In [30]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\merged-csv-files.csv",low_memory=False,
df
```

Out[30]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.5	69.49	5.04
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94
...	...	...	...	...	...	...	...	...	...	...	...
712519	84070017	US	USA	840	NaN	Southeast Utah	Utah	US	38.99617072	-110.70139579999999	Southeast Utah, Utah, US
712520	84070018	US	USA	840	NaN	Southwest Utah	Utah	US	37.85447192	-111.4418764	Southwest Utah, Utah, US
712521	84070019	US	USA	840	NaN	TriCounty	Utah	US	40.12491499	-109.5174415	TriCounty, Utah, US
712522	84070020	US	USA	840	NaN	Weber-Morgan	Utah	US	41.27116049	-111.9145117	Weber-Morgan, Utah, US
712523		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

712524 rows × 15 columns

In [31]:
```
df.head()
```

Out[31]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 inc
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.5	69.49	5.04	35526	737	
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709	
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282	
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23	
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201	

In [32]:
```
df.tail()
```

Out[32]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered
712519	84070017	US	USA	840	NaN	Southeast Utah	Utah	US	38.99617072	-110.70139579999999	Southeast Utah, Utah, US
712520	84070018	US	USA	840	NaN	Southwest Utah	Utah	US	37.85447192	-111.4418764	Southwest Utah, Utah, US
712521	84070019	US	USA	840	NaN	TriCounty	Utah	US	40.12491499	-109.5174415	TriCounty, Utah, US
712522	84070020	US	USA	840	NaN	Weber-Morgan	Utah	US	41.27116049	-111.9145117	Weber-Morgan, Utah, US
712523		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [33]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 712524 entries, 0 to 712523
Data columns (total 15 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Country/Region         678120 non-null  object
 1   Confirmed              712523 non-null  object
 2   Deaths                 712523 non-null  object
 3   Recovered              712523 non-null  object
 4   Active                 710643 non-null  object
 5   New cases              711395 non-null  object
 6   New deaths             712523 non-null  object
 7   New recovered          712523 non-null  object
 8   Deaths / 100 Cases     712523 non-null  object
 9   Recovered / 100 Cases  712523 non-null  object
 10  Deaths / 100 Recovered 628297 non-null  object
 11  Confirmed last week    628297 non-null  object
 12  1 week change          628108 non-null  object
 13  1 week % increase      628108 non-null  object
 14  WHO Region             187 non-null     object
dtypes: object(15)
memory usage: 81.5+ MB
```

In [34]:
```
df.describe()
```

Out[34]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 w cha
count	678120	712523	712523	712523	710643	711395	712523	712523	712523	712523	628297	628297	628
unique	3796	567	11168	4140	11428	18098	5140	7788	12782	3595	3673	421	11
top	Greenland	US	USA	840	0	0	Texas	US	0	Europe	Greene, Mississippi, US	7/27/20	
freq	189	612504	612128	612128	11650	19129	48128	627920	34590	25568	188	3340	253

In [35]:
```
df.isnull().sum()
```

Out[35]:
```
Country/Region             34404
Confirmed                      1
Deaths                         1
Recovered                      1
Active                      1881
New cases                   1129
New deaths                     1
New recovered                  1
Deaths / 100 Cases             1
Recovered / 100 Cases          1
Deaths / 100 Recovered     84227
Confirmed last week        84227
1 week change              84416
1 week % increase          84416
WHO Region                712337
dtype: int64
```

In [36]:
```
df.dropna()
```

Out[36]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.5	69.49	5.04	35526	737
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201
...	...	...	...	...	...	...	...	...	...	...	...	...	...
182	West Bank and Gaza	10621	78	3752	6791	152	2	0	0.73	35.33	2.08	8916	1705
183	Western Sahara	10	1	8	1	0	0	0	10	80	12.5	10	0
184	Yemen	1691	483	833	375	10	4	36	28.56	49.26	57.98	1619	72
185	Zambia	4552	140	2815	1597	71	1	465	3.08	61.84	4.97	3326	1226
186	Zimbabwe	2704	36	542	2126	192	2	24	1.33	20.04	6.64	1713	991

187 rows × 15 columns

In [38]:
```
df.shape
```

Out[38]:
```
(712524, 15)
```

In [40]:
```
df1 = df[df.isna().any(axis=1)]
df1
```

Out[40]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Dea Recov
187	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO Region	
188	NaN	Afghanistan	33.93911	67.709953	2020-01-22	0	0	0	0	Eastern Mediterranean	
189	NaN	Albania	41.1533	20.1683	2020-01-22	0	0	0	0	Europe	
190	NaN	Algeria	28.0339	1.6596	2020-01-22	0	0	0	0	Africa	
191	NaN	Andorra	42.5063	1.5218	2020-01-22	0	0	0	0	Europe	
...	...	...	...	...	...	...	...	...	...	...	
712519	84070017	US	USA	840	NaN	Southeast Utah	Utah	US	38.99617072	-110.70139579999999	Soutl Utah,
712520	84070018	US	USA	840	NaN	Southwest Utah	Utah	US	37.85447192	-111.4418764	Soutl Utah,
712521	84070019	US	USA	840	NaN	TriCounty	Utah	US	40.12491499	-109.5174415	TriCo Utal
712522	84070020	US	USA	840	NaN	Weber-Morgan	Utah	US	41.27116049	-111.9145117	W Mo Utal
712523		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

712337 rows × 15 columns

In [44]:
```
df.rename(columns={"ISO 3166-1 alpha-3 CODE": "Iso_Code"},
inplace=True)
df
```

Out[44]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.5	69.49	5.04
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94
...	...	...	...	...	...	...	...	...	...	...	...
712519	84070017	US	USA	840	NaN	Southeast Utah	Utah	US	38.99617072	-110.70139579999999	Southeast Utah, Utah, US
712520	84070018	US	USA	840	NaN	Southwest Utah	Utah	US	37.85447192	-111.4418764	Southwest Utah, Utah, US
712521	84070019	US	USA	840	NaN	TriCounty	Utah	US	40.12491499	-109.5174415	TriCounty, Utah, US
712522	84070020	US	USA	840	NaN	Weber-Morgan	Utah	US	41.27116049	-111.9145117	Weber-Morgan, Utah, US
712523		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

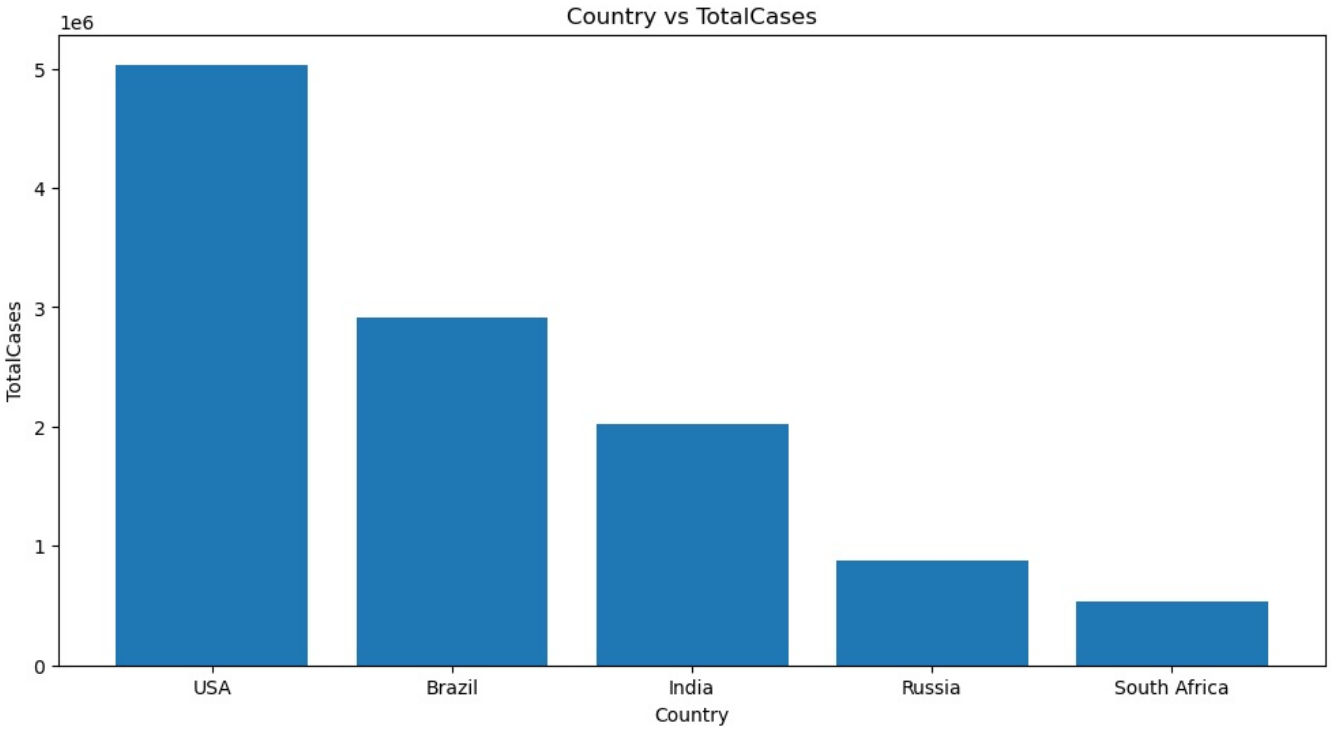
712524 rows × 15 columns

In [54]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nlargest(5, 'TotalCases')
Country.head()
```

Out[54]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	S
0	USA	North America	3.311981e+08	5032179	NaN	162804.0	NaN	2576668.0	NaN	2292707.0	
1	Brazil	South America	2.127107e+08	2917562	NaN	98644.0	NaN	2047660.0	NaN	771258.0	
2	India	Asia	1.381345e+09	2025409	NaN	41638.0	NaN	1377384.0	NaN	606387.0	
3	Russia	Europe	1.459409e+08	871894	NaN	14606.0	NaN	676357.0	NaN	180931.0	
4	South Africa	Africa	5.938157e+07	538184	NaN	9604.0	NaN	387316.0	NaN	141264.0	

In [63]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nlargest(5, 'TotalCases')
Country.head()
plt.figure(figsize =(12,6))
plt.bar(x=Country["Country/Region"],height=Country["TotalCases"])
plt.xlabel("Country")
plt.ylabel("TotalCases")
plt.title("Country vs TotalCases")
#plt.xticks(rotation = 90)
plt.show()
```



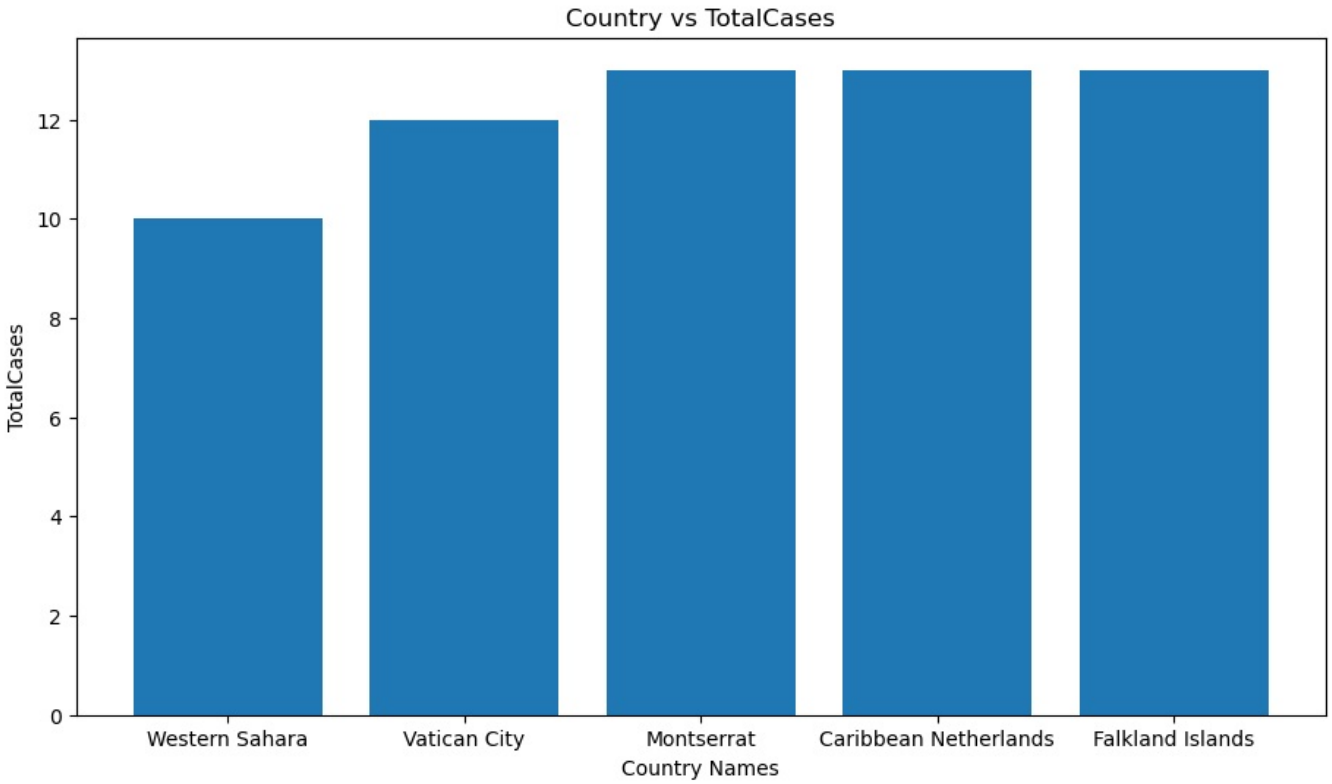
In [65]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nsmallest(5, 'TotalCases')
Country.head()
```

Out[65]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	S
208	Western Sahara	Africa	598682.0	10	NaN	1.0	NaN	8.0	NaN	1.0	
207	Vatican City	Europe	801.0	12	NaN	NaN	NaN	12.0	NaN	0.0	
204	Montserrat	North America	4992.0	13	NaN	1.0	NaN	10.0	NaN	2.0	
205	Caribbean Netherlands	North America	26247.0	13	NaN	NaN	NaN	7.0	NaN	6.0	
206	Falkland Islands	South America	3489.0	13	NaN	NaN	NaN	13.0	NaN	0.0	

In [74]:
```
import numpy as np
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nsmallest(5, 'TotalCases')
Country.head()
plt.figure(figsize =(11,6))
plt.bar(x=Country["Country/Region"],height=Country["TotalCases"])
plt.xlabel("Country Names")
plt.ylabel("TotalCases")
plt.title("Country vs TotalCases")
#plt.xticks(rotation = 90)
plt.show()
```

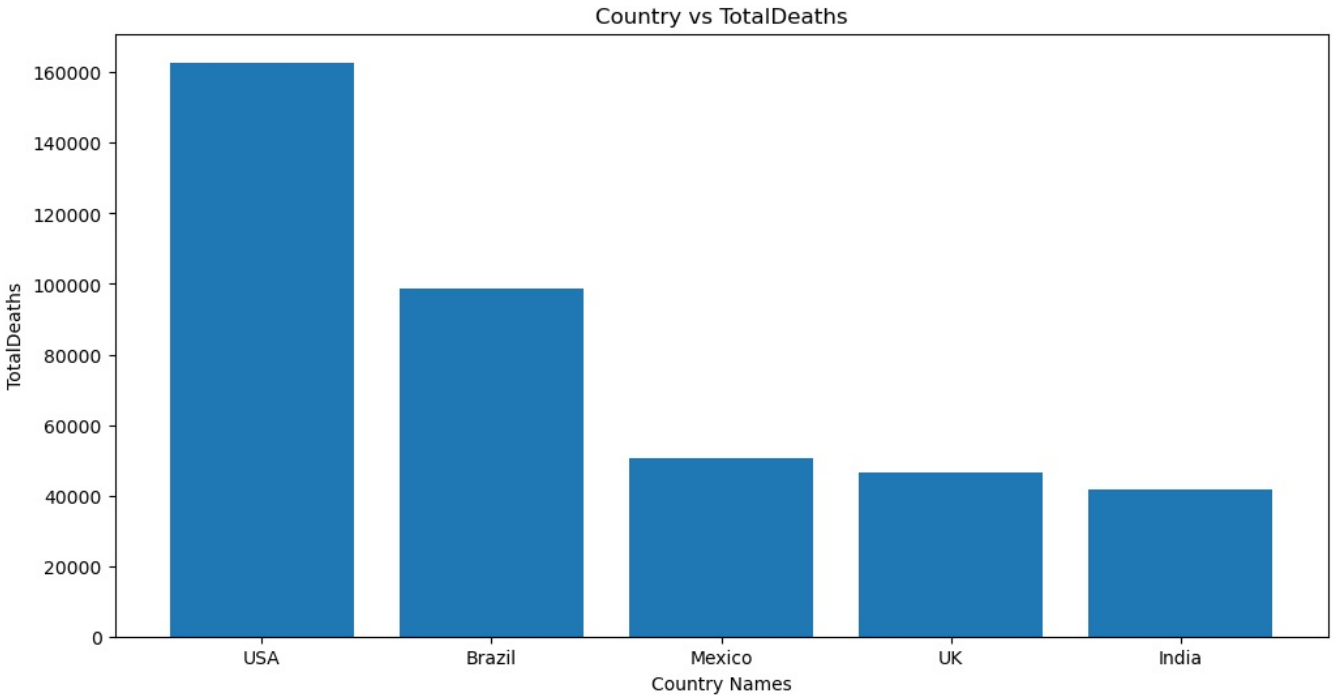


In [76]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nlargest(5, 'TotalDeaths')
Country.head()
```

Out[76]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases
0	USA	North America	3.311981e+08	5032179	NaN	162804.0	NaN	2576668.0	NaN	2292707.0
1	Brazil	South America	2.127107e+08	2917562	NaN	98644.0	NaN	2047660.0	NaN	771258.0
5	Mexico	North America	1.290662e+08	462690	6590.0	50517.0	819.0	308848.0	4140.0	103325.0
11	UK	Europe	6.792203e+07	308134	NaN	46413.0	NaN	NaN	NaN	NaN
2	India	Asia	1.381345e+09	2025409	NaN	41638.0	NaN	1377384.0	NaN	606387.0

In [77]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
Country = df.nlargest(5, 'TotalDeaths')
Country.head()
plt.figure(figsize =(12,6))
plt.bar(x=Country["Country/Region"],height=Country["TotalDeaths"])
plt.xlabel("Country Names")
plt.ylabel("TotalDeaths")
plt.title("Country vs TotalDeaths")
#plt.xticks(rotation = 90)
plt.show()
```

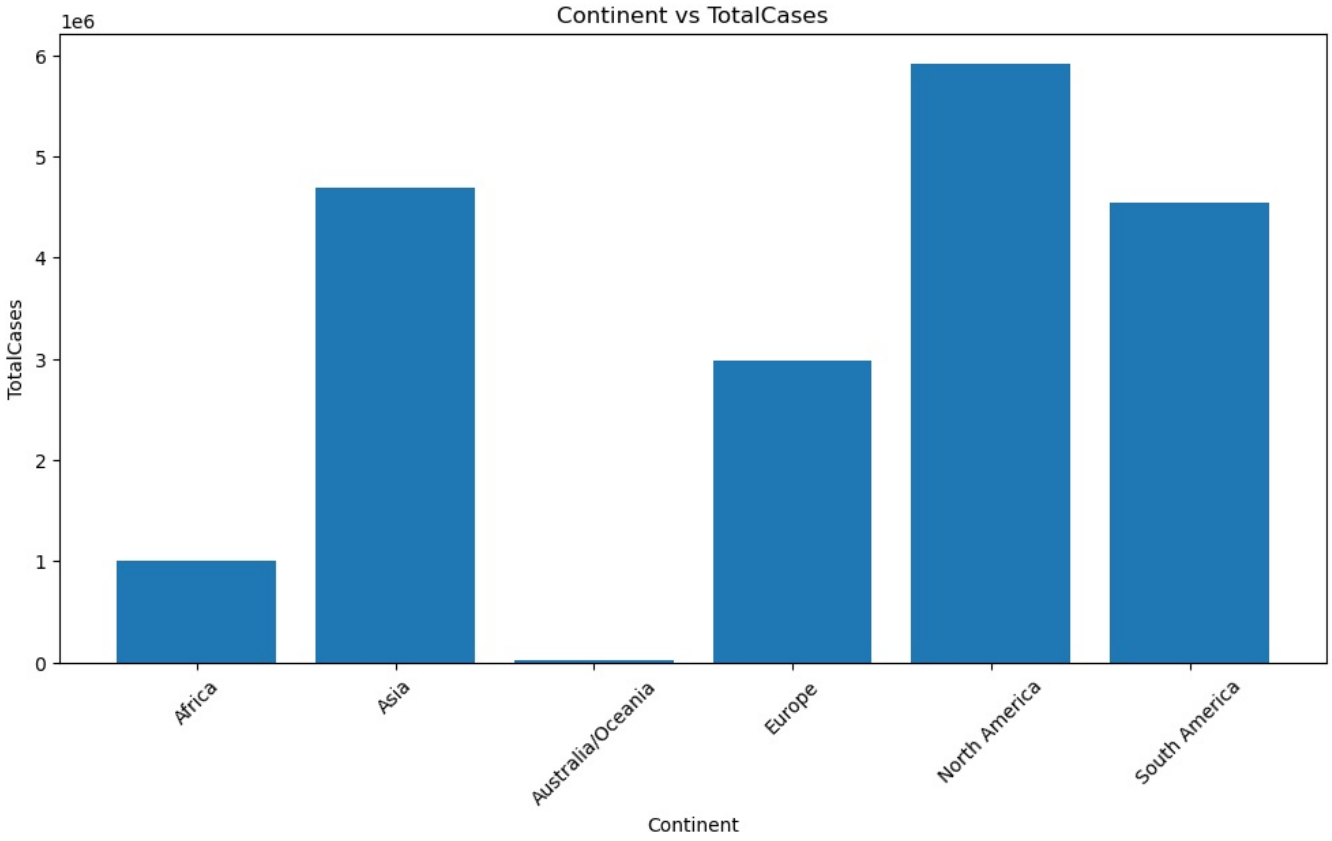


In [79]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
df2 = df.groupby('Continent')['TotalCases'].sum().reset_index()
df2
```

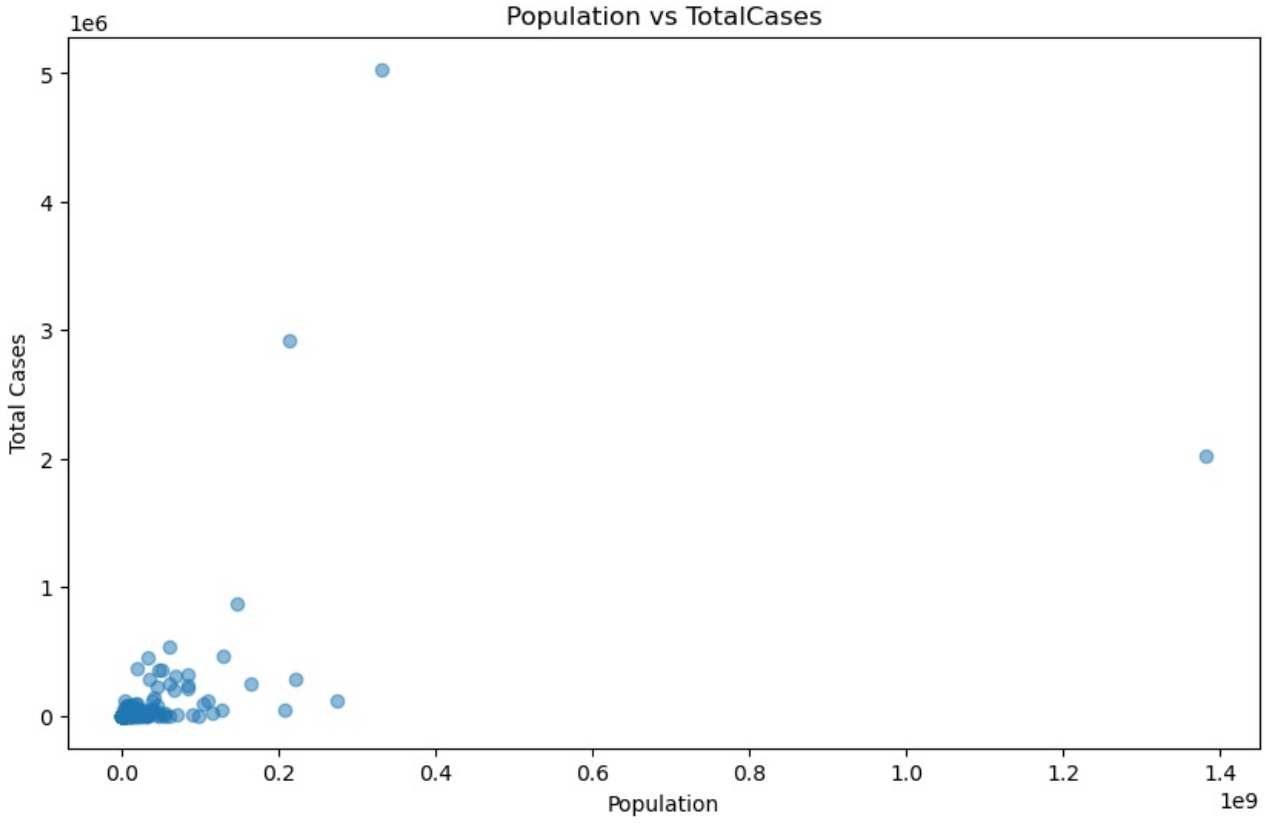
Out[79]:

	Continent	TotalCases
0	Africa	1011867
1	Asia	4689794
2	Australia/Oceania	21735
3	Europe	2982576
4	North America	5919209
5	South America	4543273

In [80]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\nandini sharma\Desktop\corona virus project\worldometer_data.csv")
df2 = df.groupby('Continent')['TotalCases'].sum().reset_index()
df2
plt.figure(figsize =(12,6))
plt.bar(x=df2['Continent'],height=df2['TotalCases'])
plt.xlabel("Continent")
plt.ylabel("TotalCases")
plt.title("Continent vs TotalCases")
plt.xticks(rotation = 45)
plt.show()
```



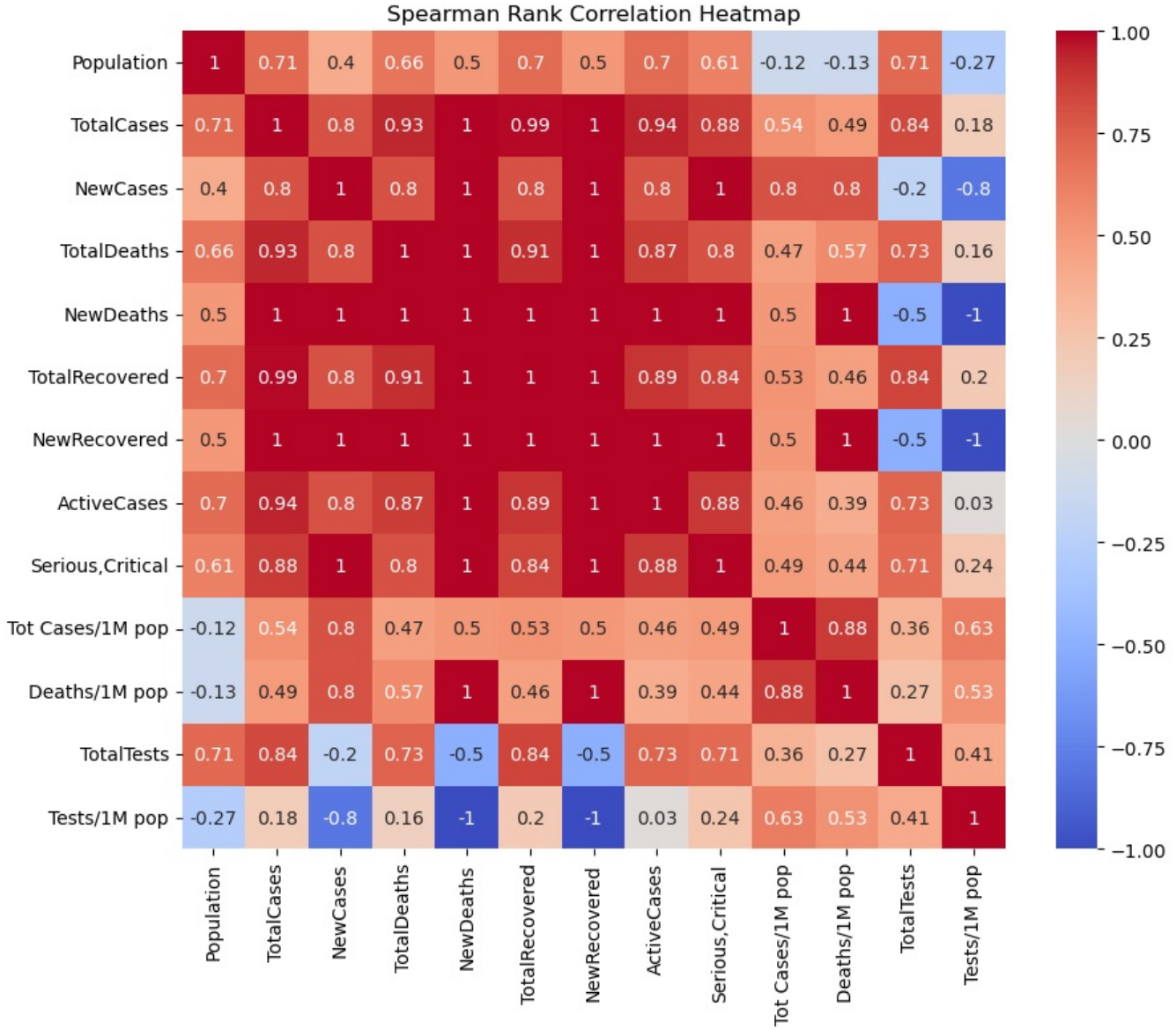
In [82]:
```
plt.figure(figsize=(10, 6))
# Create a scatter plot
plt.scatter(df["Population"], df["TotalCases"], alpha=0.5)
# Set labels and title
plt.xlabel("Population")
plt.ylabel("Total Cases")
plt.title("Population vs TotalCases")
# Show the plot
plt.show()
```



In [86]:
```
# Calculate the Spearman rank correlation matrix
correlation_matrix = df.corr(method='spearman')
# Set up the figure size
plt.figure(figsize=(10, 8))
# Create the heatmap
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", center=0)
# Set title
plt.title("Spearman Rank Correlation Heatmap")
```

```
# Show the plot
plt.show()
```

```
C:\Users\nandini sharma\AppData\Local\Temp\ipykernel_2088\2106339403.py:2: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid
columns or specify the value of numeric_only to silence this warning.
  correlation_matrix = df.corr(method='spearman')
```



In [ ]: