

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv(r'C:\Users\user\Desktop\Phonicide_by_countries.csv')

In [4]:
df

Out [5]:
   Location  Region  Subregion  Rate  Count  Year
0  Afghanistan  Asia  Southern Asia  6.7   2474  2018
1  Albania      Europe  Southern Europe  2.1    61  2020
2  Algeria      Africa  Northern Africa  1.3   580  2020
3  Andorra      Europe  Southern Europe  2.6    2  2020
4  Angola       Africa  Middle Africa  4.8  1217  2012
...         ...     ...         ...     ...     ...
190 Venezuela  NIS America  South America  36.7  10598  2018
191 Vietnam    Asia  South-Eastern Asia  1.5   1358  2011
192 Yemen      Asia  Western Asia  6.8   1703  2013
193 Zambia     Africa  Eastern Africa  5.4   853  2015
194 Zimbabwe   Africa  Eastern Africa  7.5   981  2012

195 rows x 6 columns

In [6]:
df.isnull().sum()

Out [6]:
Location      0
Region        0
Subregion     0
Rate          0
Count         0
Year          0
dtype: int64

In [7]:
df.duplicated().any()

Out [7]:
False

In [8]:
df.columns

Out [8]:
Index(['Location', 'Region', 'Subregion', 'Rate', 'Count', 'Year'], dtype='object')

In [9]:
df['Region'].replace('Americas', 'N/S America', inplace=True)

In [10]:
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  --
 0   Location    195 non-null    object
 1   Region      195 non-null    object
 2   Subregion   195 non-null    object
 3   Rate        195 non-null    float64
 4   Count       195 non-null    int64
 5   Year        195 non-null    int64
dtypes: float64(1), int64(2), object(3)
memory usage: 9.3+ KB

In [11]:
df.nunique()

Out [11]:
Location      195
Region         5
Subregion     19
Rate          103
Count         151
Year          16
dtype: int64

In [12]:
df1 = df.sort_values(by = 'Count', ascending = False)

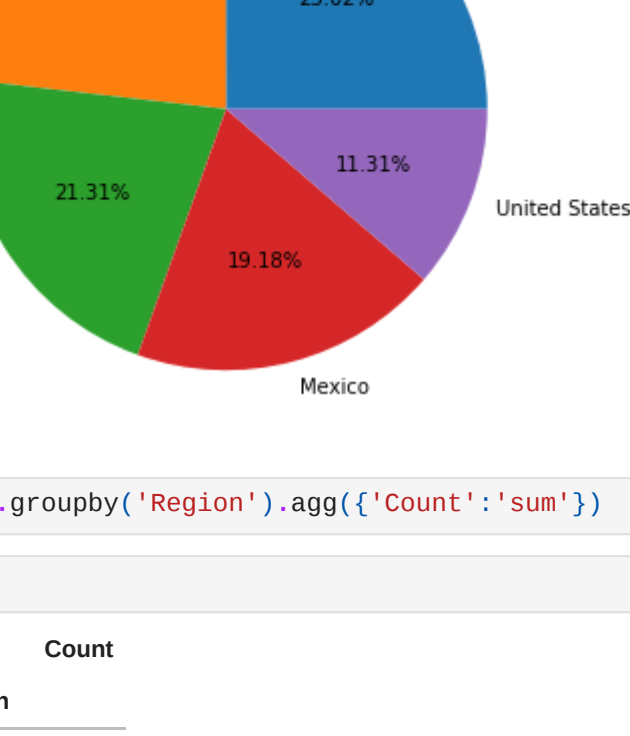
Out [12]:
   Location  Region  Subregion  Rate  Count  Year
26  Brazil    N/S America  South America  22.5  47722  2020
125 Nigeria   Africa  Western Africa  22.0  44200  2019
78  India     Asia  Southern Asia  3.0  40551  2020
111 Mexico    N/S America  Central America  28.4  36579  2020
186 United States  N/S America  Northern America  6.5  21570  2020
...     ...     ...         ...     ...     ...
73  Holy See   Europe  Southern Europe  0.0    0  2015
38  Channel Islands  Europe  Northern Europe  0.0    0  2010
145 Saint Helena  Africa  Western Africa  0.0    0  2009
151 San Marino   Europe  Southern Europe  0.0    0  2011
21  Bermuda     N/S America  Northern America  0.0    0  2019

195 rows x 6 columns

In [13]:
df1 = df[['Location', 'Count']].sort_values(by = 'Count', ascending=False).head()
df1

Out [13]:
   Location  Count
26  Brazil  47722
125 Nigeria 44200
78  India  40551
111 Mexico 36579
186 United States 21570

In [14]:
df1.plot(x= 'Location', y= 'Count', kind='pie', labels = df1.Location, autopct='%1.2f%%', figsize=(10,6))
plt.legend().set_visible(False)

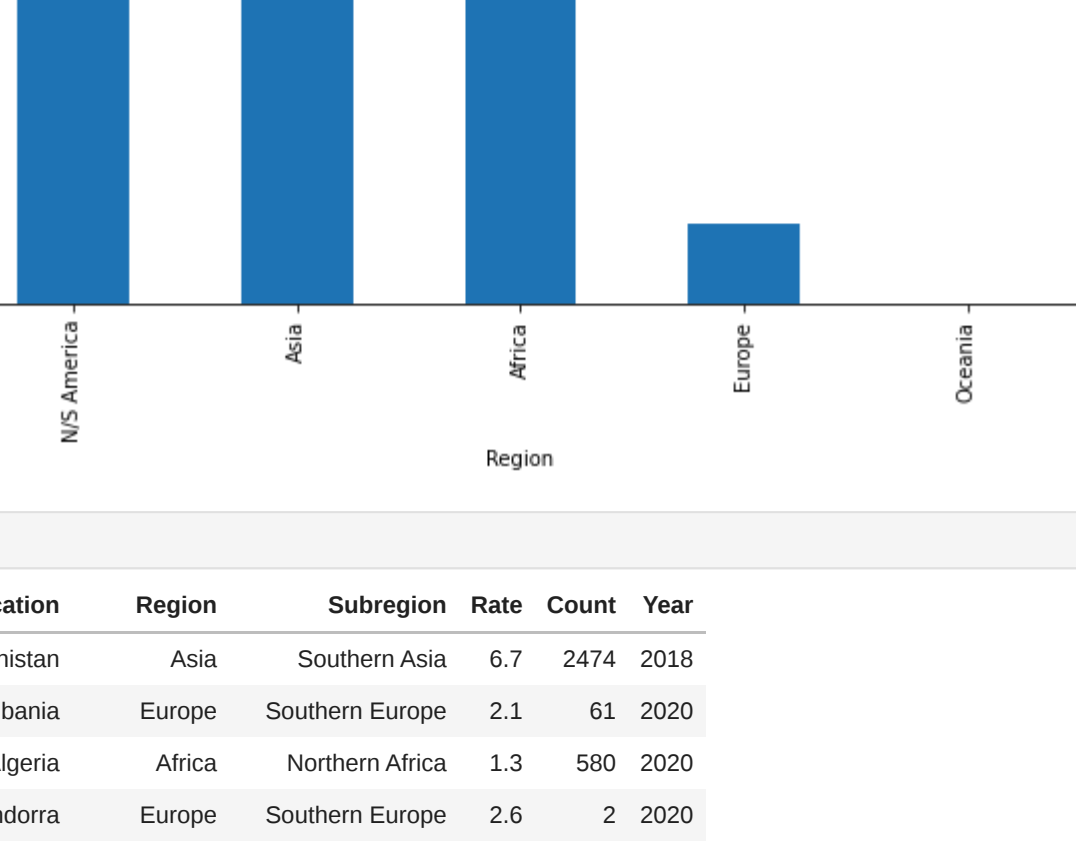


In [15]:
df2 = df.groupby('Region').agg({'Count':'sum'})

Out [16]:
df2
   Region  Count
Africa    99491
Asia     105552
Europe    18869
NIS America 153597
Oceania     347

In [17]:
df2.sort_values(by='Count', ascending=False, inplace=True)

In [18]:
df2.plot(kind='bar',figsize=(10,6))
plt.show()



In [19]:
df

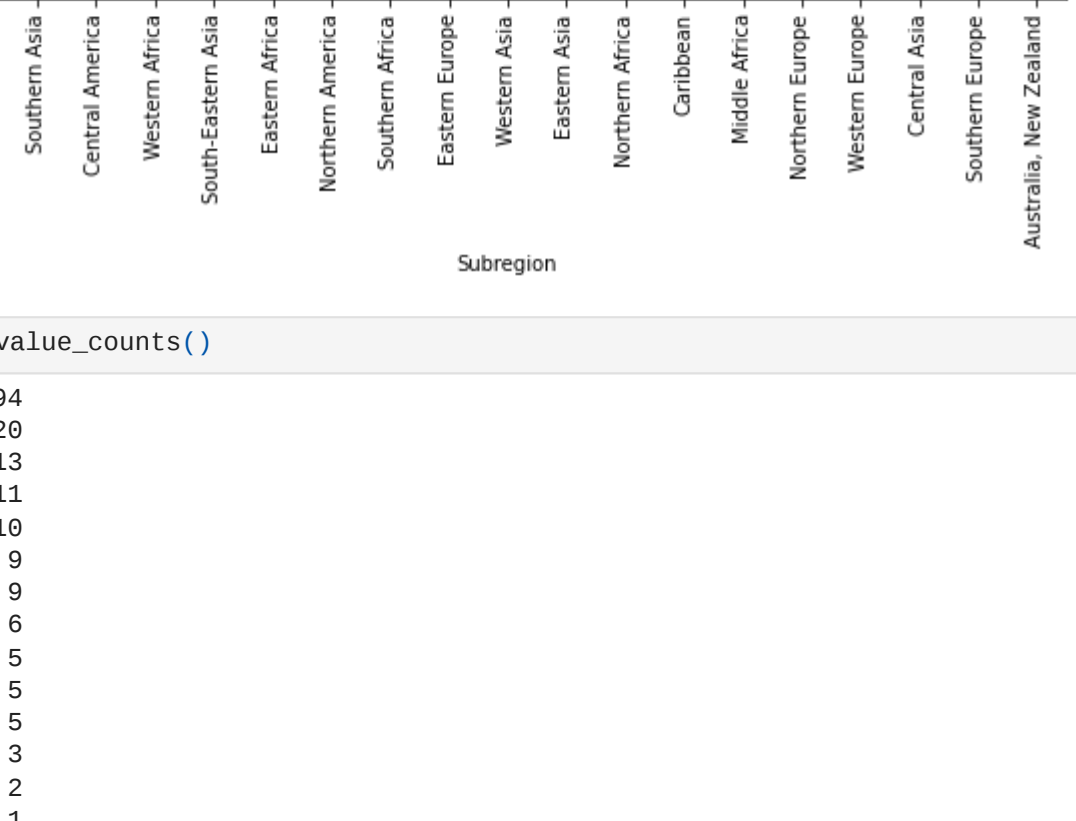
Out [19]:
   Location  Region  Subregion  Rate  Count  Year
0  Afghanistan  Asia  Southern Asia  6.7   2474  2018
1  Albania      Europe  Southern Europe  2.1    61  2020
2  Algeria      Africa  Northern Africa  1.3   580  2020
3  Andorra      Europe  Southern Europe  2.6    2  2020
4  Angola       Africa  Middle Africa  4.8  1217  2012
...         ...     ...         ...     ...     ...
190 Venezuela  NIS America  South America  36.7  10598  2018
191 Vietnam    Asia  South-Eastern Asia  1.5   1358  2011
192 Yemen      Asia  Western Asia  6.8   1703  2013
193 Zambia     Africa  Eastern Africa  5.4   853  2015
194 Zimbabwe   Africa  Eastern Africa  7.5   981  2012

195 rows x 6 columns

In [20]:
df3 = df.groupby('Subregion').agg({'Count':'sum'})
df3.sort_values(by='Count', ascending=False, inplace=True)
df3

Out [20]:
df3
   Subregion  Count
South America  78872
Southern Asia  58631
Central America  47371
Western Africa  46318
South-Eastern Asia  25314
Eastern Africa  23669
Northern America  22317
Southern Africa  21479
Eastern Europe  14604
Western Asia  11638
Eastern Asia  8563
Northern Africa  5538
Caribbean  5037
Middle Africa  2477
Northern Europe  2097
Western Europe  2075
Central Asia  1406
Southern Europe  1093
Australia, New Zealand  347

In [21]:
df3.plot(kind='bar',figsize=(10,6))
plt.show()



In [22]:
df4 = df.value_counts()

Out [22]:
Year  Count
2020   94
2019   28
2018   13
2017   11
2016   10
2015    9
2014    6
2013    5
2012    5
2011    3
2010    3
2009    2
2008    1
2007    1
2006    1
2005    1
Name: Year, dtype: int64

In [23]:
df4 = df[(df['Region']=='Asia') | (df['Region']=='Europe')]

In [24]:
df4

Out [24]:
   Location  Region  Subregion  Rate  Count  Year
0  Afghanistan  Asia  Southern Asia  6.7   2474  2018
1  Albania      Europe  Southern Europe  2.1    61  2020
3  Andorra      Europe  Southern Europe  2.6    2  2020
8  Armenia      Europe  Western Asia  1.8    52  2020
11 Austria      Europe  Western Europe  0.7    65  2020
...     ...     ...         ...     ...     ...
183 United Arab Emirates  Asia  Western Asia  0.7    64  2019
184 United Kingdom  Europe  Northern Europe  1.1   755  2018
189 Uzbekistan  Asia  Central Asia  1.2   396  2019
191 Vietnam    Asia  South-Eastern Asia  1.5   1358  2011
192 Yemen      Asia  Western Asia  6.8   1703  2013

102 rows x 6 columns

In [25]:
df4 = df4[(df4['Year'] >=2017)][['Region', 'Year', 'Count']]

In [26]:
df4 = df4.groupby(['Region', 'Year']).agg({'Count':'sum'})
df4

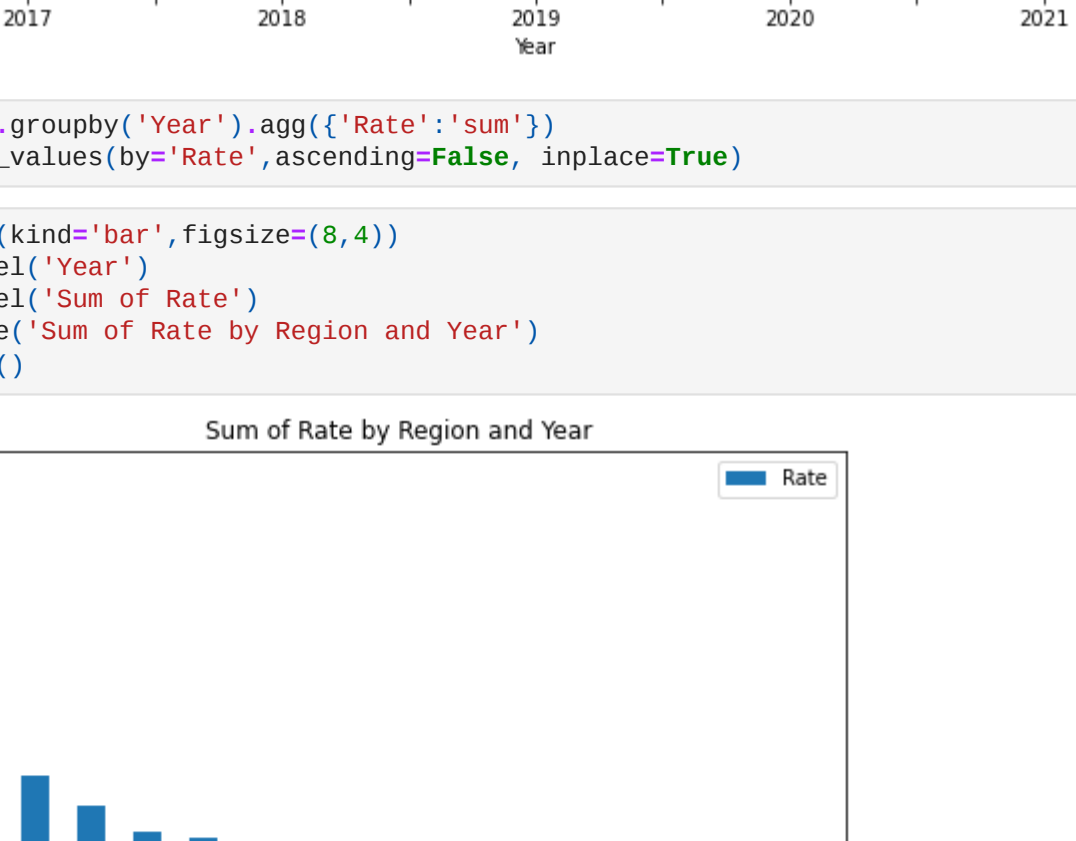
Out [26]:
df4
   Region  Year  Count
Asia  2017  1787
     2018  16923.0
     2019  6458
     2020  53516
     2021  15299
Europe 2017  3670
     2018  883
     2019  249
     2020  15066

In [27]:
df_unstacked = df4.unstack(level=0)
df_unstacked

Out [27]:
df_unstacked
   Region  Year  Count
Asia  2017  1787.0  3670.0
     2018  16923.0  883.0
     2019  6458.0  249.0
     2020  53516.0  15066.0
Europe 2017  3670.0  NaN
     2018  883.0  NaN
     2019  249.0  NaN
     2020  15299.0  NaN

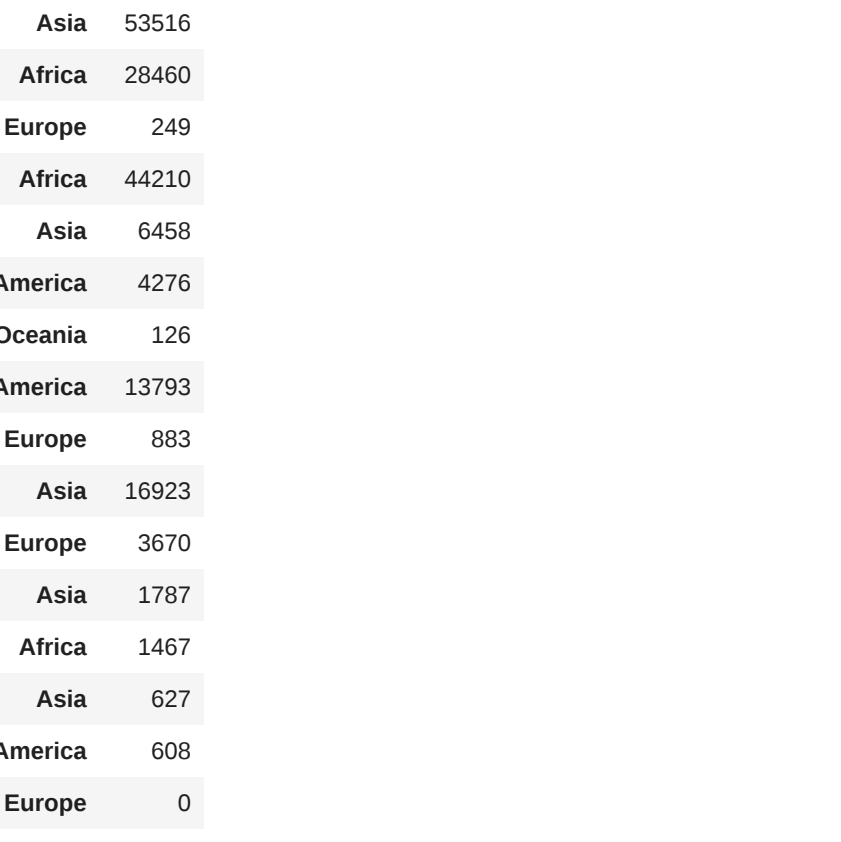
In [28]:
df_unstacked.index = df_unstacked.index.astype(int).astype(str)

In [29]:
df_unstacked.plot(figsize=(10,6))
plt.xlabel('Year')
plt.ylabel('Count')
plt.title('Count of Asia and Europe over the years')
plt.show()



In [30]:
df5 = df.groupby('Year').agg({'Rate':'sum'})
df5.sort_values(by='Rate', ascending=False, inplace=True)

In [31]:
df5.plot(kind='bar',figsize=(8,4))
plt.xlabel('Year')
plt.ylabel('Sum of Rate')
plt.title('Sum of Rate by Region and Year')
plt.show()



In [32]:
df6 = df.groupby(['Year', 'Region']).agg({'Count':'sum'})
df6.sort_values(by='Year', ascending=False, inplace=True)
df6

Out [32]:
df6
   Year  Region  Count
2021  Asia  15299
2020  Oceania  221
     NIS America  134784
     Europe  15066
     Asia  53516
     Africa  28460
2019  Europe  249
     Africa  44210
     Asia  6458
     NIS America  4278
     Oceania  126
2018  NIS America  13793
     Europe  883
     Asia  16823
2017  Europe  3670
     Asia  1787
     Africa  1467
2016  Asia  827
     NIS America  608
     Europe  0
    
```