

UMAP Microcredential

Final Capstone Project

Virtualize relationship between Area, GPP and Population by province



THE POWER OF VISUAL CHARTS

QUESTION



why Thailand is growing too slow

Poor Government management?

Import Library

- numpy
- pandas
- matplotlib
- seaborn

```
• ✓import numpy as np  
    import pandas as pd  
    pd.__version__
```

'1.4.0'

```
%matplotlib inline  
import matplotlib.pyplot as plt  
import seaborn as sns
```

Prepare Data

- Province Name(EN, TH)
- Population from 2012-2021
- GPP 2019

1	Province,Population (millions),GPP (billions ?),% of national GDP,GPP (billions US\$ nominal),GPP (billions I\$ PPP),GPP per capita (?),GPP per capita (US\$ nominal),GPP per capita (I\$ PPP)	
2	Krung Thep Maha Nakhon,8,912,"5,022.02",31,88,173,76,434,4,"573,907","19,749","55,297"	
3	Samut Prakan,2,171,717,05,4,71,25,69,64,22,"343,215","12,176","34,092"	
4	Pathum Thani,1,729,380,69,2,39,13,02,32,55,"254,627","8,039","22,509"	
5	Samut Sakhon,1,042,398,10,2,40,13,10,32,75,"411,326","12,914","36,159"	
6	Nakhon Pathom,1,186,332,63,2,06,11,26,28,15,"308,167","18,074","28,207"	
7	Nonthaburi,1,722,316,63,1,92,10,49,26,22,"204,484","6,513","18,236"	
8	Saraburi,0,758,236,64,1,45,7,93,18,82,"330,750","10,844","30,363"	
9	Sing Buri,0,190,26,58,0,16,0,89,2,22,"129,095","5,146","14,409"	
10	Chainat,0,314,31,85,0,20,1,13,2,89,"101,282","4,316","12,089"	
11	Ang Thong,0,259,27,79,0,18,0,99,2,47,"107,129","4,390","12,292"	
12	Lopburi,0,777,111,92,0,66,3,58,8,98,"144,041","5,045","14,126"	
13	Phra Nakhon Si Ayutthaya,0,907,403,60,2,38,13,31,33,27,"465,972","14,988","41,966"	
14	Chonburi,1,819,976,46,6,10,33,25,83,..12,"581,475","18,558","51,948"	
15	Chachoengsao,0,823,341,12,2,29,12,47,31,17,"427,409","15,453","43,268"	
16	Rayong,0,980,984,98,6,12,33,73,84,32,"1,095,667","34,995","97,986"	
17	Trat,0,274,46,96,0,25,1,38,3,45,"171,189","5,637","15,784"	
18	Chanthaburi,0,553,138,44,0,75,4,08,11,11,"254,582","7,423","20,784"	
19	Nakhon Nayok,0,278,26,84,0,18,0,96,2,40,"96,589","4,217","11,888"	
20	Prachinburi,0,621,297,25,1,89,10,29,25,72,"486,601","16,862","47,214"	
21	Sa Kaeo,0,630,45,25,0,34,1,63,4,07,"72,555","3,170","8,876"	
22	Khon Kaen,1,736,204,12,1,25,6,81,17,02,"117,560","4,466","12,505"	
23	Udon Thani,1,255,111,26,0,66,3,68,9,06,"88,673","3,344","9,363"	
24	Loei,0,543,52,67,0,33,1,78,4,45,"97,903","3,751","10,503"	
25	Nong Khai,0,456,48,05,0,26,1,37,3,42,"89,913","3,494","9,783"	
26	Mukdahan,0,345,25,89,0,17,0,82,2,07,"74,729","2,524","7,067"	
27	Nakhon Phanom,0,564,42,89,0,28,1,43,3,57,"76,000","3,142","8,798"	
28	Sakon Nakhon,0,925,55,63,0,36,1,91,4,77,"68,887","2,567","7,188"	
29	Kalasin,0,914,55,84,0,36,1,91,4,77,"61,084","2,887","8,084"	
30	Nakhon Ratchasima,2,515,274,90,1,75,9,53,23,82,"110,301","4,299","12,012"	
31	Chaiyaphum,0,954,68,09,0,39,2,12,5,30,"63,010","2,749","7,697"	
32	Yasothon,0,481,26,04,0,17,0,90,2,25,"54,183","2,437","6,824"	
33	Ubon Ratchathani,1,738,120,49,0,73,4,00,10,00,"70,551","2,885","7,854"	
34	Roi Et,1,069,73,48,0,48,2,43,6,07,"68,751","2,883","8,072"	
35	Buriram,1,247,84,33,0,55,2,66,6,65,"67,621","2,665","7,462"	
36	Surin,1,107,72,88,0,47,2,44,6,10,"65,818","2,776","7,772"	
37	Maha Sarakham,0,826,56,00,0,36,1,92,4,80,"67,784","3,082","8,630"	
38	Uthai Thani,0,826,56,00,0,36,1,92,4,80,"67,784","3,082","8,630"	

Area.xlsx		
	A	B
1	ProvinceEN	Area (km2)
2	Krung Thep Ma	1,564
3	Amnat Charoer	3,290
4	Ang Thong	950
5	Bueng Kan	4,003
6	Buriram	10,080
7	Chachoengsao	5,169
8	Chai Nat	2,506
9	Chaiyaphum	12,698
10	Chanthaburi	6,415
11	Chiang Mai	22,135
12	Chiang Rai	11,503
13	Chonburi	4,508
14	Chumphon	5,998
15	Kalasin	6,936
16	Kamphaeng Ph	8,512
17	Kanchanaburi	19,385
18	Khon Kaen	10,659
19	Krabi	5,323
20	Iamnang	12,488

GDP and GPP?

- GDP(Gross Domestic Product)
to measure Country is growing?
- GPP(Gross Provincial Product)
like GDP but calculate by province

Clean Data

- ~~drop na~~
- Finding dat na value from another source instead.
- Change Data type

```
#merge 3 table together to plot and see relationship
scatter = area.merge(total[['province', '2019']], how="left", left_on='ProvinceEN', right_on='province')
scatter = scatter.merge(gpp2019[['province', 'GPP(billions)']], how="left", left_on='ProvinceEN', right_on='province')
scatter = scatter.rename(columns={'ProvinceEN': 'province', '2019': 'population'})
scatter = scatter[['province', 'Area (km2)', 'population', 'GPP(billions)']]
scatter = scatter.set_index('province')
scatter[scatter['population'].isna()]
#ooh! it a NaN
```

province	Area (km2)	population	GPP(billions)
Buriram	10080	NaN	84.33
Chonburi	4508	NaN	97646
Lopburi	6493	NaN	111.92
Nong Bua Lamphu	4099	NaN	25.19
Phang Nga	5495	NaN	71.76
Prachinburi	5026	NaN	297.25
Sisaket	8936	NaN	69.57

```
#ref for NAN: https://housingkc.nha.co.th/files/article/attachments/ffaef8523776e58763edecf0906c208e.pdf
#Finding population Nan and change it!
scatter.loc['Buriram', 'population'] = 1595747
scatter.loc['Chonburi', 'population'] = 1558381
scatter.loc['Lopburi', 'population'] = 755556
scatter.loc['Nong Bua Lamphu', 'population'] = 512780
scatter.loc['Phang Nga', 'population'] = 268788
scatter.loc['Prachinburi', 'population'] = 494688
scatter.loc['Sisaket', 'population'] = 1472859
scatter[scatter['population'].isna()]
```

province	Area (km2)	population	GPP(billions)
----------	------------	------------	---------------

```
#Bangkok have comma in GPP col so change it!
scatter.loc['Krung Thep Maha Nakhon', 'GPP(billions)'] = 5022.02
scatter.loc['Krung Thep Maha Nakhon', 'GPP(billions)']

51] ... 5022.02

52] #Find na in GPP col
scatter[scatter['GPP(billions)'].isna()]

53] ...
```

province	Area (km2)	population	GPP(billions)
Chai Nat	2506	326611.0	NaN
Phatthalung	3861	524865.0	NaN

```
#Change it!
scatter.loc['Chai Nat', 'GPP(billions)'] = 31.85
scatter.loc['Phatthalung', 'GPP(billions)'] = 36.48
scatter[scatter['GPP(billions)'].isna()]

54] ...
```

province	Area (km2)	population	GPP(billions)
----------	------------	------------	---------------

```
# For sure set all column to float type
scatter = scatter.astype(float)
scatter.head()
```

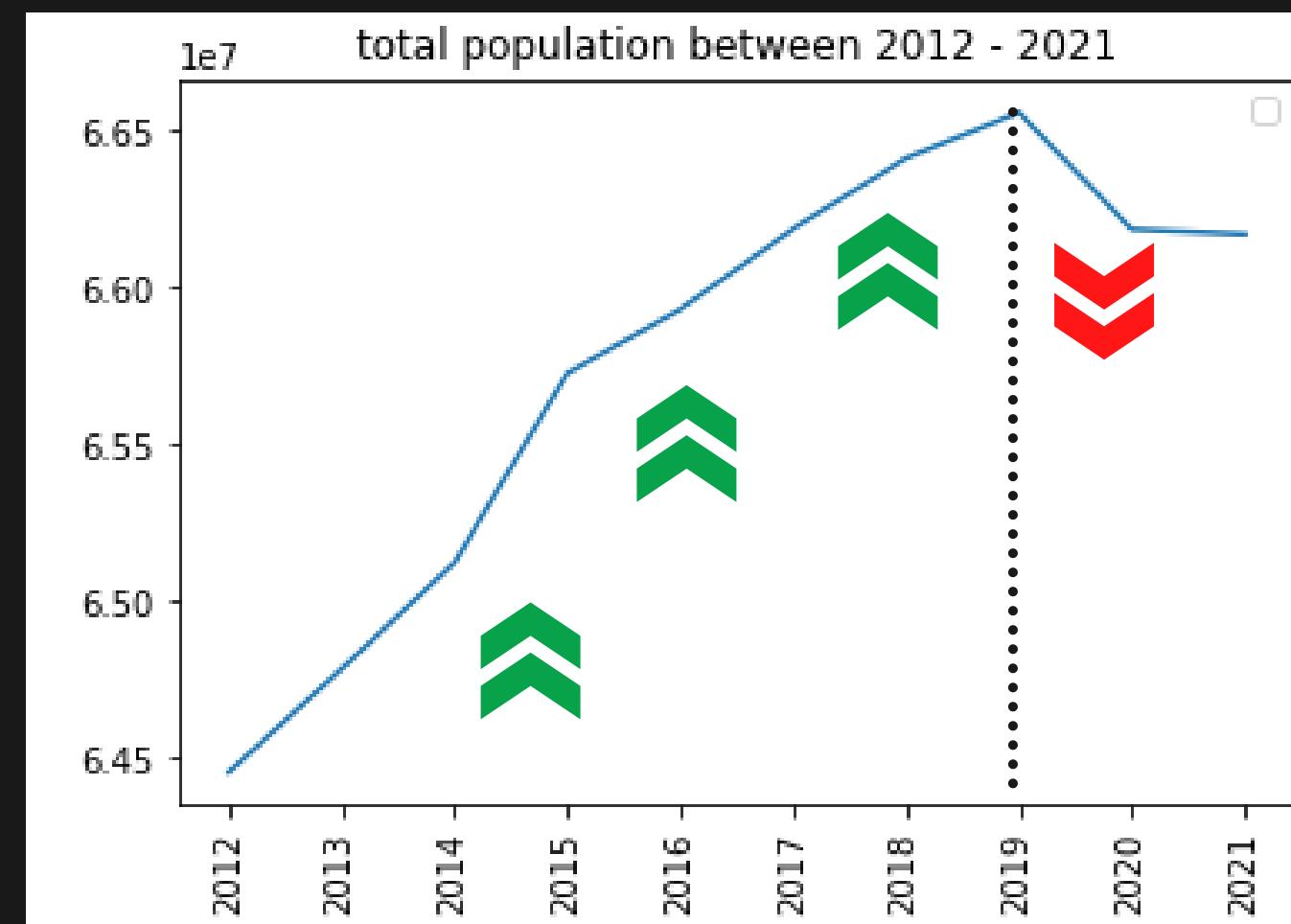
```
55] ...
```

province	Area (km2)	population	GPP(billions)
Krung Thep Maha Nakhon	1564.0	5666264.0	5022.02
Amnat Charoen	3290.0	378438.0	17.65
Ang Thong	950.0	279654.0	27.79
Bueng Kan	4003.0	424091.0	27.17
Buriram	10080.0	1595747.0	84.33

Start Visualizing

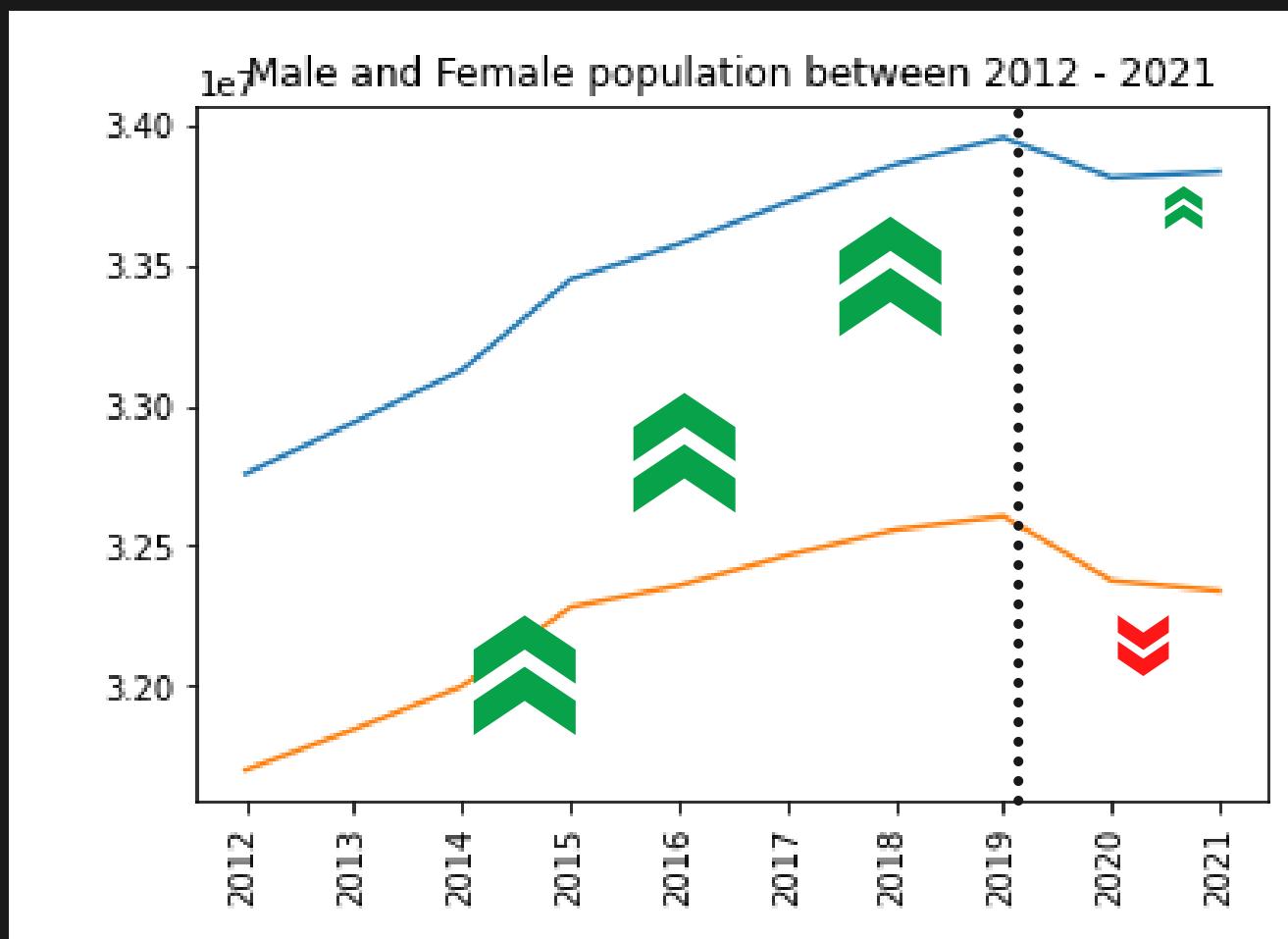
Population 2012-2021?

```
year = ['2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021']
# total-=total.astype(int)
plt.plot(year, total.iloc[:, 1:].sum())
plt.legend()
plt.xticks(rotation=90)
plt.title("total population between 2012 - 2021")
plt.show()
```



Population (Male/Female)

```
year = ['2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021']
# total-=total.astype(int)
plt.plot(year, total.iloc[:, 1:].sum())
plt.legend()
plt.xticks(rotation=90)
plt.title("total population between 2012 - 2021")
plt.show()
```



Male > Female ?
2020 -> 2021 male increase
female still decrease

Data In 2019

- merge Data
- concatenate
- choose top and bottom by GPP
(billion \$)
- Columns: Area, population GPP

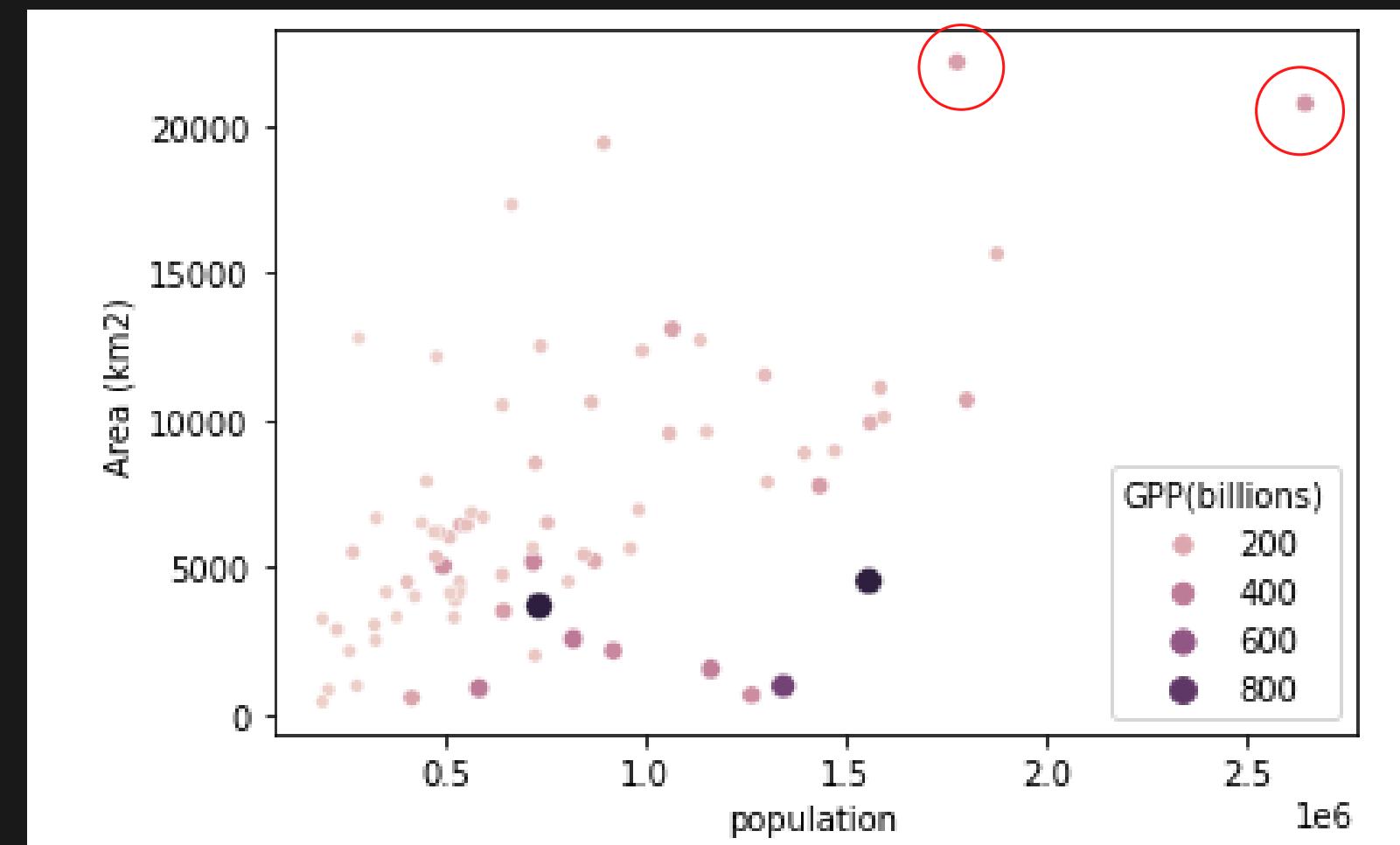
```
scatter = scatter.sort_values('GPP(billions)')
topGPP = scatter.nlargest(11, 'GPP(billions)', keep='all')
bottomGPP = scatter.nsmallest(11, 'GPP(billions)', keep='all')
scatt = pd.concat([topGPP, bottomGPP])
scatt
```

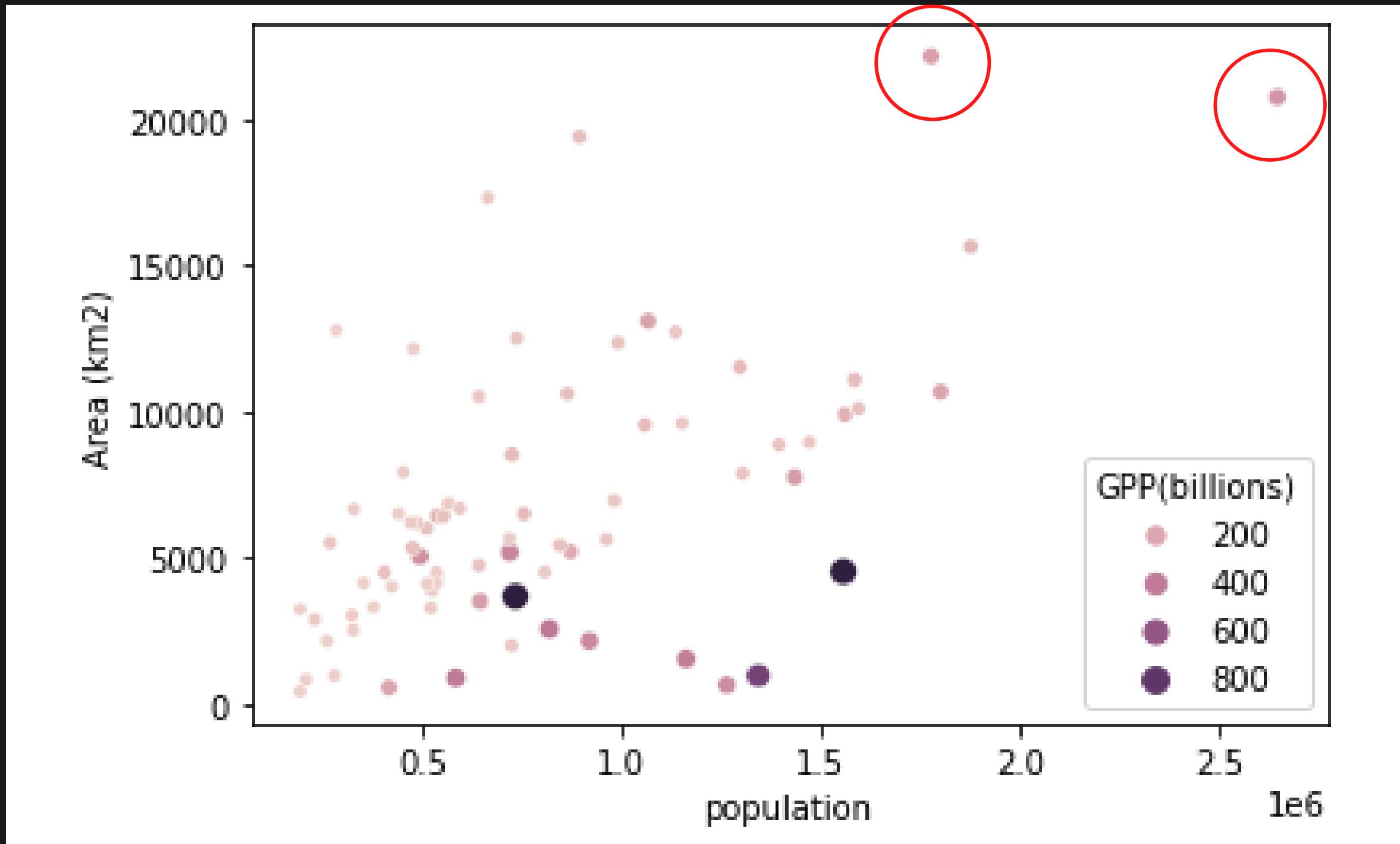
	Area (km2)	population	GPP(billions)
province			
Krung Thep Maha Nakhon	1564.0	5666264.0	5022.02
Rayong	3666.0	734753.0	984.98
Chonburi	4508.0	1558301.0	976.46
Samut Prakan	947.0	1344875.0	717.05
Phra Nakhon Si Ayutthaya	2548.0	820188.0	403.60
Samut Sakhon	866.0	584703.0	398.10
Pathum Thani	1520.0	1163604.0	380.69
Chachoengsao	5169.0	720113.0	341.12
Nakhon Pathom	2142.0	920030.0	332.63
Nonthaburi	637.0	1265387.0	316.63
Prachinburi	5026.0	494680.0	297.25
Mae Hong Son	12765.0	284138.0	13.00
Amnat Charoen	3290.0	378438.0	17.65
Samut Songkhram	414.0	193305.0	21.88
Nong Bua Lamphu	4099.0	512780.0	25.19
Mukdahan	4126.0	353174.0	25.80
Yasothon	4131.0	537299.0	26.04
Sing Buri	817.0	208446.0	26.50
Ranong	3230.0	193370.0	26.77
Nakhon Nayok	2141.0	260751.0	26.84
Bueng Kan	4003.0	424091.0	27.17
Ang Thong	950.0	279654.0	27.79

Data In 2019(Plot)

Use seaborn to virtualize instead of matplotlib
and remove Krung Thep Mahanakon in Graph

```
#plot except KrungThep
sns.scatterplot(data=scatter.sort_values('GPP(billions)', ascending=False).iloc[1:,:], x='population', y='Area (km2)',hue='GPP(billions)', size='GPP(billions)')
```





High Area and Population But Low GPP?

Table of area

Top 10 widest in Thailand

- All of province's name is North Sector only!

1st : Chiangmai

- Have University
- tourist attraction
- HIGH GPP

2nd is a high transportation system

- High transportation system
- Good quality academy
- What about another province?
- Can we do something to increase GDP?

```
scatter.nlargest(10, 'Area (km2)')
```

✓ 0.4s

province	Area (km2)	population	GPP(billions)
Chiang Mai	22135.0	1779254.0	231.73
Nakhon Ratchasima	20736.0	2648927.0	274.90
Kanchanaburi	19385.0	895525.0	97.29
Tak	17303.0	665620.0	47.80
Ubon Ratchathani	15626.0	1878146.0	120.49
Surat Thani	13079.0	1068010.0	211.05
Mae Hong Son	12765.0	284138.0	13.00
Chaiyaphum	12698.0	1137357.0	60.09
Lampang	12488.0	738316.0	68.20
Phetchabun	12340.0	992451.0	76.80

What about another province?

Top 10 widest in Thailand

- All of the rows in the table is a **high terrain**
- low GPP in table 80-90 % is **conservation Area**



Can we do something to increase GDP?

Top 10 widest in Thailand

- Make conservation area to a tourist attraction
- If it already, then need to promote and support by government



Why Thailand is growing too slow

Maybe because of we need support from government

Covid-19?

disparity(education, opportunity, etc)