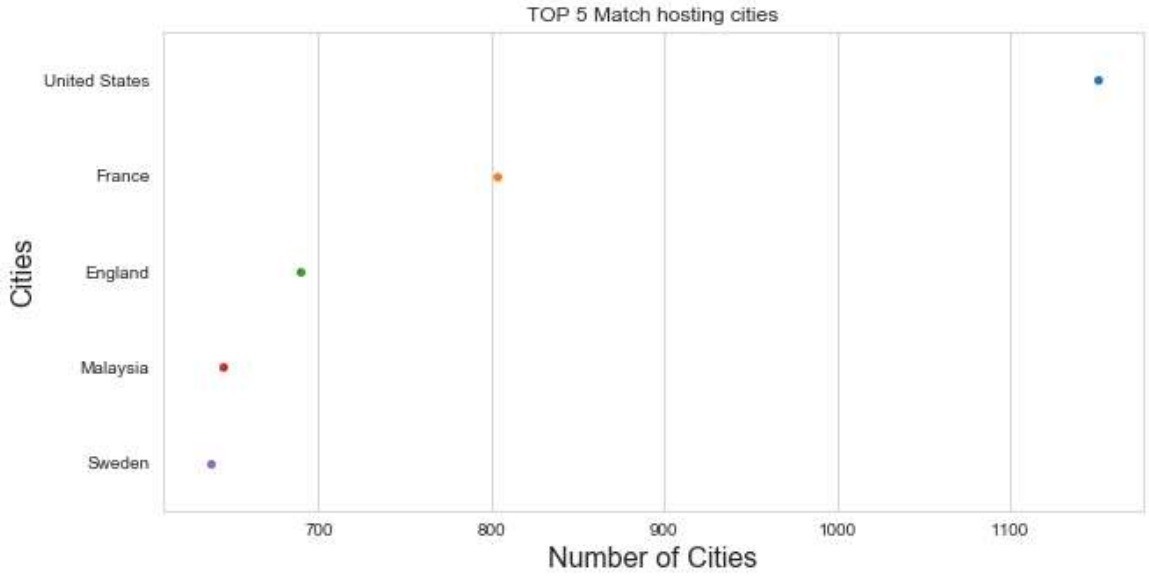
In [123]:

```
tournament1 = matches['country'].value_counts()
tournament1 = tournament1[:5]

plt.figure(figsize = (10,5))
barplot = sns.swarmplot(y=tournament1.index, x=tournament1.values)
barplot.set_ylabel('Cities', size=16)
barplot.set_xlabel('Number of Cities', size=16)
barplot.set_title("TOP 5 ""Match hosting cities""")
```

Out[123]:

```
Text(0.5, 1.0, 'TOP 5 Match hosting cities')
```

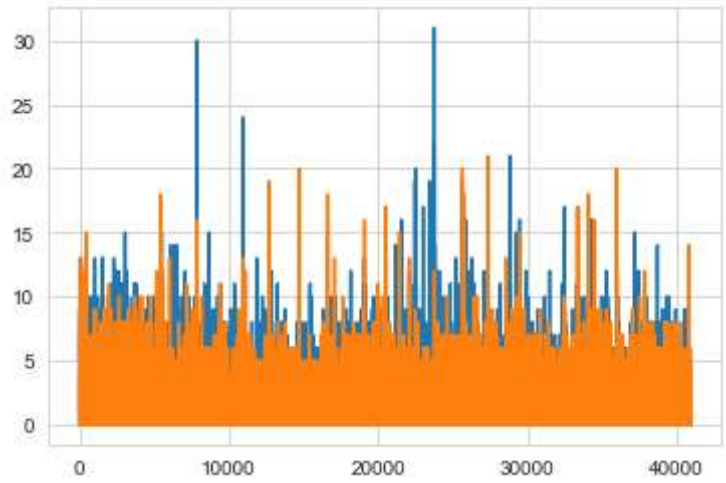


In [109]:

```
plt.plot(matches['home_score'])
plt.plot(matches['away_score'])
```

Out[109]:

```
[<matplotlib.lines.Line2D at 0x17f286f1198>]
```



In [112]:

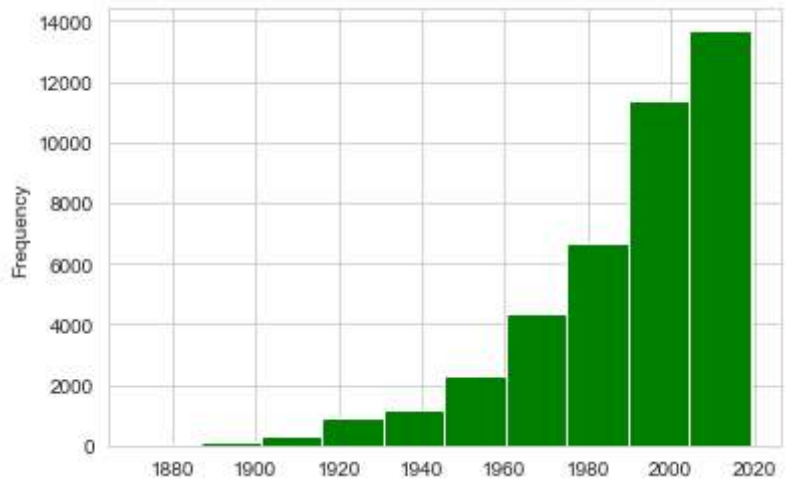
```
matches['date'] = pd.to_datetime(matches['date'])
matches['year'] = matches['date'].dt.year
```

In [199]:

```
matches['year'].plot(kind='hist',color='green')
```

Out[199]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x17f2d795dd8>
```



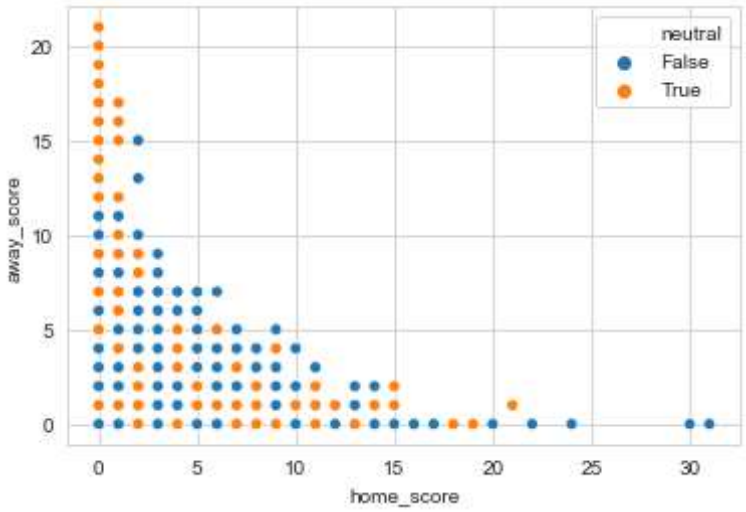
In [189]:

```
print("Total number of tournaments:{0}".format(len(matches['tournament'].unique())))
print("Total number of countries participated: {0}".format(len(matches['country'].uniqu
e())))
print("Total number of cities hosted for matches: {0}".format(len(matches['city'].uniqu
e())))
```

```
Total number of tournaments:109
Total number of countries participated: 266
Total number of cities hosted for matches: 1965
```

In [196]:

```
sns.scatterplot(x="home_score", y="away_score",hue="neutral",data=matches)
plt.show()
```



In [218]:

```
print(len(matches.loc[matches['tournament'] == 'FIFA World Cup']))
```

```
900
```

In [248]:

```
wins = matches.home_score>matches.away_score
lost = matches.home_score<matches.away_score
draw = matches.home_score==matches.away_score
```

In [251]:

```
print("Total match wins = {0}".format(len(matches[wins])))
print("Total match lost = {0}".format(len(matches[lost])))
print("Total match drawn = {0}".format(len(matches[draw])))
```

```
Total match wins = 19912
Total match lost = 11577
Total match drawn = 9456
```