COVID-19 Data Analysis & Visualization using Python

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
df = pd.read_csv(r"C:\Users\Abhay_Thakur_PC\Downloads\country_wise_latest.csv")
df.head(5)
```

[2]:

	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 week % increase	WHO Region
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.50	69.49	5.04	35526	737	2.07	Eastern Mediterranean
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709	17.00	Europe
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282	18.07	Africa
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23	2.60	Europe
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201	26.84	Africa

[3]:
```
df.info()
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to
186 Data columns (total 15
columns):
 #   Column               Non-Null     Dtype
                          Count
---  ------               ------------ -----
                          --
 0   Country/Region       187 non-null object
 1   Confirmed            187 non-null int64
 2   Deaths               187 non-null int64
 3   Recovered            187 non-null int64
 4   Active               187 non-null int64
 5   New cases            187 non-null int64
 6   New deaths           187 non-null int64
 7   New recovered        187 non-null int64
 8   Deaths / 100 Cases   187 non-null float64
 9   Recovered / 100      187 non-null float64
     Cases
 10  Deaths / 100         187 non-null float64
     Recovered
 11  Confirmed last week  187 non-null int64
 12  1 week change        187 non-null int64
 13  1 week % increase    187 non-null float64
 14  WHO Region           187 non-null object
dtypes: float64(4), int64(9),
object(2) memory usage: 22.0+ KB
```

[4]:
```
df.describe()
```

[4]:

```
         Confirmed        Deaths    Recovered      Active    New cases\
count 1.870000e+02   187.000000 1.870000e+021.870000e+02  187.000000

mean   8.813094e+04  3497.518717 5.063148e+043.400194e+04 1222.957219

std    3.833187e+05 14100.002482 1.901882e+052.133262e+05 5710.374790

min    1.000000e+01     0.000000 0.000000e+000.000000e+00    0.000000

25%    1.114000e+03    18.500000 6.265000e+021.415000e+02    4.000000

50%    5.059000e+03   108.000000 2.815000e+031.600000e+03   49.000000

75%    4.046050e+04   734.000000 2.260600e+049.149000e+03  419.500000

max    4.290259e+06 148011.0000001.846641e+062.816444e+06
                                                 56336.000000
                                  Deaths / 100 Cases   Recovered / 100
       New deaths  New recovered Cases                                 \
```

```
       count  187.000000    187.000000        187.000000          187.000000
       mean    28.957219    933.812834          3.019519           64.820535
       std    120.037173   4197.719635          3.454302           26.287694
       min      0.000000      0.000000          0.000000            0.000000
       25%      0.000000      0.000000          0.945000           48.770000
       50%      1.000000     22.000000          2.150000           71.320000
       75%      6.000000    221.000000          3.875000           86.885000
       max   1076.000000  33728.000000         28.560000          100.000000

             Deaths / 100 RecoveredConfirmed last week1 week change \
       count               187.00       1.870000e+02   187.000000

       mean                   inf         7.868248e+04  9448.459893
       std                    NaN         3.382737e+05 47491.127684
       min                   0.00         1.000000e+01   -47.000000
       25%                   1.45         1.051500e+03    49.000000
       50%                   3.62         5.020000e+03   432.000000
       75%                   6.44         3.708050e+04  3172.000000
       max                    inf         3.834677e+06455582.000000
             1 week % increase
       count 187.000000 mean
       13.606203 std
       24.509838 min    -
       3.840000 25%
       2.775000
       50%   6.890000    75%
       16.855000         max
       226.320000
```

[5]:
```
df.isnull().sum()
```

[5]:
```
Country/Region         0
Confirmed              0
Deaths                 0
Recovered              0
Active                 0
New cases              0
New deaths             0
New recovered          0
Deaths / 100 Cases     0
Recovered / 100 Cases  0
Deaths / 100 Recovered 0
Confirmed last week    0
1 week change          0
```

```
1 week % increase      0
WHO Region             0
dtype: int64
```

[6]:
```
df.fillna(0,inplace=True)
```

[7]:
```
total_cases = df['Confirmed'].sum()
total_deaths = df['Deaths'].sum()
total_recovered = df['Recovered'].sum()
death_rate = (total_deaths / total_cases) * 100
recovery_rate = (total_recovered / total_cases) * 100
```

[8]:
```
print(" Global Summary   ")
print(f"Total Cases: {total_cases:,}")
print(f"Total Deaths: {total_deaths:,}")
print(f"Total Recovered: {total_recovered:,}")
print(f"Death Rate: {death_rate:.2f}%")
print(f"Recovery Rate: {recovery_rate:.2f}%")
```

```
 Global Summary
Total Cases: 16,480,485
Total Deaths: 654,036
Total Recovered: 9,468,087
Death Rate: 3.97%
Recovery Rate: 57.45%
```

[9]:
```
top_countries = df.sort_values(by="Confirmed",ascending=False).head(10)
```

[10]:
```
top_countries
```

[10]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases
173	US	4290259	148011	1325804	2816444	56336
23	Brazil	2442375	87618	1846641	508116	23284
79	India	1480073	33408	951166495499		44457
138	Russia	816680	13334	602249	201097	5607
154	South Africa	452529	7067	274925170537		7096
111	Mexico	395489	44022	303810	47657	4973
132	Peru	38971718418		272547	98752	13756
35	Chile	347923	9187	319954	18782	2133
177	United Kingdom	30170845844		1437254427		688
81	Iran	29360615912		255144	22550	2434

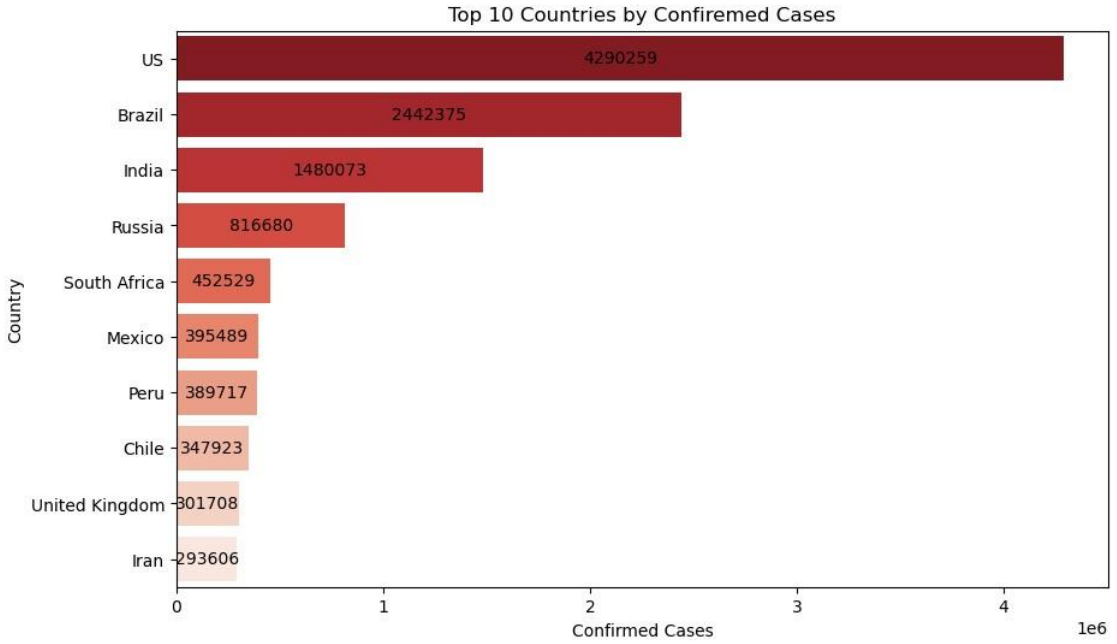
	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	\
173	1076	27941	3.45	30.90	
23	614	33728	3.59	75.61	
79	637	33598	2.26	64.26	
138	85	3077	1.63	73.74	
154	298	9848	1.56	60.75	
111	342	8588	11.13	76.82	
132	575	4697	4.73	69.93	
35	75	1859	2.64	91.96	
177	7	3	15.19	0.48	
81	212	1931	5.42	86.90	

	Deaths / 100 Recovered	Confirmed last week	1 week change	\
173	11.16	3834677	455582	
23	4.74	2118646	323729	
79	3.51	1155338	324735	
138	2.21	776212	40468	
154	2.57	373628	78901	
111	14.49	349396	46093	
132	6.76	357681	32036	
35	2.87	333029	14894	
177	3190.26	296944	4764	
81	6.24	276202	17404	

	1 week % increase	WHO Region
173	11.88	Americas
23	15.28	Americas
79	28.11	South-East Asia
138	5.21	Europe
154	21.12	Africa
111	13.19	Americas
132	8.96	Americas

```
35              4.47              Americas
177             1.60                Europe
81              6.30  Eastern Mediterranean
```

[11]:
```
plt.figure(figsize=(10,6))
ax = sns.barplot(x="Confirmed",y="Country/Region",hue="Country/
 ↪Region",data=top_countries , palette="Reds_r",legend=False)
plt.title("Top 10 Countries by Confiremed Cases ")
plt.xlabel("Confirmed Cases")
plt.ylabel("Country")

for i in ax.containers:
    ax.bar_label(i, fmt="%.0f", label_type="center",fontsize=10)
plt.show()
```



[12]:
```
df["Death Rate %"] = (df["Deaths"] / df["Confirmed"]) * 100
top_death_rate = df.sort_values(by="Death Rate %" ,
ascending=False).head(10)
```

[13]:
```
top_death_rate
```

[13]:
```
    Country/Region         Deaths Recovered  Active New cases  \
                   Confirmed
184          Yemen       1691   483       833    375        10
177  United Kingdom   30170845844       1437  254427       688
16         Belgium      664289822        1745239154       402
```

85	Italy	246286	35112	19859	312581	168		
61	France	220352	30212	81212	108928	2551		
77	Hungary	4448	596	3329	523	13		
120	Netherlands	534	136	160	189	47064	419	
111	Mexico	395489	44022	303810	47657	4973		
157	Spain	272421	28432	150376	93613	0		
183	Western Sahara	10	1	8	1	0		

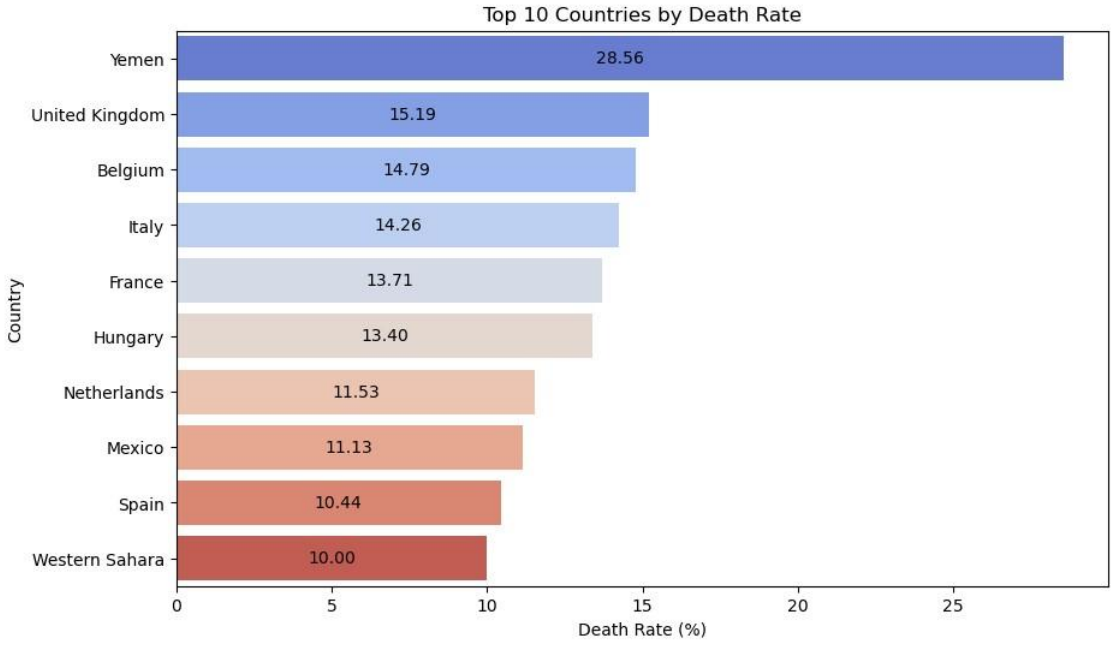
	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	\
184	4	36	28.56	49.26	
177	7	3	15.19	0.48	
16	1	14	14.79	26.27	
85	5	147	14.26	80.64	
61	17	267	13.71	36.86	
77	0	0	13.40	74.84	
120	1	0	11.53	0.35	
111	342	8588	11.13	76.82	
157	0	0	10.44	55.20	
183	0	0	10.00	80.00	

	Deaths / 100 Recovered	Confirmed last week	1 week change	\
184	57.98	1619	72	
177	3190.26	296944	4764	
16	56.28	64094	2334	
85	17.68	244624	1662	
61	37.20	214023	6329	
77	17.90	4339	109	
120	3259.26	52132	1281	
111	14.49	349396	46093	
157	18.91	264836	7585	

```
183                12.50                 10              0
```

	1 week % increase	WHO Region	Death Rate %
184	4.45	Eastern Mediterranean	28.562980
177	1.60	Europe	15.194824
16	3.64	Europe	14.785934
85	0.68	Europe	14.256596
61	2.96	Europe	13.710790
77	2.51	Europe	13.399281
120	2.46	Europe	11.532773
111	13.19	Americas	11.131030
157	2.86	Europe	10.436787
183	0.00	Africa	10.000000

[14]:
```
plt.figure(figsize=(10,6))
ax = sns.barplot(x="Death Rate %",y="Country/Region",hue="Country/
 ↪Region",data=top_death_rate,palette="coolwarm")
plt.title("Top 10 Countries by Death Rate ")
plt.xlabel("Death Rate (%)")
plt.ylabel("Country")
for i in ax.containers:
    ax.bar_label(i, fmt="%.2f", label_type="center",fontsize=10)
plt.show()
```

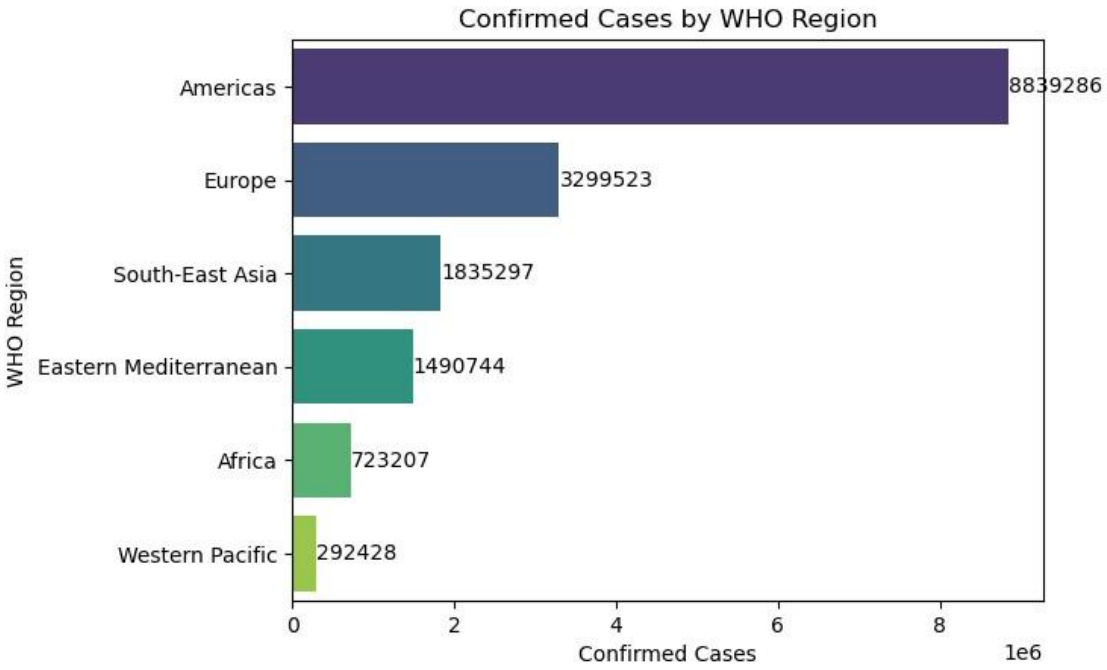


[15]:
```
region_cases = df.groupby("WHO Region")["Confirmed"].sum().reset_index().
↪sort_values(by="Confirmed", ascending=False)
```

[16]:
```
region_cases
```

[16]:
```
              WHO Region  Confirmed
1 Americas 8839286 3 Europe
3299523
4      South-East Asia  1835297
2 Eastern Mediterranean  1490744
0               Africa    723207
5       Western Pacific     292428
```

[53]:
```
ax =sns.barplot(x="Confirmed",  ="WHO Region",hue="WHO Region",
↪data=region_cases, palette="viridis")
plt.title("Confirmed Cases by WHO Region ")
plt.xlabel("Confirmed Cases")
plt.ylabel("WHO Region")
for i in ax.containers:
    ax.bar_label(i, fmt="%.0f", label_type="edge",fontsize=10)
plt.show()
```



[18]:
```
top_recovered = df.sort_values(by='Recovered',
ascending=False).head(10) top_recovered
```

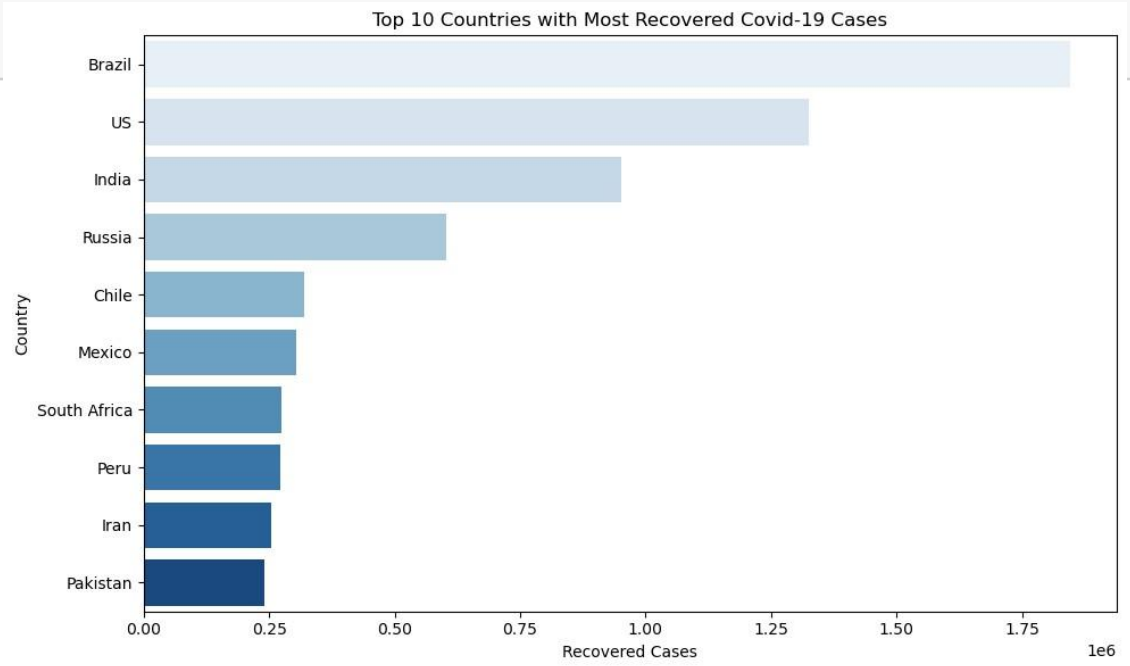
[18]:

	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
23	Brazil	2442375	87618	1846641	508116	23284	614	33728	3.59	75.61	4.74	2118646	323729
173	US	4290259	148011	1325804	2816444	56336	1076	27941	3.45	30.90	11.16	3834677	455582
79	India	1480073	33408	951166	495499	44457	637	33598	2.26	64.26	3.51	1155338	324735
138	Russia	816680	13334	602249	201097	5607	85	3077	1.63	73.74	2.21	776212	40468
35	Chile	347923	9187	319954	18782	2133	75	1859	2.64	91.96	2.87	333029	14894
111	Mexico	395489	44022	303810	47657	4973	342	8588	11.13	76.82	14.49	349396	46093
154	South Africa	452529	7067	274925	170537	7096	298	9848	1.56	60.75	2.57	373628	78901
132	Peru	389717	18418	272547	98752	13756	575	4697	4.73	69.93			
81	Iran	293606	15912	255144	22550	2434	212	1931	5.42	86.90			
128	Pakistan	274289	5842	241026	27421	1176	20	3592	2.13	87.87			

132	6.76	357681	32036
81	6.24	276202	17404
128	2.42	266096	8193

	1 week % increase	WHO Region	Death Rate %
23	15.28	Americas	3.587410
173	11.88	Americas	3.449932
79	28.11	South-East Asia	2.257186
138	5.21	Europe	1.632708
35	4.47	Americas	2.640527
111	13.19	Americas	11.131030
154	21.12	Africa	1.561668
132	8.96	Americas	4.725993
81	6.30	Eastern Mediterranean	5.419508
128	3.08	Eastern Mediterranean	2.129870

[52]:
```
plt.figure(figsize=(10,6))
sns.barplot(x="Recovered",y="Country/Region",hue="Country/
 ↪Region",data=top_recovered,palette="Blues")
plt.title("Top 10 Countries with Most Recovered Covid-
19 Cases") plt.xlabel("Recovered Cases")
plt.ylabel("Country") plt.tight_layout() plt.show()
```



[20]:
```
top_active = df.sort_values(by='Active', ascending=False).head(10)
top_active
```

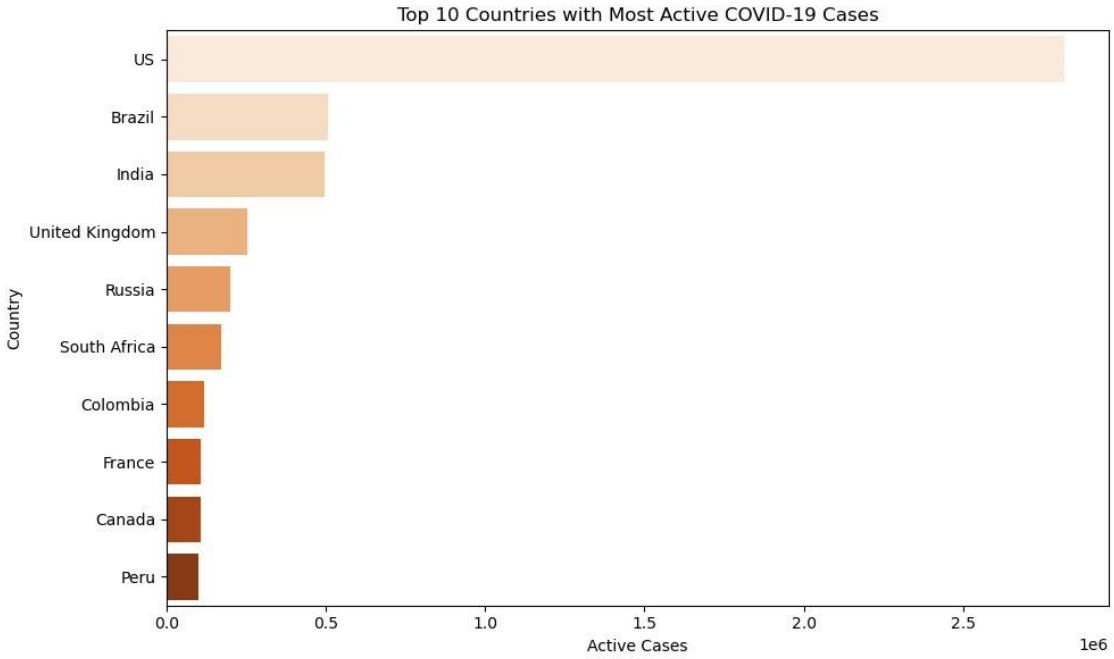
[20]:

	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
173	US	4290259	148011	13258042	816444	56336	1076	27941	3.45	30.90	11.16	3834677	455582
23	Brazil	244237	587618	1846641	508116	23284	614	33728	3.59	75.61	4.74	2118646	323729
79	India	1480073	33408	951166	495499	44457	637	33598	2.26	64.26	3.51	1155338	324735
177	United Kingdom	3017084	5844	1437254	427	688	7	3	15.19	0.48			
138	Russia	8166801	3334	6022492	01097	5607	85	3077	1.63	73.74			
154	South Africa	452529	7067	2749251	70537	7096	298	9848	1.56	60.75			
37	Colombia	257101	8777	1311611	17163	16306	508	11494	3.41	51.02			
61	France	2203523	0212	8121210	8928	2551	17	267	13.71	36.86			
32	Canada	116458	8944		0107514	682	11	0	7.68	0.00			
132	Peru	3897171	8418	272547	98752	13756	575	4697	4.73	69.93			

177	3190.26	296944	4764
138	2.21	776212	40468
154	2.57	373628	78901
37	6.69	204005	53096
61	37.20	214023	6329
32	inf	112925	3533
132	6.76	357681	32036

	1 week % increase	WHO Region	Death Rate %
173	11.88	Americas	3.449932
23	15.28	Americas	3.587410
79	28.11	South-East Asia	2.257186
177	1.60	Europe	15.194824
138	5.21	Europe	1.632708
154	21.12	Africa	1.561668
37	26.03	Americas	3.413833
61	2.96	Europe	13.710790
32	3.13	Americas	7.680022
132	8.96	Americas	4.725993

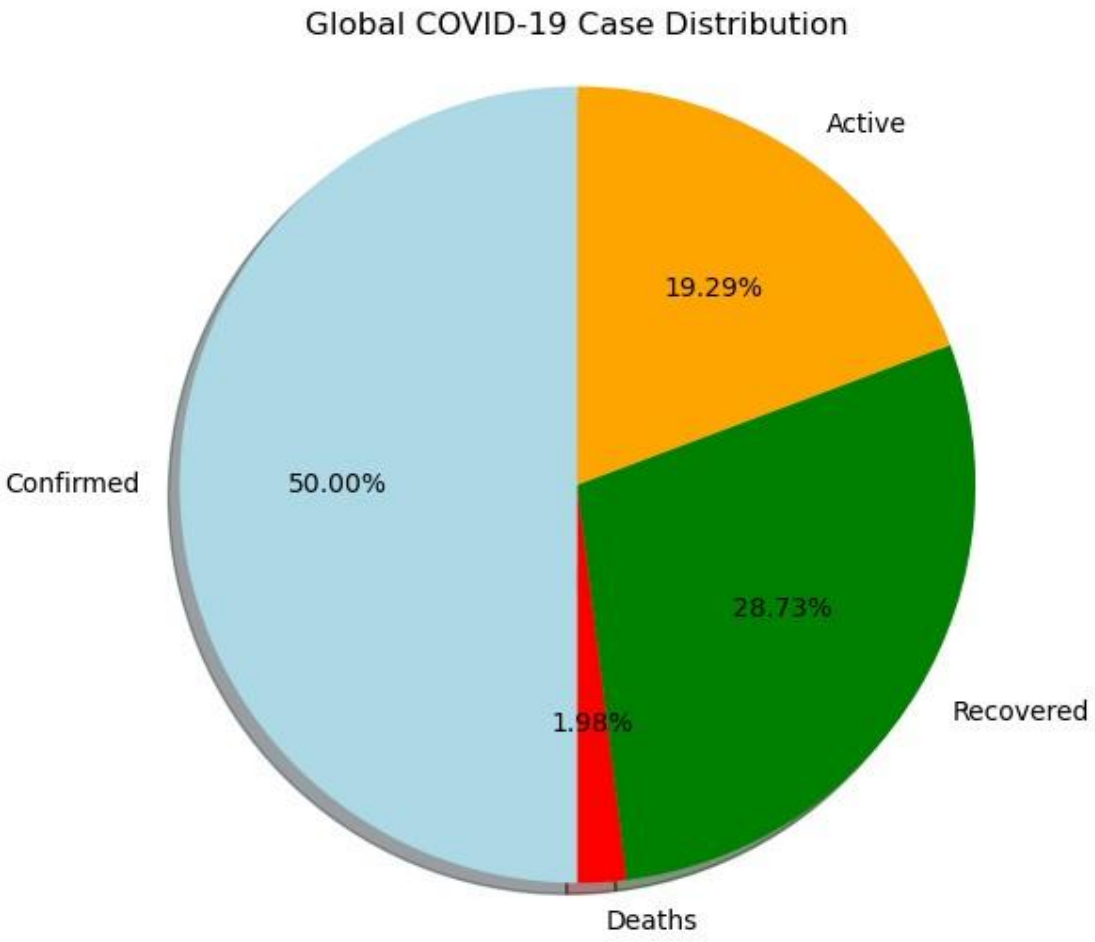
[51]:
```
top_active = df.sort_values(by='Active', ascending=False).head(10)

plt.figure(figsize=(10,6))
sns.barplot(
    x='Active',
    y='Country/Region',
    data=top_active,
    hue='Country/Region',
    palette='Oranges',
    legend=False
)
plt.title("Top 10 Countries with Most Active COVID-19 Cases ")
plt.xlabel("Active Cases")
plt.ylabel("Country")
plt.tight_layout()
plt.show()
```



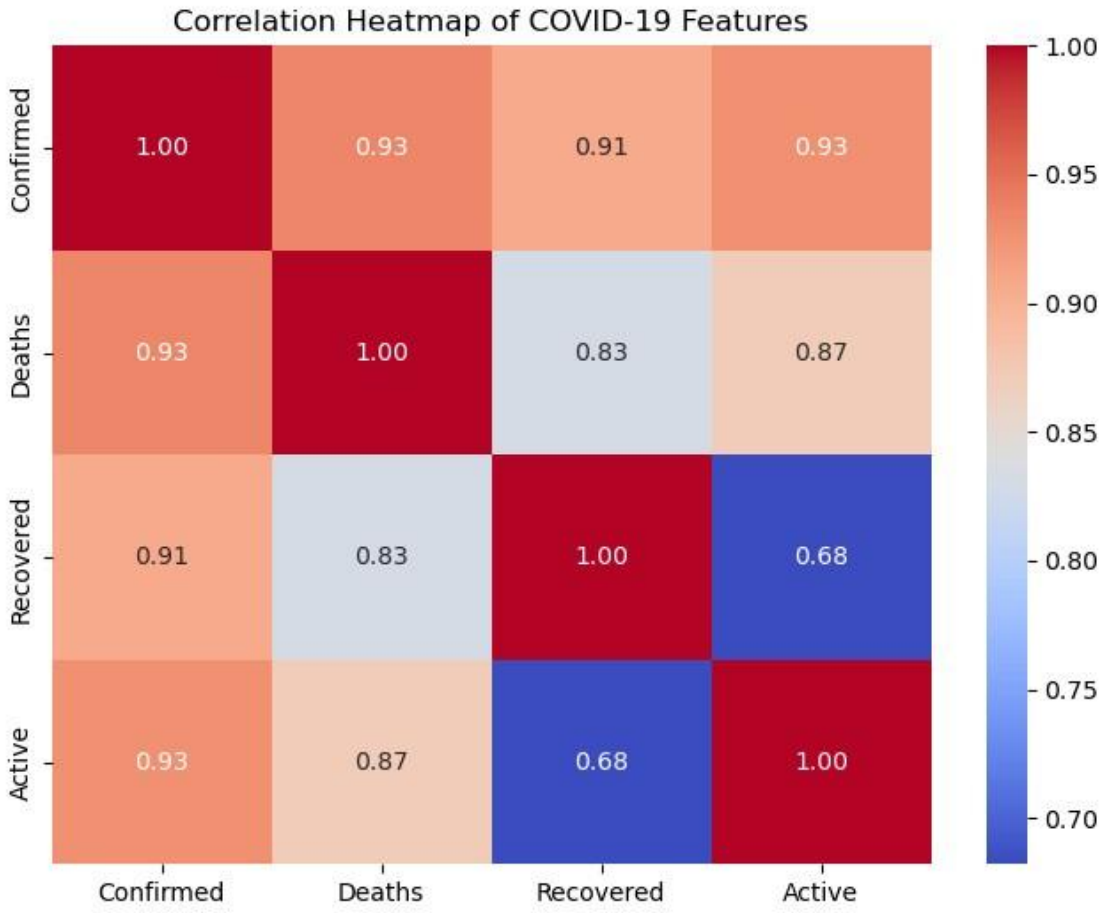
[23]:

[24]:
```
labels = ['Confirmed', 'Deaths', 'Recovered', 'Active']
sizes = [total_confirmed, total_deaths, total_recovered, total_active]
colors = ['lightblue', 'red', 'green', 'orange']
```

[25]:
```
plt.figure(figsize=(6, 6))
plt.pie(sizes, labels=labels, colors=colors, startangle=90, autopct='%1.2f%%',
 ↪shadow=True)
plt.title('Global COVID-19 Case Distribution')
plt.axis('equal')
plt.show()
```



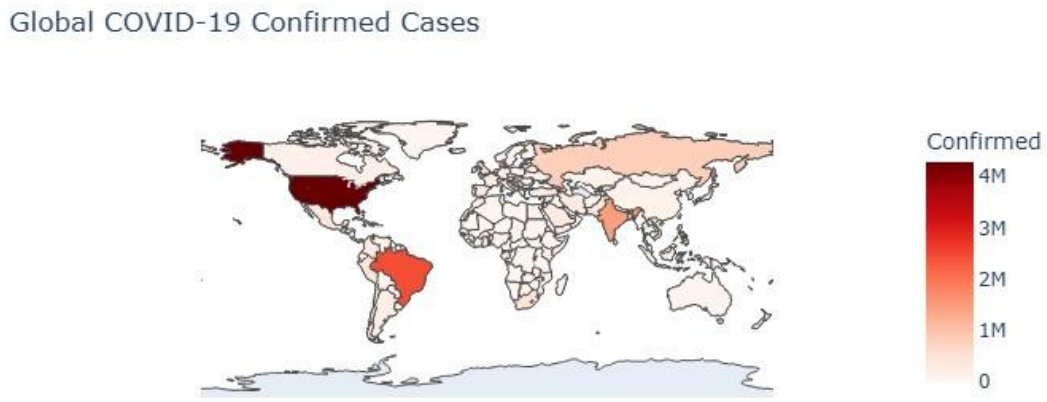
[26]:
```
 plt.figure(figsize=(8,6))
sns.heatmap(df[['Confirmed','Deaths','Recovered','Active']].
 ↪corr(),annot=True,cmap='coolwarm',fmt='.2f')
plt.title('Correlation Heatmap of COVID-19
Features') plt.show()
```



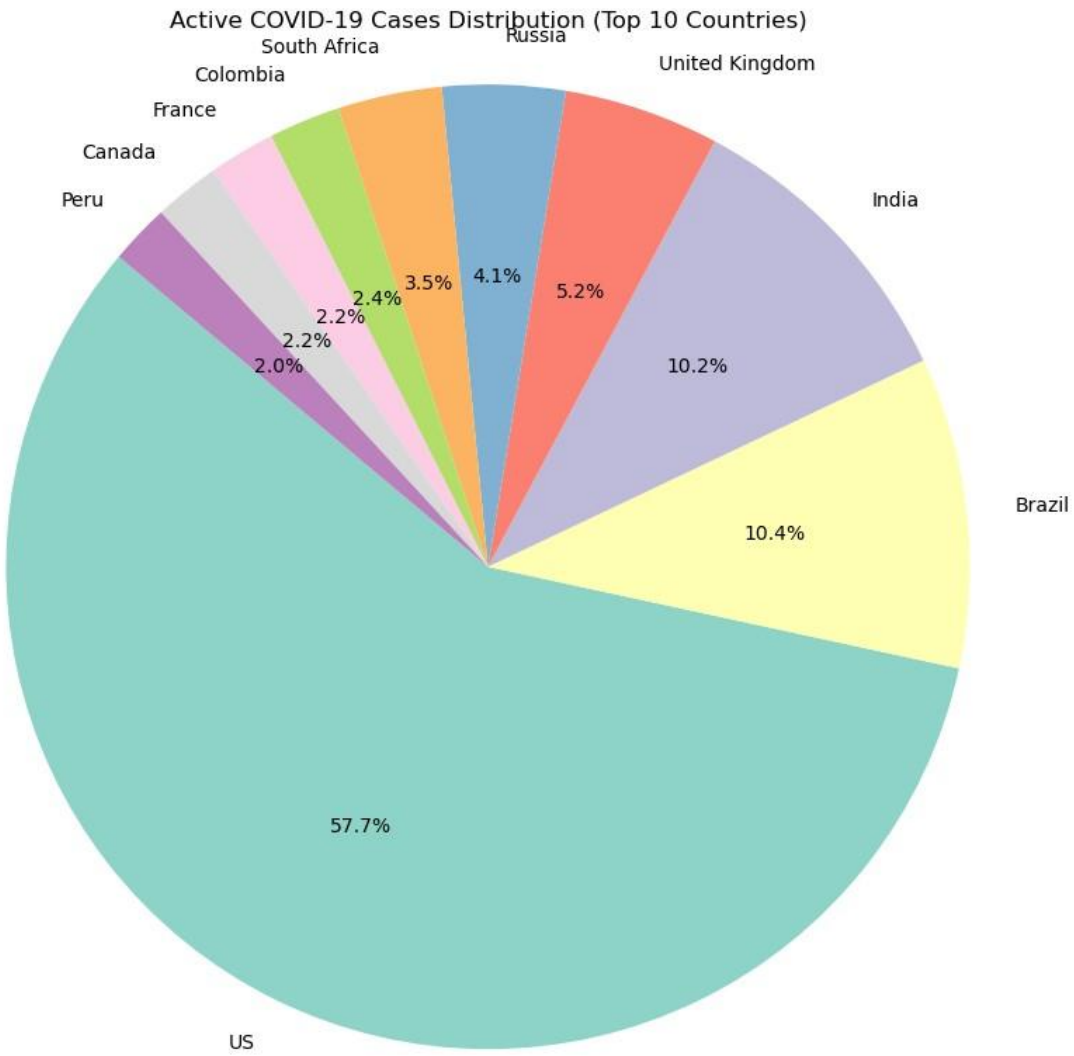
[27]:
```
pip install plotly
```

```
Requirement already satisfied: plotly in
c:\users\abhay_thakur_pc\anaconda3\lib\site-packages (5.22.0)
Requirement already satisfied: tenacity>=6.2.0 in
c:\users\abhay_thakur_pc\anaconda3\lib\site-packages (from plotly)
(8.2.2) Requirement already satisfied: packaging in
c:\users\abhay_thakur_pc\anaconda3\lib\site-packages (from plotly)
(23.2) Note: you may need to restart the kernel to use updated
packages.
```

[28]:
```
import plotly.express as px

fig = px.choropleth(df,
    locations="Country/Region",
    locationmode="country names",
    color="Confirmed",
    hover_name="Country/Region",
    color_continuous_scale="Reds",
    title="Global COVID-19 Confirmed Cases"
)

fig.update_layout(geo=dict(showframe=False, showcoastlines=True))
fig.show()
```

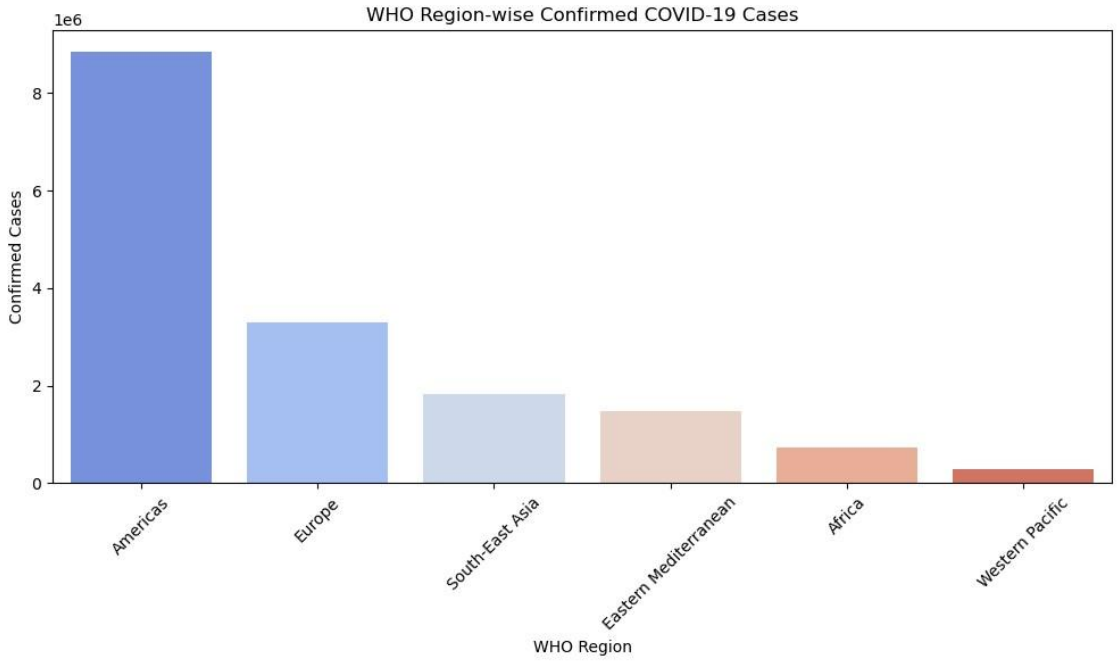


[31]:
```
top_active = df.sort_values(by='Active', ascending=False).head(10)

plt.figure(figsize=(8,8))
plt.pie(top_active['Active'],labels=top_active['Country/Region'],a
utopct='%1.
 ↪1f%%',startangle=140,colors=sns.color_palette('Set3'))
plt.title('Active COVID-19 Cases Distribution (Top 10
Countries)') plt.axis('equal') plt.tight_layout()
plt.show()
```

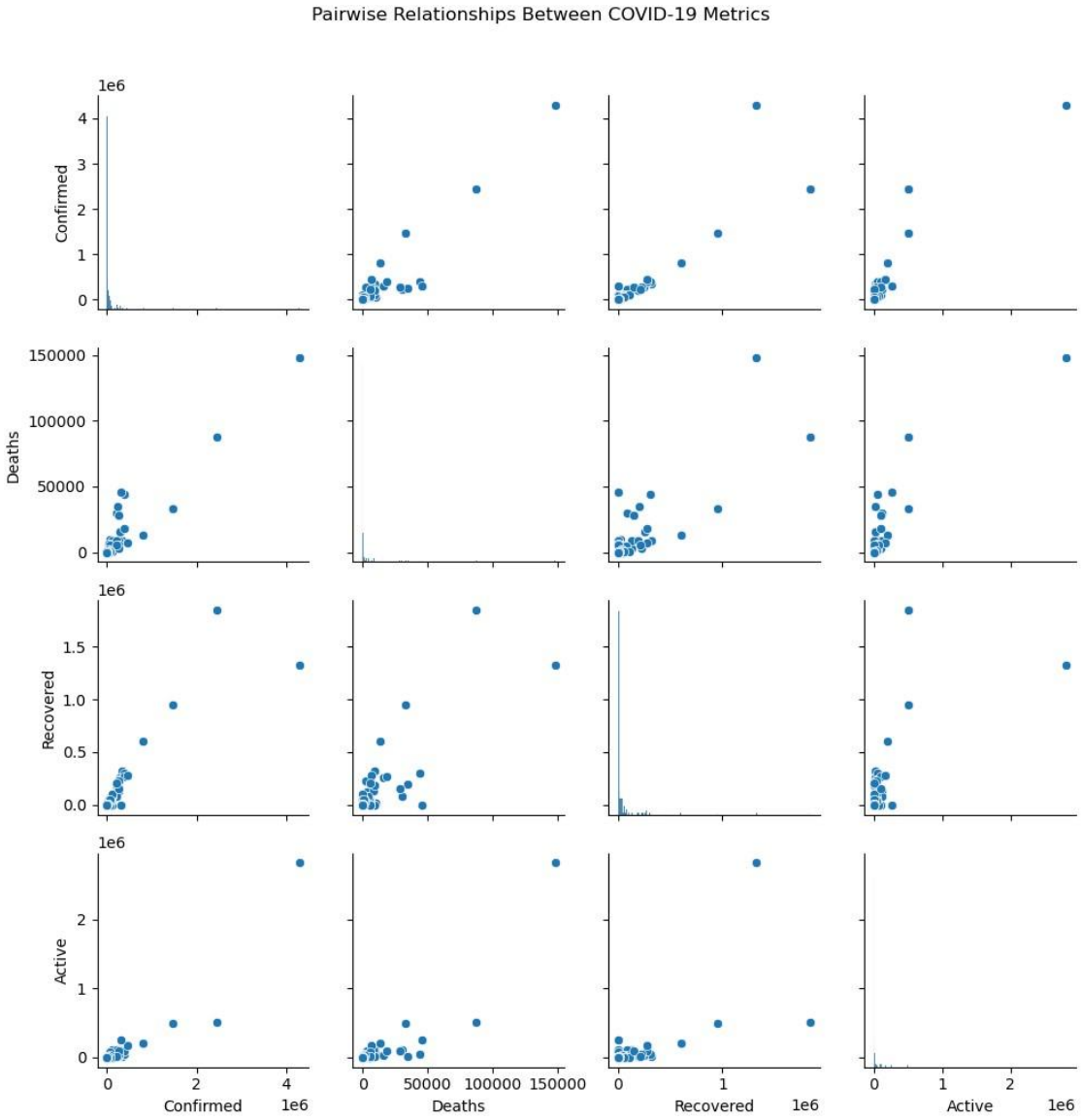


[48]:
```
who_data = df.groupby('WHO Region')[['Confirmed', 'Deaths',
'Recovered',␣
                    ↪'Active']].sum().sort_values(by='Confirmed',
                                ascending=False).reset_index()

plt.figure(figsize=(10,6)) sns.barplot(x='WHO Region',
y='Confirmed', hue='WHO Region', data=who_data,␣
↪palette='coolwarm', legend=False)

plt.title('WHO Region-wise Confirmed COVID-19
Cases') plt.xlabel('WHO Region')
plt.ylabel('Confirmed Cases')
```

```
plt.xticks(rotation=45) plt.tight_layout()
plt.show()
```



[54]:
```
sns.pairplot(df[['Confirmed', 'Deaths', 'Recovered',
'Active']]) plt.suptitle('Pairwise Relationships Between
COVID-19 Metrics', y=1.02) plt.tight_layout() plt.show()
```


Pairwise Relationships Between COVID-19 Metrics
Confirmed
Deaths
Recovered
Active
1e6

[55]:
```
fig = px.scatter(df,
                 x="Country/Region",
                 y="Confirmed",
                 size="Confirmed",
                 color="Country/Region",
                 hover_name="Country/Region",
                 size_max=60,
                 title="Total Confirmed Cases by All Country ",
                 labels={"Confirmed": "Total Confirmed Cases"},
                 height=600)
fig.update_layout(showlegend=False)
fig.show()
```