In [3]:
```
import pandas as pd
import numpy as np
from IPython.display import Image
import pydotplus

import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
```

Finding 2 - Explore Geography and Infrastructure of Movie Industries for Different Countries

Visualization Analysis using Choropleth Mapping

In [4]:
```
df = pd.read_csv("movies_cleaned2.csv")
```

In [5]:
```
df_choro = df.copy()
```

To get the Choropleth to populate, each country needs its respective ISO code. We downloaded a file that has a list of Country names and their Iso codes

In [6]:
```
df_iso = pd.read_csv('ISOcodes.csv', encoding='latin-1')
```

In [7]:
```
df_iso.rename(columns={'Country Name, Full ':'Country'},inplace=True)
```

In [8]:
```
df_choro.rename(columns={'production_country':'Country'},inplace=True)
```

Convert Numbers to Numeric

In [9]:
```
df_choro['revenue'] = pd.to_numeric(df_choro['revenue'],errors='coerce')
df_choro['budget'] = pd.to_numeric(df_choro['budget'],errors='coerce')
df_choro['popularity'] = pd.to_numeric(df_choro['popularity'],errors='coerce')
df_choro['vote_average'] = pd.to_numeric(df_choro['vote_average'],errors='coerce')
```

In [10]:
```
df_choro1 = df_choro.groupby('Country').agg({'revenue':'sum','budget':'sum',\
                                             'popularity':'mean','vote_average':'mean',\
                                             'production_company':'nunique'})
```

In [11]:
```
df_choro1.reset_index(inplace=True)
```

We know that Budget and Revenue are highly correlated. It seems Infrastructure (Production Companies) sometimes play a part

In [12]:
```
df_choro1.corr()
```

Out[12]:

	revenue	budget	popularity	vote_average	production_company
revenue	1.000000	0.999857	0.114110	-0.020408	0.955848
budget	0.999857	1.000000	0.115091	-0.020547	0.959012
popularity	0.114110	0.115091	1.000000	-0.081682	0.129841
vote_average	-0.020408	-0.020547	-0.081682	1.000000	-0.017908
production_company	0.955848	0.959012	0.129841	-0.017908	1.000000

In [13]:
```
df_choro1.sort_values(by=['revenue','production_company'],ascending=[False,False]).head(20)
```

Out[13]:

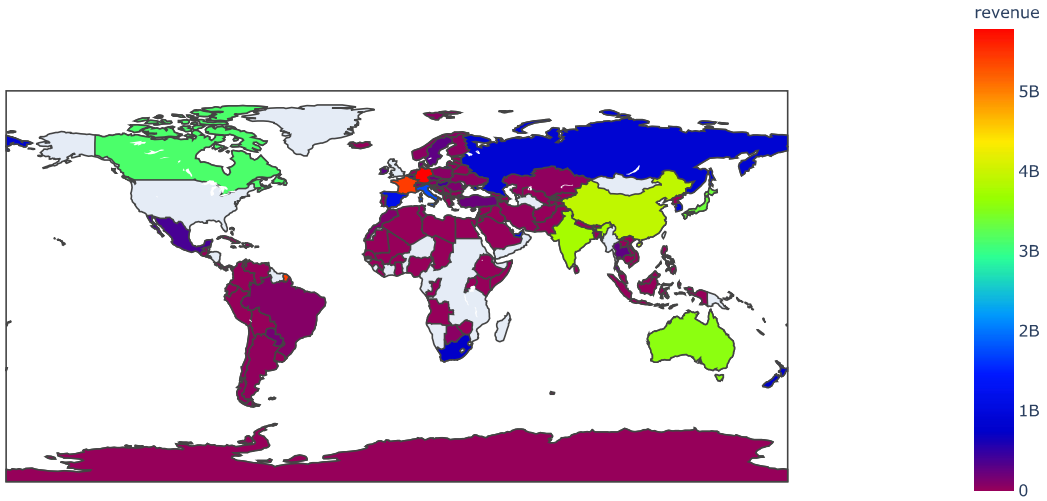
	Country	revenue	budget	popularity	vote_average	production_company
128	USA	4.509030e+11	162276833962	4.267380	5.598638	5050
132	United Kingdom	1.577196e+10	5816362384	3.195716	5.983153	1114
49	Germany	5.786016e+09	2565370488	2.145768	5.942456	509
46	France	5.398091e+09	4347991984	2.613919	6.122363	809
28	China	3.892239e+09	1381309467	2.765970	6.073239	97
58	India	3.721754e+09	1358932797	1.475195	5.954199	299
8	Australia	3.538882e+09	1601383811	3.246859	5.764858	220
66	Japan	3.331304e+09	943605933	2.176299	6.232245	330
25	Canada	3.105380e+09	1809761464	2.573823	5.424642	546
55	Hong Kong	2.117819e+09	1132893341	2.918977	6.236383	135
64	Italy	1.723248e+09	1094220482	1.996076	5.671873	590
118	Spain	1.155817e+09	804357750	2.412007	5.664852	343
131	United Arab Emirates	9.882652e+08	284000000	8.706423	6.200000	7
116	South Korea	9.150571e+08	445004762	2.330549	6.495896	152
106	Russia	8.609034e+08	999353007	1.218013	5.623256	211
91	New Zealand	7.509484e+08	282512027	3.203800	6.010127	40
115	South Africa	7.203206e+08	457547487	3.469435	5.512000	55
83	Malta	5.318650e+08	200000000	13.229975	3.350000	2
56	Hungary	4.098421e+08	291250144	1.220373	5.807826	66
85	Mexico	3.784256e+08	173691456	1.975590	5.761033	122

In [14]:
```
df_map = df_choro1.merge(df_iso, left_on='Country', right_on='Country')
```

We can change the color to any criteria (budget, revenue, popularity, etc.)

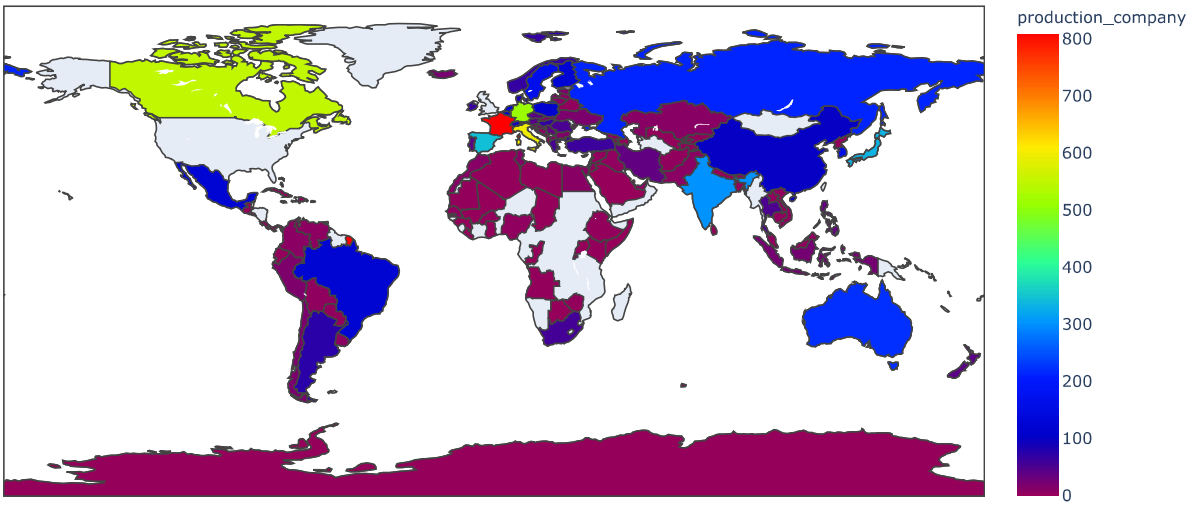
Revenue by Country (Excluding USA and UK)

In [15]:
```
fig = px.choropleth(df_map[(df_map.Country != 'USA') & (df_map.Country !='United Kingdom')],\
                    locations="ISO3-digit Alpha",
                    color="revenue",
                    hover_name="Country",
                    color_continuous_scale=px.colors.sequential.Rainbow)
fig.show()
```



Number of Major Production Companies per Country (Excluding USA and UK)

In [16]:
```
fig = px.choropleth(df_map[(df_map.Country != 'USA') & (df_map.Country !='United Kingdom')],\
                    locations="ISO3-digit Alpha",
                    color="production_company",
                    hover_name="Country",
                    color_continuous_scale=px.colors.sequential.Rainbow)
fig.show()
```



Explanation

We see that Canada, West Europe, East Asia and Australia have the highest movie revenues behind USA and UK.

- Exclued USA and UK for scaling
- The "red" colored country in South America is French Guana, which is part of France
- There was one movie registered to Antarctica - a documentary

Countries like France, Germany and Canada who have a strong infrastructure in the movie industry generate the some of the most revenue. However, the correlation between Production Companies and Revenue don't always hold true. This relationship might be true for lower revenue generating countries.

When we look at the second Choropleth, China, India and Australia have few production companies, yet generate high revenue. In contrast, some countries in Northern Europe and South America have many production companies, but generate a low revenue.