

Covid Data from Worldometer

In [31]:

#libraries

import pandas **as** pd

import numpy **as** np

import seaborn **as** sns

import matplotlib.pyplot **as** plt

df = pd.read_csv("covidReport_clean.csv")

In [32]:

df

Out[32]:

	CountryOther	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	!
0	USA	111,820,082	NaN	1,219,487	NaN	109,814,428	NaN	786,167	
1	India	45,035,393	NaN	533,570	NaN	NaN	NaN	NaN	
2	France	40,138,560	NaN	167,642	NaN	39,970,918	NaN	0	
3	Germany	38,828,995	NaN	183,027	NaN	38,240,600	NaN	405,368	
4	Brazil	38,743,918	NaN	711,380	NaN	36,249,161	NaN	1,783,377	
...	
234	Total:	70,200,879	NaN	1,367,332	NaN	66,683,585	NaN	2,149,962	
235	Total:	14,895,771	NaN	33,015	NaN	14,752,388	NaN	110,368	
236	Total:	12,860,924	NaN	258,892	NaN	12,090,808	NaN	511,224	
237	Total:	721	NaN	15	NaN	706	NaN	0	
238	Total:	704,753,890	0.0	7,010,681	0.0	675,619,811	790.0	22,123,398	

239 rows × 21 columns

In [33]:

#check if any row is not a country using pattern: only alphabetic or spaces allowed

pattern = r'^[A-Za-z .-]+\$'

not_country = df[~df['CountryOther'].str.match(pattern, na=False)]

not_country

Out[33]:

	CountryOther	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	!
231	Total:	131,889,132	NaN	1,695,941	NaN	127,665,129	NaN	2,528,062	
232	Total:	221,500,265	NaN	1,553,662	NaN	205,673,091	NaN	14,273,512	
233	Total:	253,406,198	NaN	2,101,824	NaN	248,754,104	NaN	2,550,270	
234	Total:	70,200,879	NaN	1,367,332	NaN	66,683,585	NaN	2,149,962	
235	Total:	14,895,771	NaN	33,015	NaN	14,752,388	NaN	110,368	
236	Total:	12,860,924	NaN	258,892	NaN	12,090,808	NaN	511,224	
237	Total:	721	NaN	15	NaN	706	NaN	0	
238	Total:	704,753,890	0.0	7,010,681	0.0	675,619,811	790.0	22,123,398	

8 rows × 21 columns

In [34]:

#only country data

df = df[df['CountryOther'].str.match(pattern, na=False)]

df

Out[34]:

	CountryOther	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	1
0	USA	111,820,082	NaN	1,219,487	NaN	109,814,428	NaN	786,167	
1	India	45,035,393	NaN	533,570	NaN	NaN	NaN	NaN	
2	France	40,138,560	NaN	167,642	NaN	39,970,918	NaN	0	
3	Germany	38,828,995	NaN	183,027	NaN	38,240,600	NaN	405,368	
4	Brazil	38,743,918	NaN	711,380	NaN	36,249,161	NaN	1,783,377	
...	
226	Tokelau	80	NaN	NaN	NaN	NaN	NaN	80	
227	Vatican City	29	NaN	NaN	NaN	29	NaN	0	
228	Western Sahara	10	NaN	1	NaN	9	NaN	0	
229	MS Zaandam	9	NaN	2	NaN	7	NaN	0	
230	China	503,302	NaN	5,272	NaN	379,053	NaN	118,977	

231 rows × 21 columns

```
In [37]:
df['TotalCases'] = df['TotalCases'].astype('str').str.replace(',','')
df['TotalCases'] = pd.to_numeric(df['TotalCases'], errors = 'coerce')
df = df[df['TotalCases'].notna() & (df['TotalCases'] > 0)]
df
```

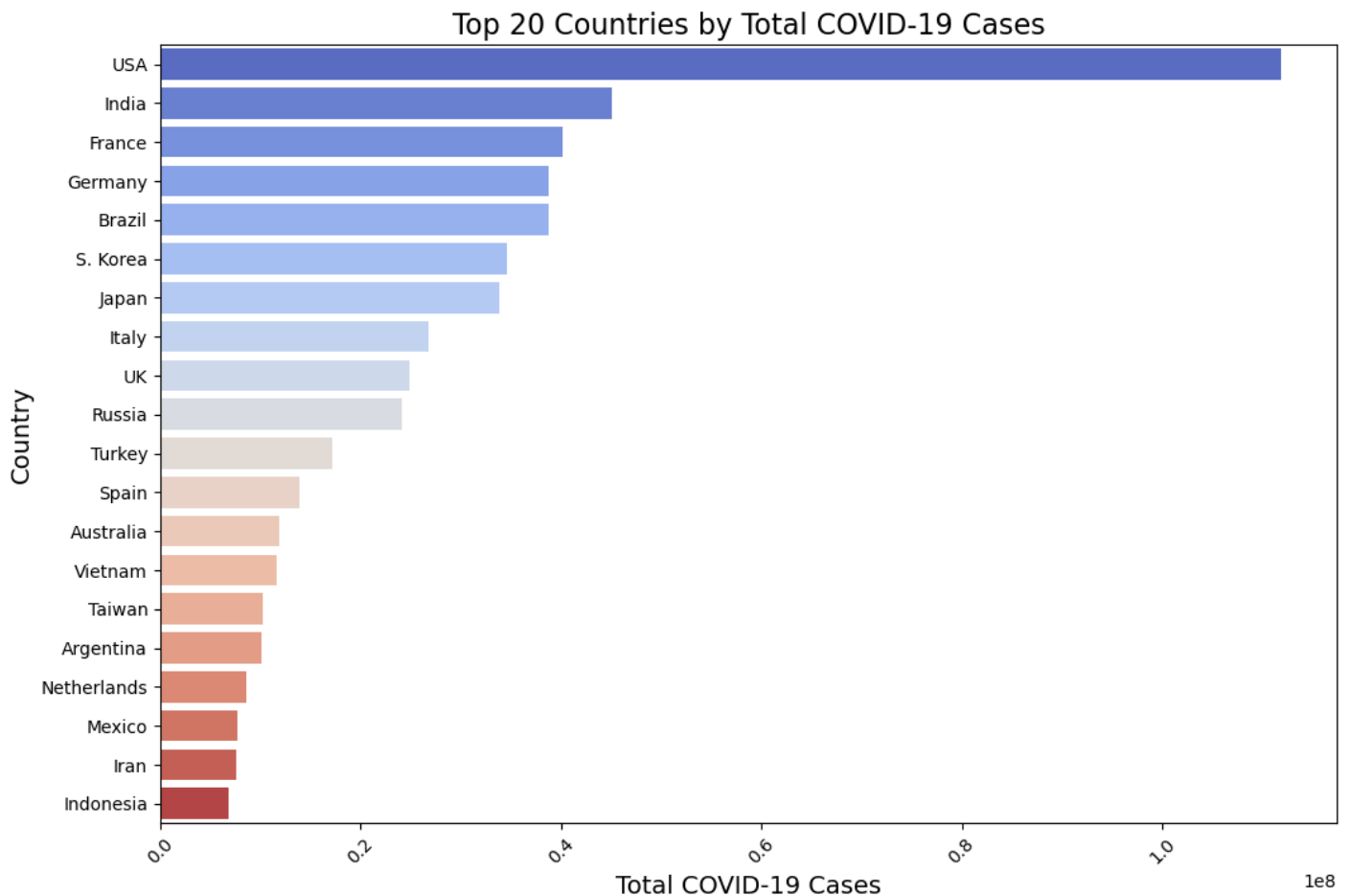
Out[37]:

	CountryOther	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	S
0	USA	111820082	NaN	1,219,487	NaN	109,814,428	NaN	786,167	
1	India	45035393	NaN	533,570	NaN	NaN	NaN	NaN	
2	France	40138560	NaN	167,642	NaN	39,970,918	NaN	0	
3	Germany	38828995	NaN	183,027	NaN	38,240,600	NaN	405,368	
4	Brazil	38743918	NaN	711,380	NaN	36,249,161	NaN	1,783,377	
...	
226	Tokelau	80	NaN	NaN	NaN	NaN	NaN	80	
227	Vatican City	29	NaN	NaN	NaN	29	NaN	0	
228	Western Sahara	10	NaN	1	NaN	9	NaN	0	
229	MS Zaandam	9	NaN	2	NaN	7	NaN	0	
230	China	503302	NaN	5,272	NaN	379,053	NaN	118,977	

231 rows × 21 columns

In [42]:

```
plt.figure(figsize = (12,8))
sns.barplot(
    x = 'TotalCases',
    y = 'CountryOther',
    data=df.sort_values('TotalCases', ascending=False).head(20),
    hue='CountryOther',
    palette='coolwarm',
    dodge=False, # Prevent bar separation
    legend=False
)
plt.title('Top 20 Countries by Total COVID-19 Cases', fontsize=16)
plt.xlabel('Total COVID-19 Cases', fontsize=14)
plt.ylabel('Country', fontsize=14)
plt.xticks(rotation=45)
plt.show()
```



```
In [43]:
df['TotalDeaths'] = df['TotalDeaths'].astype(str).str.replace(',', '')
df['TotalDeaths'] = pd.to_numeric(df['TotalDeaths'], errors='coerce')
df = df[df['TotalDeaths'].notna() & (df['TotalDeaths'] > 0)]
```

```
# Sort and select top N countries
df_sorted = df.sort_values('TotalDeaths', ascending=False)
top_countries = df_sorted.head(10)
In [45]:
top_countries
```

Out[45]:

	CountryOther	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	Se
0	USA	111820082	NaN	1219487.0	NaN	109,814,428	NaN	786,167	
4	Brazil	38743918	NaN	711380.0	NaN	36,249,161	NaN	1,783,377	
1	India	45035393	NaN	533570.0	NaN	NaN	NaN	NaN	
9	Russia	24124215	NaN	402756.0	NaN	23,545,818	NaN	175,641	
17	Mexico	7702809	NaN	334958.0	NaN	6,899,865	NaN	467,986	
8	UK	24910387	NaN	232112.0	NaN	24,678,275	NaN	0	
34	Peru	4572667	NaN	222161.0	NaN	4,350,506	NaN	0	
7	Italy	26723249	NaN	196487.0	NaN	26,361,218	NaN	165,544	
3	Germany	38828995	NaN	183027.0	NaN	38,240,600	NaN	405,368	
2	France	40138560	NaN	167642.0	NaN	39,970,918	NaN	0	

10 rows × 21 columns

In [46]:

```
sns.set_style('whitegrid')
```

```
# Plot Pie Chart using Matplotlib with Seaborn style
```

```
plt.figure(figsize=(12, 8))
```

```
plt.pie(
```

```
    top_countries['TotalDeaths'],
```

```
    labels=top_countries['CountryOther'],
```

```
    autopct='%1.1f%%',
```

```
    startangle=140,
```

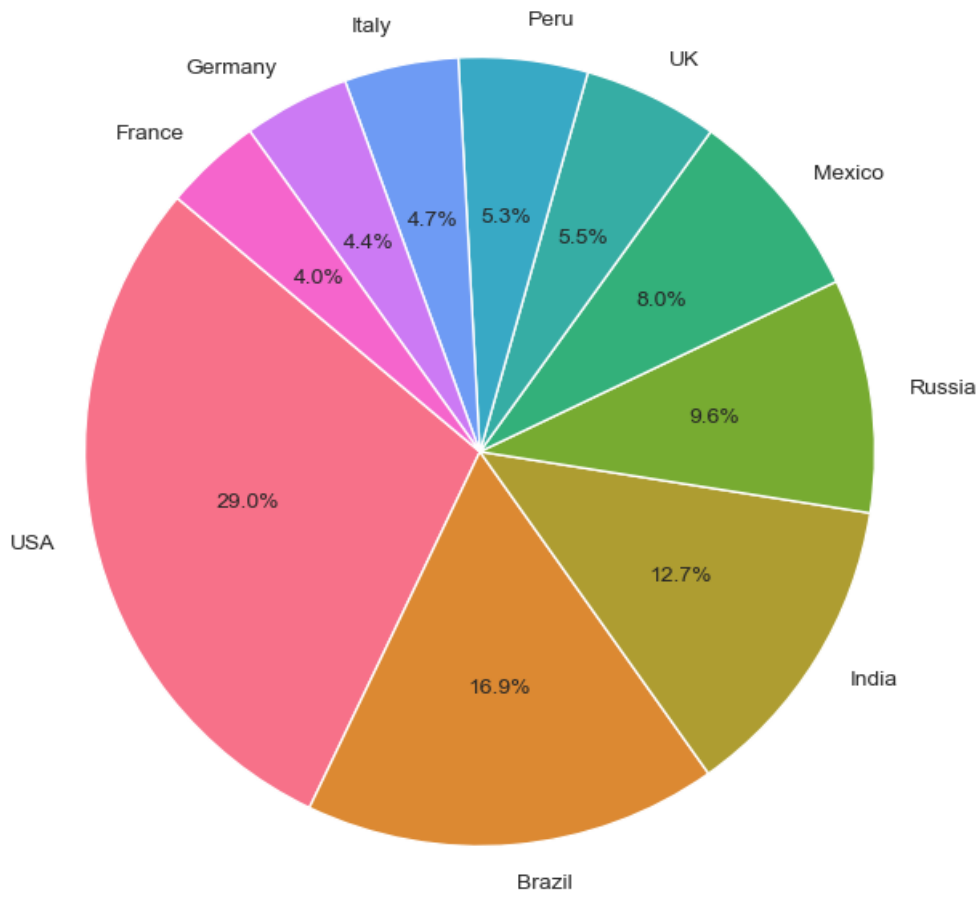
```
    colors=sns.color_palette('husl', len(top_countries))
```

```
)
```

```
plt.title('Top 10 Countries by Total COVID-19 Deaths', fontsize=16)
```

```
plt.show()
```

Top 10 Countries by Total COVID-19 Deaths



In []: