

Mục tiêu

Sử dụng thư viện Seaborn để vẽ và tinh chỉnh được một số biểu đồ trên bộ dữ liệu GDPlist

Bộ dữ liệu GDPlist thống kê giá trị GDP của 125 quốc gia trên 6 châu lục.

Trong bài thực hành này, chúng ta sẽ sử dụng thư viện Seaborn vừa học được để vẽ một số biểu đồ trên bộ dữ liệu này

Khai báo thư viện cần dùng

```
import pandas as pd
import seaborn as sns
```

Đọc dữ liệu

```
df = pd.read_csv("GDPlist.csv", encoding = "ISO-8859-1")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 125 entries, 0 to 124
Data columns (total 3 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Country               125 non-null   object
 1   Continent             125 non-null   object
 2   GDP (millions of US$) 125 non-null   int64
dtypes: int64(1), object(2)
memory usage: 3.1+ KB
```

```
df.describe()
```

GDP (millions of US\$)

count	1.250000e+02
mean	5.554278e+05
std	1.649266e+06
min	1.060000e+04
25%	2.574200e+04
50%	9.608900e+04
75%	3.332380e+05

```
df.head()
```

	Country	Continent	GDP (millions of US\$)
0	Afghanistan	Asia	18181
1	Albania	Europe	12847
2	Algeria	Africa	190709
3	Angola	Africa	100948
4	Argentina	South America	447644

Loại bỏ dữ liệu bị khuyết

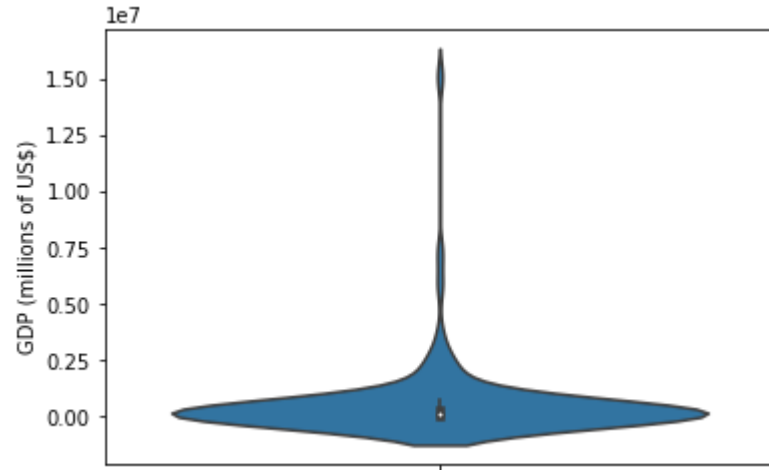
```
df = df.dropna()
```

Vẽ biểu đồ phân bố

Biểu đồ phân bố giá trị GDP toàn cầu

```
sns.violinplot(y=df["GDP (millions of US$)"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0a55e26750>
```



Có thể nhận thấy giá trị GDP phân tập trung ở giá trị < 2500000, số lượng các nước có GDP cao hơn ít hơn rất nhiều

Biểu đồ phân bố giá trị GDP châu Á

```
sns.violinplot(y=df[df["Continent"]=="Asia"]["GDP (millions of US$)"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0a53852250>
```

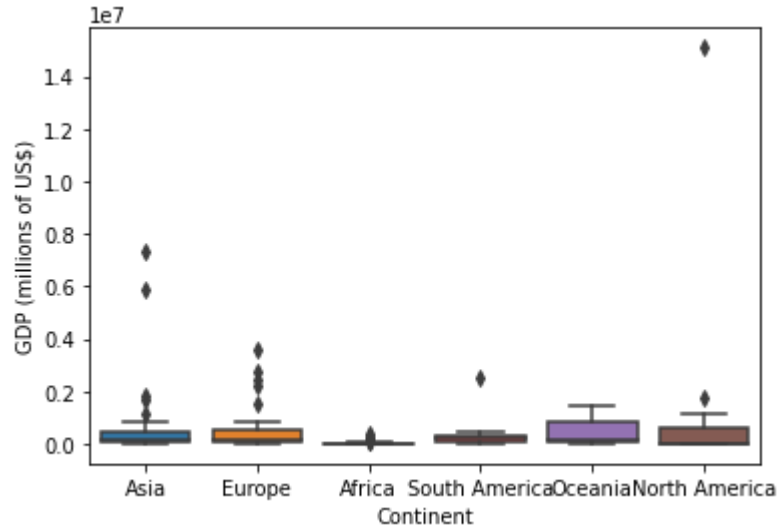
GDP các nước châu Á có vẻ đồng đều hơn so với trên thế giới

Vẽ biểu đồ boxplot

Biểu đồ box plot nhóm theo châu lục

```
sns.boxplot(x="Continent", y = "GDP (millions of US$)", data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0a537f28d0>
```



Nhận thấy rằng Bắc Mỹ có một quốc gia có GDP rất cao, GDP châu Phi khá thấp, GDP của châu Đại Dương có vẻ đồng đều hơn

Tổng kết

Qua bài này, chúng ta đã ôn lại cách sử dụng biểu đồ phân bố và biểu đồ boxplot