| In [3]: In [4]: | <pre>import pandas as pd import seaborn as sns import matplotlib.pyplot as plt df = pd.read_csv(r'C:\Users\user\Desktop\P\homicide_by_countries.csv')</pre> |
|---------------------------------|--|
| In [5]: Out[5]: | LocationRegionSubregionRateCountYear0 AfghanistanAsiaSouthern Asia6.7247420181 AlbaniaEuropeSouthern Europe2.16120202 AlgeriaAfricaNorthern Africa1.35802020 |
| | 3 Andorra Europe Southern Europe 2.6 2 2020 4 Angola Africa Middle Africa 4.8 1217 2012 190 Venezuela Americas 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 |
| In [6]: | 193 Zambia Africa Eastern Africa 5.4 853 2015 194 Zimbabwe Africa Eastern Africa 7.5 981 2012 195 rows × 6 columns df.isnull().sum() |
| Out[6]: In [7]: | Location 0 Region 0 Subregion 0 Rate 0 Count 0 Year 0 dtype: int64 df.duplicated().any() |
| Out[7]: In [8]: Out[8]: In [9]: | df.columns Index(['Location', 'Region', 'Subregion', 'Rate', 'Count', 'Year'], dtype='object') df['Region'].replace('Americas','N/S America',inplace=True) |
| In [10]: | <pre>df.info() <class 'pandas.core.frame.dataframe'=""> RangeIndex: 195 entries, 0 to 194 Data columns (total 6 columns): # Column Non-Null Count Dtype</class></pre> |
| In [11]: | 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 df.nunique() |
| Out[11]: In [12]: | Location 195 Region 5 Subregion 19 Rate 103 Count 151 Year 16 dtype: int64 df1 = df.sort_values(by = 'Count', ascending =False) |
| Out[12]: | Location Region Subregion Rate Count Year |
| | 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 |
| In [13]: | 151 San Marino Europe Southern Europe 0.0 0 2011 21 Bermuda N/S America Northern America 0.0 0 2019 195 rows × 6 columns df1 = df[['Location', 'Count']].sort_values(by = 'Count', ascending=False).head() df1 |
| Out[13]: | Location Count Brazil 47722 125 Nigeria 44200 78 India 40651 111 Mexico 36579 186 United States 21570 |
| In [14]: | <pre>df1.plot(x= 'Location', y= 'Count', kind='pie', labels = df1.Location, autopct='%1.2f%%', figsize=(10,6)) plt.legend().set_visible(False)</pre> <pre>Nigeria</pre> <pre>Brazil</pre> |
| | 23.18% 25.02% 11.31% United States |
| In [15]: | India 19.18% Mexico df2 = df.groupby('Region').agg({'Count':'sum'}) |
| <pre>In [16]: Out[16]:</pre> | |
| In [17]: In [18]: | <pre>Europe 19869 N/S America 153597 Oceania 347 df2.sort_values(by='Count', ascending=False,inplace=True) df2.plot(kind='bar',figsize=(10,6))</pre> |
| L | plt.show() 160000 - Count 140000 - 120000 - |
| | 100000 - 80000 - 60000 - 40000 - |
| | N/S America Oceania Region |
| In [19]: Out[19]: | |
| | 3 Andorra Europe Southern Europe 2.6 2 2020 4 Angola Africa Middle Africa 4.8 1217 2012 190 Venezuela N/S 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 |
| In [20]: | 193 Zambia Africa Eastern Africa 5.4 853 2015 194 Zimbabwe Africa Eastern Africa 7.5 981 2012 195 rows × 6 columns df3 = df.groupby('Subregion').agg({'Count':'sum'}) df3.sort_values(by='Count', ascending=False, inplace=True) |
| Out[20]: | Count Subregion 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() 80000 - Count 60000 - Count |
| | 50000 - 40000 - 30000 - |
| | South America - Southern Asia - Central America - Western Africa - Northern America - Southern Africa - Northern Africa - Middle Africa - Central Asia - Central Asia - Southern Europe - Western Europe - Central Asia - Southern Europe - Southern Europe - America - Northern Africa - Southern Furope - Central Asia - Southern Europe - Central Asia - Central Asia - Southern Europe - Central Asia - Cen |
| In [22]: Out[22]: | Subregion df.Year.value_counts() 2020 94 2019 20 2018 13 |
| | 2012 11 2017 10 2016 9 2015 9 2009 6 2014 5 2013 5 2011 5 2010 3 |
| In [23]: In [24]: | 2006 |
| 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 Asia Western Asia 1.8 52 2020 11 Austria Europe Western Europe 0.7 65 2020 |
| | Image: Composition of the co |
| In [25]: In [26]: | 102 rows × 6 columns df4= df4[(df4['Year'] >=2017)][['Region', 'Year', 'Count']] df4 = df4.groupby(['Region', 'Year']).agg({'Count': 'sum'}) df4 |
| Out[26]: | Count Region Year Asia 2017 1787 2018 16923 2019 6458 2020 53516 |
| | 2021 15299 Europe 2017 3670 2018 883 2019 249 2020 15066 |
| In [27]: Out[27]: | df_unstacked = df4.unstack(level=0) df_unstacked Count Region Asia Europe Year 2017 1787.0 3670.0 |
| In [28]: | 2018 16923.0 883.0 2019 6458.0 249.0 2020 53516.0 15066.0 2021 15299.0 NaN 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() Count of Asia and Europe over the Years None,Region |
| | 50000 - (Count, Asia) (Count, Europe) |
| | 20000 |
| In [30]: In [31]: | df5 = df.groupby('Year').agg({'Rate':'sum'}) df5.sort_values(by='Rate', ascending=False, inplace=True) 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() Sum of Rate by Region and Year 600 - Rate 500 - Rate |
| | 400 - 100 - 100 - |
| In [32]: Out[32]: | df6 = df.groupby(['Year', 'Region']).agg({'Count':'sum'}) df6.sort_values(by='Year', ascending=False, inplace=True) df6 Count |
| 1• | Year Region 2021 Asia 15299 2020 Oceania 221 N/S America 134784 Europe 15066 Asia 53516 |
| | Asia 53516 Africa 28460 2019 Europe 249 Africa 44210 Asia 6458 N/S America 4276 Oceania 126 |
| | Oceania 126 2018 N/S America 13793 Europe 883 Asia 16923 2017 Europe 3670 Asia 1787 Africa 1467 |
| | 2016 Asia 627 N/S America 608 Europe 0 Africa 4987 2015 Asia 241 |
| | Europe 0 2014 N/S America 13 Asia 3029 2013 Asia 5782 2012 Africa 15264 |
| | N/S America 53 Asia 61 2011 Asia 1626 Africa 855 Europe 0 2010 Europe 1 |
| | 2009 |
| In [33]: | df6.plot(kind='bar',figsize=(12,8)) plt.xlabel('Year, Region') plt.ylabel('Sum of Count') plt.title('Sum of Count by Region and Year') plt.show() Sum of Count by Region and Year |
| | 140000 - Count 120000 - 100000 - |
| | 80000 - 60000 - 60000 - |
| | 20000 - |
| In [34]: | (2021, Asia) (2020, Oceania) (2020, Oceania) (2020, Oceania) (2020, Oceania) (2020, Africa) (2020, Africa) (2019, Europe) (2019, Europe) (2019, Africa) (2019, Africa) (2019, Africa) (2019, Africa) (2011, Europe) (2011, Africa) (2010, Africa) (201 |
| In [35]: Out[35]: | df7.sort_values(by='Count',ascending=False,inplace=True) df7.reset_index() Subregion Count Southern Asia 6514.56 South America 6067.08 |
| | Central America 5921.38 Northern America 4463.40 Southern Africa 4295.80 Western Africa 4210.73 South-Eastern Asia 2531.40 Eastern Africa 1577.93 |
| | 8 Eastern Europe 1460.40 9 Eastern Asia 1223.29 10 Northern Africa 1107.60 11 Middle Africa 619.25 12 Western Asia 581.90 |
| | 13 Central Asia 281.20 14 Western Europe 230.56 15 Caribbean 201.48 16 Australia, New Zealand 173.50 17 Northern Europe 139.80 18 Southern Europe 64.29 |
| In [36]: In [38]: | <pre>import plotly.express as px fig = px.treemap(df, path=['Subregion'], values='Count', title='Treemap') fig.show() C:\ProgramData\Anaconda3\lib\site-packages\plotly\express_core.py:1637: FutureWarning:</pre> |
| | The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead. Treemap |
| | South America Southern Asia Central America South-Eastern Asia Northern America Southern Africa |
| | Western Africa Western Africa Western Africa Western Africa Western Africa Western Africa Western Asia Western Asia Western Asia Toolhen Europe Caribbean Northern Africa Western Asia Countral Asia Footners Europe Central Asia |
| In []: | |
| | |