



SPCK CSA-06

Người thuyết trình: Minh Đức

Data được sử dụng để lập bảng

covid19.csv

Unnamed: 0	Province/State	Country/Region	date	Lat	Long	confirmed	deaths	recoveries	
0	0	NaN	Afghanistan	1/22/2020	33.000000	65.000000	0	0	0
1	1	NaN	Albania	1/22/2020	41.153300	20.168300	0	0	0
2	2	NaN	Algeria	1/22/2020	28.033900	1.659600	0	0	0
3	3	NaN	Andorra	1/22/2020	42.506300	1.521800	0	0	0
4	4	NaN	Angola	1/22/2020	-11.202700	17.873900	0	0	0
...
37039	37039	NaN	Sao Tome and Principe	6/16/2020	0.186360	6.613081	671	12	182
37040	37040	NaN	Yemen	6/16/2020	15.552727	48.516388	885	214	91
37041	37041	NaN	Comoros	6/16/2020	-11.645500	43.333300	197	3	127
37042	37042	NaN	Tajikistan	6/16/2020	38.861034	71.276093	5160	50	3624
37043	37043	NaN	Lesotho	6/16/2020	-29.609988	28.233608	4	0	2

37044 rows × 9 columns

Data được sử dụng để lập bảng
population.csv

	Country (or dependency)	Population (2020)	Yearly Change	Net Change	Density (P/Km ²)	Land Area (Km ²)	Migrants (net)	Fert. Rate	Med. Age	Urban Pop %	World Share
0	China	1438207241	0.39%	5540090	153	9388211	-348399.0	1.7	38	61%	18.47%
1	India	1377233523	0.99%	13586631	464	2973190	-532687.0	2.2	28	35%	17.70%
2	United States	330610570	0.59%	1937734	36	9147420	954806.0	1.8	38	83%	4.25%
3	Indonesia	272931713	1.07%	2898047	151	1811570	-98955.0	2.3	30	56%	3.51%
4	Pakistan	219992900	2.00%	4327022	287	770880	-233379.0	3.6	23	35%	2.83%
...
230	Montserrat	4991	0.06%	3	50	100	NaN	N.A.	N.A.	10%	0.00%
231	Falkland Islands	3458	3.05%	103	0	12170	NaN	N.A.	N.A.	66%	0.00%
232	Niue	1624	0.68%	11	6	260	NaN	N.A.	N.A.	46%	0.00%
233	Tokelau	1354	1.27%	17	136	10	NaN	N.A.	N.A.	0%	0.00%
234	Holy See	801	0.25%	2	2003	0	NaN	N.A.	N.A.	N.A.	0.00%

Giải quyết vấn đề: Những số liệu không có (N.A)

```
data1.fillna(0).replace('NaN','N.A.')  
df1=data1.dropna(how='all')  
data2.fillna(0).replace('N.A.','NaN')
```

Tìm ra những số liệu cần thiết dựa trên các vấn đề, thông tin(số ca mắc cv19 2020,...)

Country/Region	confirmed	deaths	recoveries
Afghanistan	592218	11926	73221
Albania	68066	2345	45362
Algeria	450886	37954	244751
Andorra	58933	3291	36276
Angola	4014	225	1186
...
West Bank and Gaza	29828	176	17014
Western Sahara	496	22	320
Yemen	13171	2884	665
Zambia	39816	392	24822
Zimbabwe	7349	283	1197

Country (or dependency)	Population (2020)
China	1438207241
India	1377233523
United States	330610570
Indonesia	272931713
Pakistan	219992900
...	...
Montserrat	4991
Falkland Islands	3458
Niue	1624
Tokelau	1354
Holy See	801

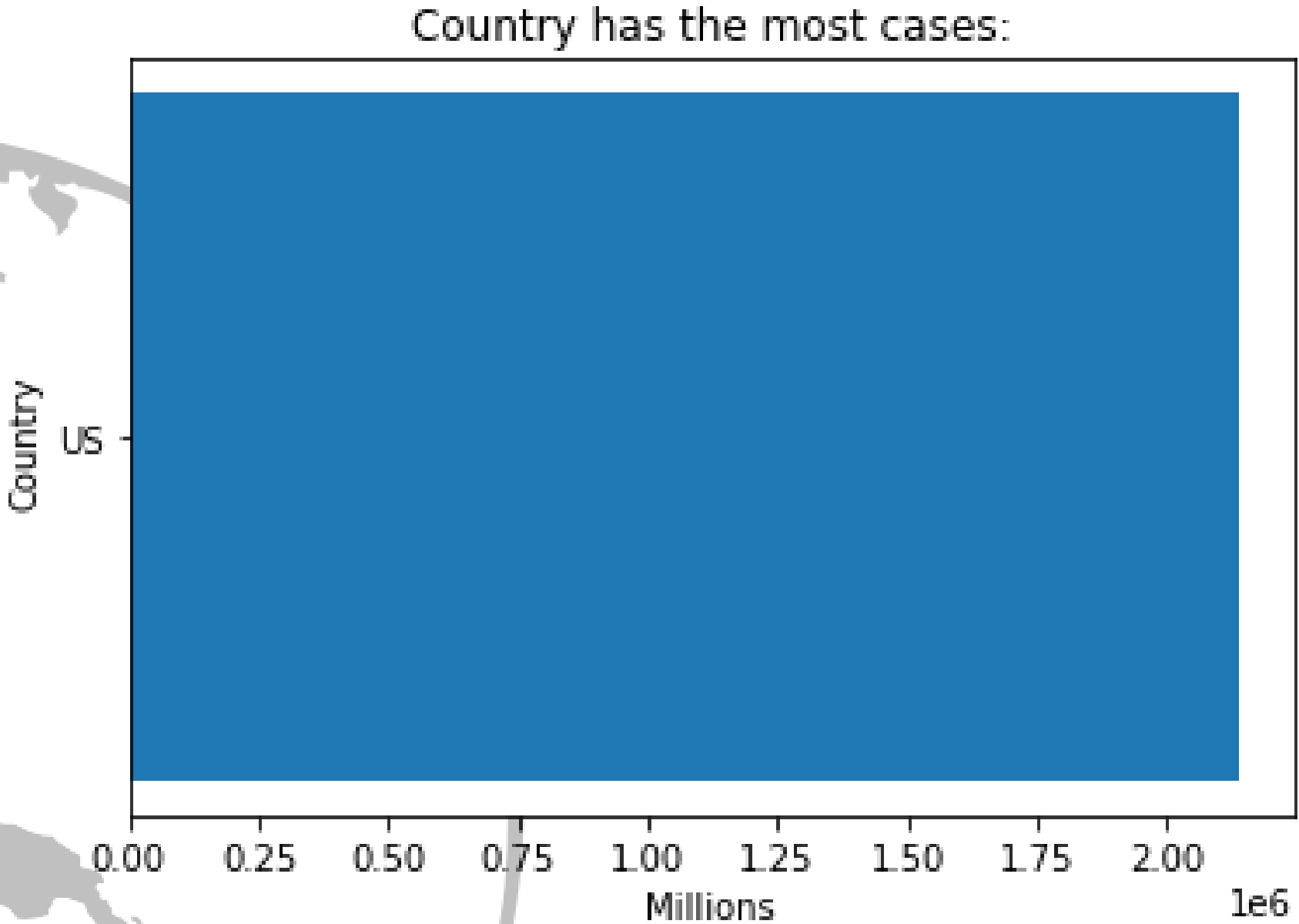
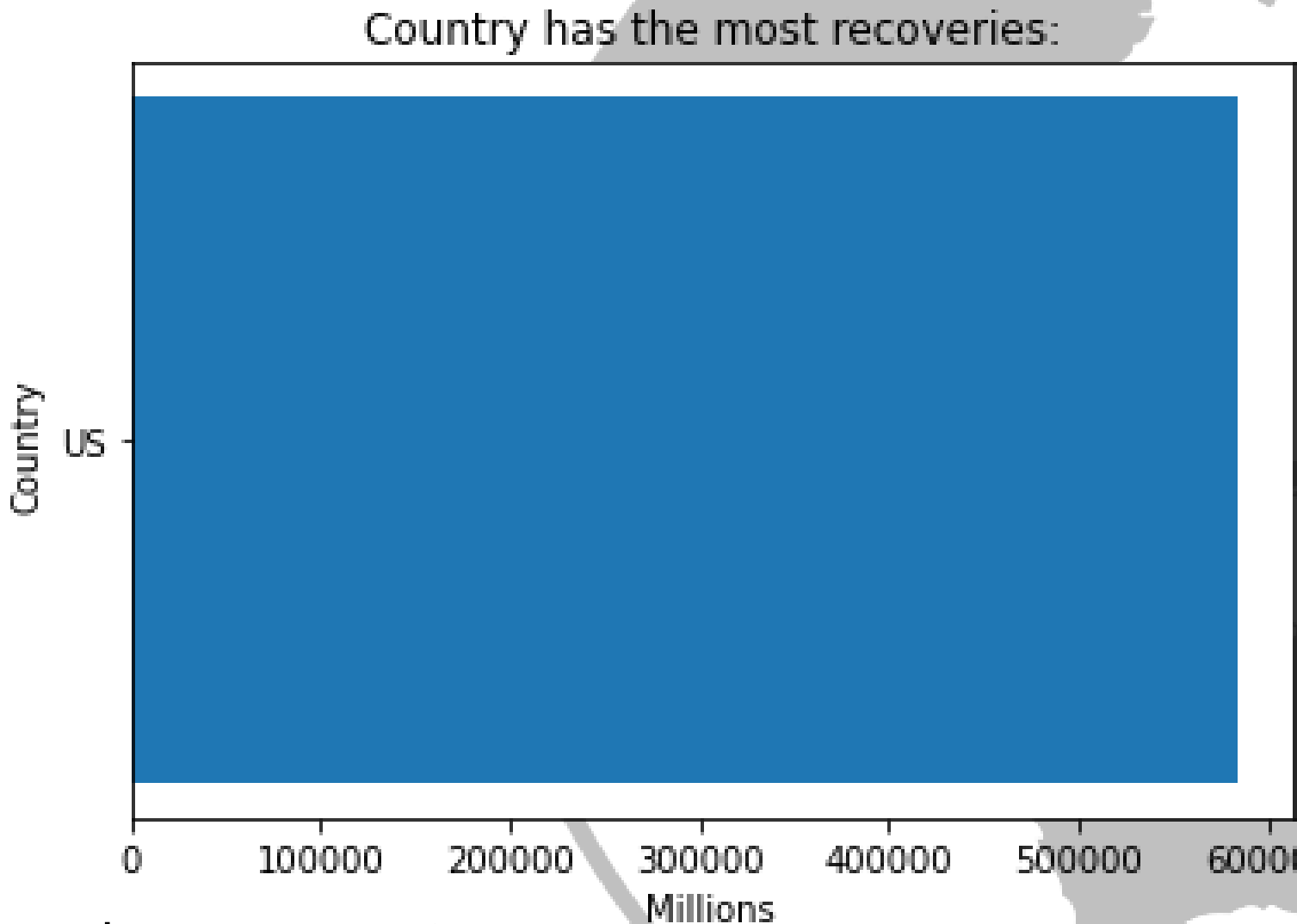
Land Area (Km ²)
9388211
2973190
9147420
1811570
770880
...
100
12170
260
10
0

Med. Age
38
28
38
30
23
...

World Share
18.47%
17.70%
4.25%
3.51%
2.83%
...
0.00%
0.00%
0.00%
0.00%
0.00%


population data:

- +)Total population: 7777967034
- +)Total land area: (130094083, 'km^2')
- +)Countries has the average(steadiest) world share from highest to lowest: United States, Indonesia,Pakistan,Brazil,Nigeria



covid19 data:

- +)Total confirmed: 329135580
- +)Total deaths: 20540202
- +)Total recoveries: 124277666
- +)Death rate: 6.240650737304062 %
- +)Recovery rate: 37.75880626457948 %



```
data2.loc[:,['Country/Region', 'confirmed','deaths','recoveries']].groupby('Country/Region').sum()
print("covid19 data:")
print("+)Total confirmed:",(data2['confirmed'].sum()))
print("+)Total deaths:",(data2['deaths'].sum()))
print("+)Total recoveries:",(data2['recoveries'].sum()))
deathrate=((data2['deaths'].sum()/data2['confirmed'].sum())*100)
print("+)Death rate:",deathrate,"%")
recoveryrate=((data2['recoveries'].sum()/data2['confirmed'].sum())*100)
print("+)Recovery rate:",recoveryrate,"%")
```

covid19 data:

+)Total confirmed: 329135580

+)Total deaths: 20540202

+)Total recoveries: 124277666

+)Death rate: 6.240650737304062 %

+)Recovery rate: 37.75880626457948 %

population data:

+)Total population: 7777967034

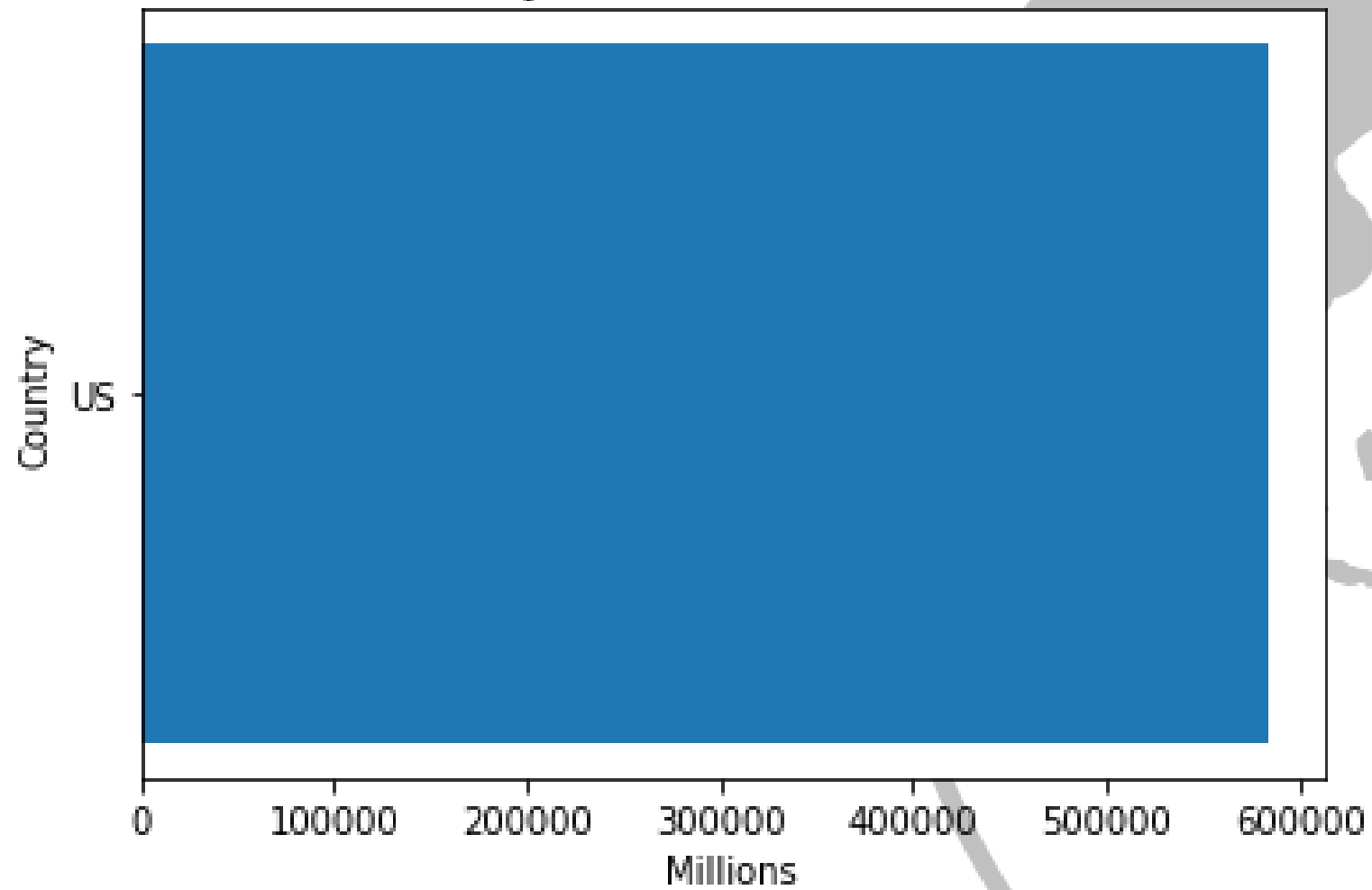
+)Total land area: (130094083, 'km^2')

+World median age:

+)Countries has the average(steadiest) world share from highest to lowest: United States, Indonesia,Pakistan,Brazil,Nigeria

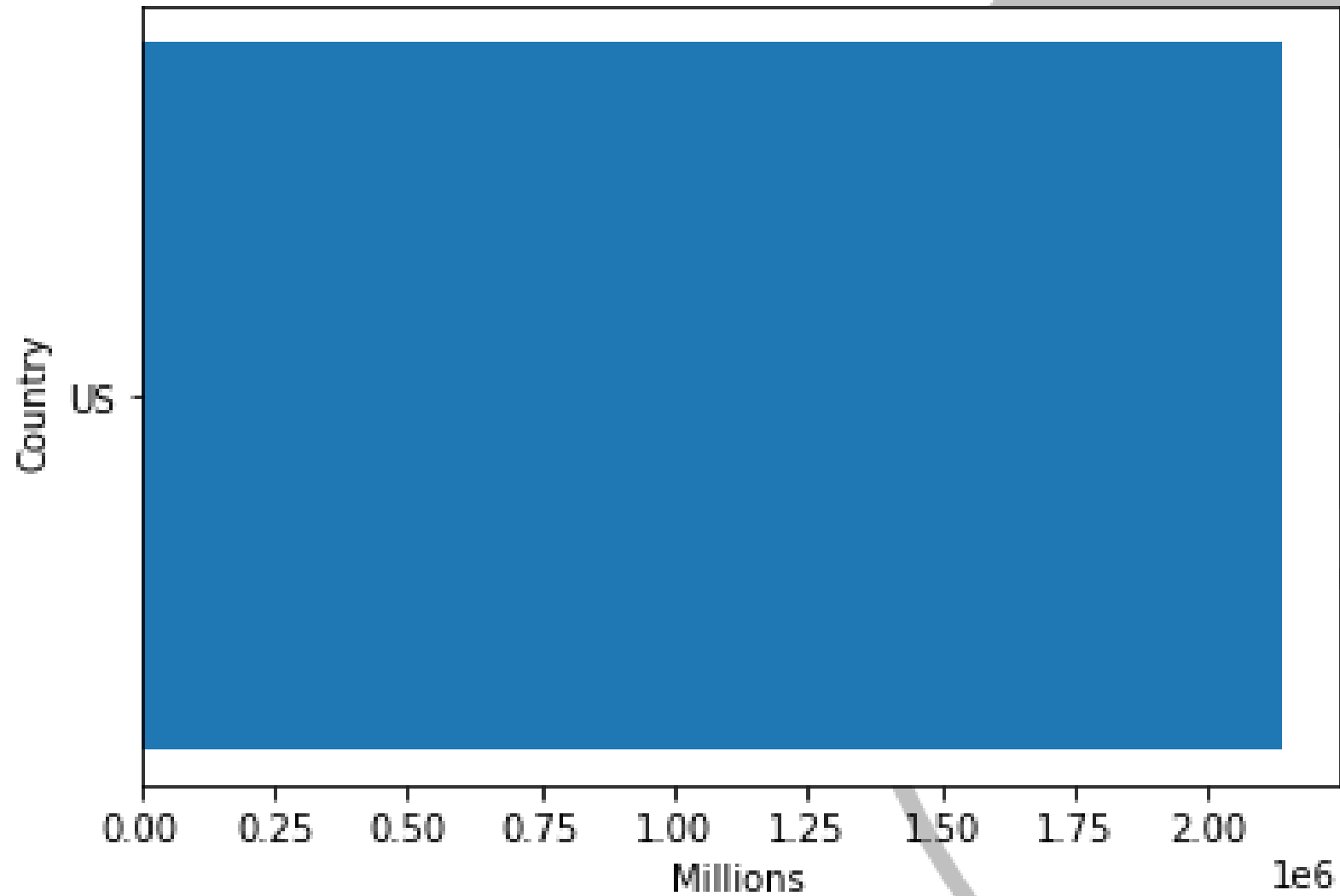
```
print("+)Total population:",(data1['Population (2020)'].sum()))
print("+)Total land area:",(data1['Land Area (Km²)'].sum(),"km^2"))
sorted_worldshare=data1.sort_values(by=['World Share'], ascending=False)
fivedworldshare=sorted_worldshare.head()
print("Countries has the average(steadiest) world share from highest to lowest:",fivedworldshare['Country (or dependency)'])
medianage=int(data1['Med. Age'].sum())
print("+)World median age :",medianage/235)
```

Country has the most recoveries:



```
sorted_recoveries=data2.sort_values(by=['recoveries'], ascending=False)
fiverecoveries=sorted_recoveries.head(5)
x=fiverecoveries['Country/Region']
y=fiverecoveries['recoveries']
#graphing
plt.barh(x, y)
plt.title('Country has the most recoveries:')
plt.xlabel("Millions")
plt.ylabel("Country")
plt.show()
print(fiverecoveries)
print(y)
```


Country has the most cases:



```
sorted_confirmed=data2.sort_values(by=['confirmed'], ascending=False)
fivedconfirmed=sorted_confirmed.head()
x=fivedconfirmed['Country/Region']
y=fivedconfirmed['confirmed']
#graphing
plt.barh(x, y)
plt.title('Country has the most cases:')
plt.xlabel("Millions")
plt.ylabel("Country")
plt.show()
print(fivedconfirmed)
```



Qua bài SPCK và Khóa lần này, em tự đánh giá :

Có tiếp thu và áp dụng kiến thức đã được học và tự học cũng như ví dụ của anh chị đưa ra trong chính sphaam nhưng:

-Bài làm thiếu sơ sài, sự sáng tạo về mặt phân tích dữ liệu và còn thiếu khá nhiều yếu tố khác nữa.

-Chưa vận dụng hết tiềm năng của một set data cũng như các phần nhỏ trong cả 2 datas ví dụ như phần median age, date,...

+)Về mặt tốt thì "có thể" dùng để tạo ra sự so sánh giữa các nước với nhau, nghiên cứu các biến đổi của từng nước cũng như tìm ra các vấn đề của các nước qua số liệu nêu trên

Cảm ơn mọi người đã lắng nghe phần trình bày của em

