Countries clustered by latitude and longitude

In [88]:
```
from IPython.display import Image
Image(filename='worldmap.jpg')
```

Out[88]:

In [128…
```
import pandas as pd
from decimal import Decimal
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.cluster import KMeans
```

In [129…
```
data = pd.read_csv("all_countries.csv")
df = data.iloc[:,1:4]
df = df.round(2)
df = df.dropna()
```

In [90]:
```
display(df)
```

	latitude	longitude	country
0	42.55	1.60	Andorra
1	23.42	53.85	United Arab Emirates
2	33.94	67.71	Afghanistan
3	17.06	-61.80	Antigua and Barbuda
4	18.22	-63.07	Anguilla
...	...	...	...
240	15.55	48.52	Yemen
241	-12.83	45.17	Mayotte
242	-30.56	22.94	South Africa

243	-13.13	27.85	Zambia
244	-19.02	29.15	Zimbabwe

244 rows × 3 columns

In [99]:
```
x = df.iloc[:,1:2]
kmeans = KMeans(5)
```

In [ ]:
```
kmeans.fit(x)
```

In [ ]:
```
identified_clusters = kmeans.fit_predict(x)
identified_clusters
```

In [102…
```
data_with_clusters = df.reset_index(drop=True).copy()
data_with_clusters['cluster'] = identified_clusters
```

All countries divided is 5 clusters (by continent)

In [103…
```
data_with_clusters
```

Out[103]:

	latitude	longitude	country	cluster
0	42.55	1.60	Andorra	1
1	23.42	53.85	United Arab Emirates	3
2	33.94	67.71	Afghanistan	3
3	17.06	-61.80	Antigua and Barbuda	4
4	18.22	-63.07	Anguilla	4
...	...	...	...	...
239	15.55	48.52	Yemen	3
240	-12.83	45.17	Mayotte	3
241	-30.56	22.94	South Africa	1
242	-13.13	27.85	Zambia	3
243	-19.02	29.15	Zimbabwe	3

244 rows × 4 columns

All countries in cluster 1

In [119…
```
cluster_1 = data_with_clusters[data_with_clusters['cluster'] == 1]
display(cluster_1)
```

	latitude	longitude	country	cluster
0	42.55	1.60	Andorra	1
5	41.15	20.17	Albania	1
8	-11.20	17.87	Angola	1
9	-75.25	-0.07	Antarctica	1
12	47.52	14.55	Austria	1

...	...	...	...	...
211	8.62	0.82	Togo	1
217	33.89	9.54	Tunisia	1
229	41.90	12.45	Vatican City	1
238	42.60	20.90	Kosovo	1
241	-30.56	22.94	South Africa	1

83 rows × 4 columns

In [133…
```
plt.scatter(data_with_clusters['longitude'], data_with_clusters['latitude'],
            c=data_with_clusters['cluster'], cmap='rainbow')
plt.xlim(-180, 180)
plt. ylim(-90, 90)


plt.show
```

Out[133]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```