



กลุ่ม จริ่งใจ

COVID-19 DATA VISUALIZA TION



รายชื่อสมาชิก

ศิริลักษณ์ กัณหารัตน์ STAT

623020451-3

จุฑาภาญจน์ ชิงจันทร์ SI

623020514-5

สุพิชญา ตั้งกิจวานิชย์ SI

623020542-0



วัตถุประสงค์

เพื่อเปรียบเทียบสัดส่วนประชากรผู้ป่วยที่ติดโควิด-19
กับประชากรทั้งหมด ปี พ.ศ.2564 โดยแบ่งตาม 4
ภูมิภาค (การเมือง)

แหล่งข้อมูล



ข้อมูลผู้ป่วยที่ติดโควิด

องค์กร : กรมควบคุมโรค
Government Data of
Thailand จาก :
<https://data.go.th>



ข้อมูลภูมิภาคในประเทศไทย

องค์กร : วิקיพีเดีย จาก :
<https://th.wikipedia.org>



ข้อมูลประชากรแต่ละจังหวัดใน ประเทศไทย ปี 2564

องค์กร : สำนักบริหารการทะเบียน จาก :
<https://stat.bora.dopa.go.th>

การเชื่อมต่อไดรฟ์เพื่อดึงข้อมูลจากไดรฟ์

```
import pandas as pd #เพื่อใช้ทำงานกับข้อมูลที่เป็นรูปแบบตาราง
import os            #เพื่อจัดการกับไฟล์ และ ตัว path ที่เราจะนำเข้า
```

```
from google.colab import drive#เชื่อมต่อข้อมูล เพราะเราจะทำการนำเข้าข้อมูล โดยให้ดึงมาจาก google drive ของเรา
drive.mount('/content/drive')
```

Mounted at /content/drive

```
path = '/content/drive/My Drive/Data_Viz_Project'
#โดยเราจะทำการระบุให้ชี้ไปที่ โฟลเดอร์ที่เราเซฟไฟล์ไว้ หรือก็คือ โฟลเดอร์ Project
```





ตารางข้อมูล

การนำเข้าตารางข้อมูล

ตารางที่ 1 ตาราง confirmed cases ชุดที่ 1



```
[ ] conf_1 = os.path.join(path, 'confirmed-cases-1.csv')
conf_1 = pd.read_csv(conf_1, parse_dates=['announce_date'])
conf_1
```

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:2882: DtypeWarning: Columns (3) have mixed types.Specify dtype in read_csv call to avoid this warning.

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation	risk	province_of_onset
0	1	2020-12-01	หญิง	61.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
1	2	2020-01-17	หญิง	74.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
2	3	2020-01-22	หญิง	73.0	ปี	Thailand	นครปฐม	คนต่างชาติเดินทางมาจากต่างประเทศ	นครปฐม
3	4	2020-01-22	ชาย	68.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร
4	5	2020-01-24	หญิง	66.0	ปี	China	นนทบุรี	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร

ตารางที่ 2 ตาราง confirmed cases ชุดที่ 2

```
[ ] conf_2 = os.path.join(path, 'confirmed-cases-2.csv')
conf_2 = pd.read_csv(conf_2, parse_dates=['announce_date'])
conf_2
```

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation	risk
0	816990	2021-12-08	ชาย	7.0	ปี	Thailand	เชียงราย	อื่นๆ
1	816991	2021-12-08	ชาย	1.0	ปี	Thailand	เชียงราย	อื่นๆ
2	816992	2021-12-08	ชาย	35.0	ปี	Thailand	เชียงราย	อื่นๆ
3	816993	2021-12-08	หญิง	33.0	ปี	Thailand	เชียงราย	อื่นๆ



ตารางที่ 3 ตาราง confirmed cases ชุดที่ 3

```
[ ] conf_3 = os.path.join(path, 'confirmed-cases-3.csv')
conf_3 = pd.read_csv(conf_3, parse_dates=['announce_date'])
conf_3
```

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:2882: DtypeWarning: Fields (int64, bool) have dtype incompatible with float64, overriding

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation
0	1859158	2021-10-26	ชาย	20.0	ปี	Thailand	ลำปาง
1	1859159	2021-10-26	หญิง	42.0	ปี	Thailand	ลำปาง
2	1859160	2021-10-26	หญิง	33.0	ปี	Thailand	ลำปาง

ตารางที่ 4 ตาราง confirmed cases ชุดที่ 4

```
conf_4 = os.path.join(path, 'confirmed-cases-4.csv')
conf_4 = pd.read_csv(conf_4, parse_dates=['announce_date'])
conf_4
```

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation
0	2869617	2022-02-28	ชาย	41.0	ปี	Thailand	พะเยา
1	2869618	2022-02-28	หญิง	45.0	ปี	Thailand	พะเยา
2	2869619	2022-02-28	หญิง	40.0	ปี	Thailand	พะเยา
3	2869620	2022-02-28	ชาย	44.0	ปี	Thailand	พะเยา

ตารางที่ 5 ตารางจำนวนประชากรในประเทศไทย

```
prop = pd.read_csv(os.path.join(path, 'prop.csv'))
prop
```

	ชื่อจังหวัด	จำนวนประชากรชาย	จำนวนประชากรหญิง	จำนวนประชากรทั้งหมด
0	กรุงเทพมหานคร	2,592,292	2,935,702	5,527,994
1	สมุทรปราการ	645,884	710,565	1,356,449
2	นนทบุรี	599,167	689,470	1,288,637
3	ปทุมธานี	563,851	626,209	1,190,060
4	พระนครศรีอยุธยา	394,024	426,488	820,512

DATA PREPROCESSING



เชื่อมตารางข้อมูลผู้ป่วยที่ติดโควิด



โดยใช้ .concat เพื่อเชื่อมตารางแนวนิ่ง

ตารางที่ 6 ตารางข้อมูลผู้ป่วย

โดยจะทำการรวมตาราง

```
data_con = pd.concat([conf_1, conf_2, conf_3, conf_4])
data_con
```

	No.	announce_date	sex	age	Unit	nationality	province_of_isolation	
0	1	2020-12-01	หญิง	61.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทาง
1	2	2020-01-17	หญิง	74.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทาง
2	3	2020-01-22	หญิง	73.0	ปี	Thailand	นครปฐม	คนต่างชาติเดินทาง
3	4	2020-01-22	ชาย	68.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติเดินทาง
4	5	2020-01-24	หญิง	66.0	ปี	China	นนทบุรี	คนต่างชาติเดินทาง



เชื่อมตารางข้อมูลผู้ป่วยกับภูมิภาค



โดยใช้ .merge เพื่อเชื่อมตารางแนวนอน โดยใช้ province_of_onset เชื่อมกับ จังหวัด

```
merged_01 = data_con.merge(dt_re,how='left',left_on='province_of_onset',right_on='จังหวัด')
merged_01
```



	No.	announce_date	sex	age	Unit	nationality	province_of_isolation	risk	province_c
0	1	2020-12-01	หญิง	61.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติ เดินทางมา จากต่าง ประเทศ	กรุงเทพ
1	2	2020-01-17	หญิง	74.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติ เดินทางมา จากต่าง ประเทศ	กรุงเทพ
2	3	2020-01-22	หญิง	73.0	ปี	Thailand	นครปฐม	คนต่างชาติ เดินทางมา จากต่าง ประเทศ	
3	4	2020-01-22	ชาย	68.0	ปี	China	กรุงเทพมหานคร	คนต่างชาติ เดินทางมา จากต่าง ประเทศ	กรุงเทพ
4	5	2020-01-24	หญิง	66.0	ปี	China	นนทบุรี	คนต่างชาติ เดินทางมา จากต่าง ประเทศ	กรุงเทพ

การกำจัด MISSING



```
[ ] merged_01.shape  
  
(3325951, 15)
```

ตรวจสอบขนาด

```
[ ] data_drop = merged_01.dropna()  
data_drop
```

ใช้ dropna เพื่อต้องการลบข้อมูลทั้งแถวออกเมื่อมี
คอลัมน์ใดคอลัมน์หนึ่งเป็น NaN

```
merged_01.isnull().any()  
  
No. False  
announce_date False  
sex True  
age True  
Unit True  
nationality True  
province_of_isolation True
```

#.isnull() เพื่อระบุว่านิพจน์มีค่า Null หรือไม่

#.any() เพื่อดูรายละเอียดการเปรียบเทียบข้อมูลให้ออกมาในรูปแบบ true false

```
[ ] removes = merged_01.shape[0] - data_drop.shape[0]  
print(f'size before drop = {merged_01.shape[0]}')  
print(f'size after drop = {data_drop.shape[0]}')  
print(f'% from drop = {(removes/merged_01.shape[0])*100}')  
#จากการคำนวณพบว่ามีข้อมูลที่หายไปถึง 27.48%
```

```
size before drop = 3325951  
size after drop = 2411894  
% from drop = 27.4825756603149
```

จากผลเราจะพิจารณาจากการอ่าน
ค่า True และ False ถ้า True คือ
มีค่า missing ดังนั้น จะเห็นได้ว่า
ข้อมูลมีค่า missing เกือบในทุก
คอลัมน์ ดังนั้น สามารถอนุมานได้
เลยว่า ข้อมูลมีค่า missing
จำนวนมาก จึงไม่เหมาะกับตัด
ข้อมูล หรือ .dropna จึงพิจารณา
ให้ใช้วิธีแทนค่าเฉพาะคอลัมน์ข้อมูล
ที่เราต้องการใช้

No.	False
announce_date	False
sex	True
age	True
Unit	True
nationality	True
province_of_isolation	True
risk	True
province_of_onset	True
district_of_onset	True
จังหวัด	True
6 ภูมิภาค (ภูมิศาสตร์)	True
4 ภูมิภาค (การเมือง)	True
6 ภูมิภาค (อุดมศึกษา)	True
5 ภูมิภาค (การท่องเที่ยว)	True
dtype: bool	

```
merged_01.isnull().sum() #sum หาคอลัมน์
```

No.	0
announce_date	0
sex	0
age	268183
Unit	0
nationality	0
province_of_isolation	3956
risk	0
province_of_onset	0
district_of_onset	502847

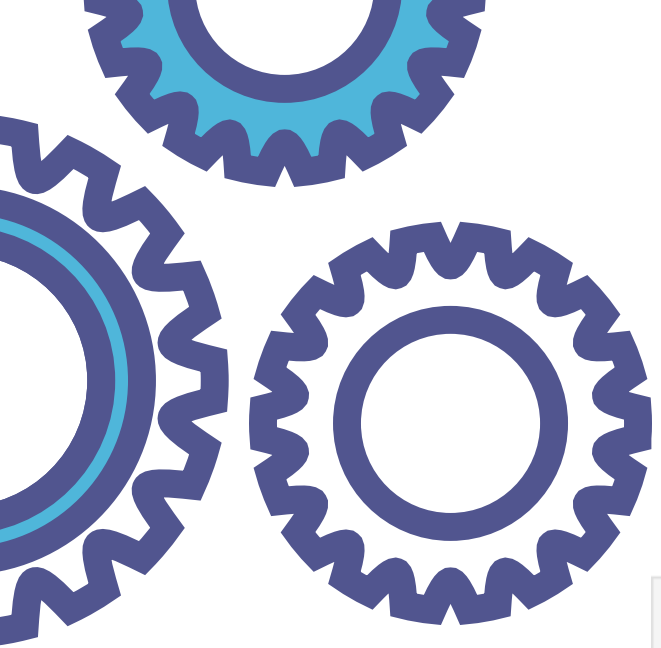
```
Age = merged_01[['age']].mean() #กำหนดให้ Age เป็นค่าเฉลี่ยของอายุ  
merged_01[['age']] = merged_01[['age']].fillna(Age) #แทนค่าเฉลี่ยลงในข้อมูลที่ว่าง
```

```
merged_01[['sex']] = merged_01[['sex']].fillna('ไม่ระบุ')  
#แทนค่า Missing ด้วยคำว่า ไม่ระบุ
```

```
merged_01[['Unit']] = merged_01[['Unit']].fillna('ปี')  
#แทนค่า Missing ด้วยคำว่า ไม่ระบุ
```

```
merged_01[['nationality']] = merged_01[['nationality']].fillna('ไม่ระบุ')  
#แทนค่า Missing ด้วยคำว่า ไม่ระบุ
```





การตัดเฉพาะข้อมูลที่ต้องการใช้

```
data_co = merged_01[['No.', 'announce_date', 'sex', 'age', 'Unit', 'nationality', 'risk', 'province_of_onset', 'จังหวัด', '4 ภูมิภาค (การเมือง)']]
data_co
```

	No.	announce_date	sex	age	Unit	nationality	risk	province_of_onset	จังหวัด	4 ภูมิภาค (การเมือง)
0	1	2020-12-01	หญิง	61.0	ปี	China	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง
1	2	2020-01-17	หญิง	74.0	ปี	China	คนต่างชาติเดินทางมาจากต่างประเทศ	กรุงเทพมหานคร	กรุงเทพมหานคร	กลาง

```
removes = data_co.shape[0] - data_co.dropna().shape[0]
print(f'size before drop = {data_co.shape[0]}')
print(f'size after drop = {data_co.dropna().shape[0]}')
print(f'% from drop = {(removes/data_co.shape[0])*100}')
#จากการคำนวณพบว่ามีข้อมูลที่หายไป 8%
```

```
size before drop = 3325951
size after drop = 3057768
% from drop = 8.06334789658657
```

```
data_covid = data_co.dropna()
data_covid.isnull().sum()
```

```
No.
announce_date
sex
age
Unit
nationality
risk
province of onset
```





จัดการค่าข้อมูลในตารางที่มีค่าซ้ำ

```
set(data_64['sex'])
```

```
{'ชาย', 'นาย', 'หญิง', 'ไม่ระบุ'}
```

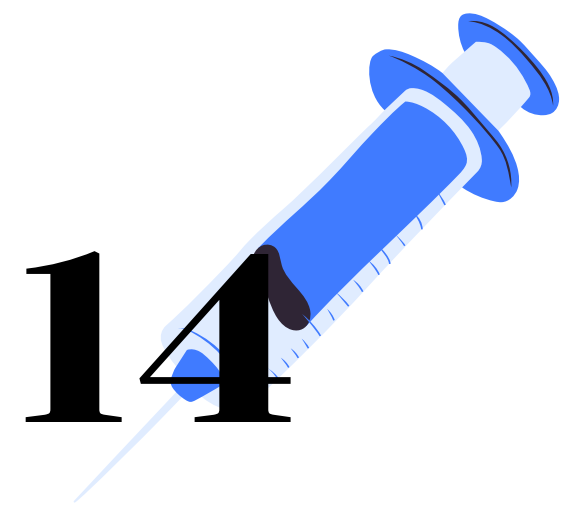
```
data_64['sex'].replace({'นาย':'ชาย'},inplace=True)
```

```
set(data_64['sex'])
```

```
{'ชาย', 'หญิง', 'ไม่ระบุ'}
```

```
data_64['province_of_onset'].replace({  
'\xa0กาญจนบุรี':'กาญจนบุรี',  
'\xa0นครปฐม':'นครปฐม',  
'\xa0สมุทรสงคราม':'สมุทรสงคราม',  
'\xa0สมุทรสาคร':'สมุทรสาคร',  
'\xa0อุทัยธานี':'อุทัยธานี',  
'กทม':'กรุงเทพมหานคร',  
'กทม.':':'กรุงเทพมหานคร',
```

จัดการกับค่าข้อมูลที่ไม่ถูกต้อง เช่น ชื่อจังหวัดที่ซ้ำ และ
เพศ โดยการแทนด้วยค่าที่ถูกต้องลงไปด้วยฟังก์ชัน
.replace()



จัดการข้อมูลเพศชายและเพศหญิง



ใช้ค่าข้อมูลตามเพศแล้ว ให้จัดกลุ่มข้อมูลตาม province_of_onset เพื่อหาผลรวมของผู้ป่วยแต่ละเพศในแต่ละจังหวัด

```
Data_male = data_64[data_64['sex'] == 'ชาย']  
Data_male
```

```
Data_Female = data_64[data_64['sex'] == 'หญิง']  
Data_Female
```

```
M_gb = Data_male.groupby('province_of_onset')['sex'].count()
```

```
M_df = pd.DataFrame(M_gb)  
M_noindex = M_df.reset_index()  
M_noindex
```

```
male = {'sex': 'ผู้ติดเชื้อเพศชาย'}  
# call rename () method  
M_df.rename(columns=male,  
            inplace=True)
```

```
M_df_noindex = M_df.reset_index()  
M_df_noindex
```

```
FM_gb = Data_Female.groupby('province_of_onset')['sex'].count()  
FM_df = pd.DataFrame(FM_gb)  
FM_noindex = FM_df.reset_index()  
FM_noindex
```

```
Famale = {'sex': 'ผู้ติดเชื้อเพศหญิง'}  
# call rename () method  
FM_df.rename(columns=Famale,  
            inplace=True)
```

```
FM_df_noindex = FM_df.reset_index()  
FM_df_noindex
```

เมื่อได้ข้อมูลแต่ละเพศตามต้องการแล้ว ต่อมานำมาต่อเป็นตารางเดียวกันโดยใช้ `province_of_onset` เชื่อม



```
data_onset = M_df_noindex.merge(FM_df_noindex,how='left',left_on='province_of_onset',right_on='province_of_onset')
data_onset
```

	province_of_onset	ผู้ติดเชื้อเพศชาย	ผู้ติดเชื้อเพศหญิง
0	กระบี่	4048	4713
1	กรุงเทพมหานคร	182945	188872
2	กาญจนบุรี	6972	8085
3	กาฬสินธุ์	3417	3714

```
data_onset.isnull().any()
```

```
province_of_onset    False
ผู้ติดเชื้อเพศชาย    False
ผู้ติดเชื้อเพศหญิง    False
dtype: bool
```

```
Province = {'province_of_onset':'ชื่อจังหวัด'}
# call rename () method
data_onset.rename(columns=Province,
                  inplace=True)
data_onset
```

	ชื่อจังหวัด	ผู้ติดเชื้อเพศชาย	ผู้ติดเชื้อเพศหญิง
0	กระบี่	4048	4713



รวมตารางข้อมูลทั้งหมด
ที่ต้องการนำไปวิเคราะห์

เชื่อมตารางข้อมูลประชากรกับข้อมูลผู้ติดเชื้อที่แบ่งตามเพศโดยใช้ชื่อจังหวัดเชื่อมตาราง

```
data_compare = prop.merge(data_onset,how='left',left_on='ชื่อจังหวัด',right_on='ชื่อจังหวัด')
data_compare
```

กำหนดตารางข้อมูลที่มีแค่ข้อมูลจังหวัดและ 4 ภูมิภาค

```
data_re = dt_re[['จังหวัด','4 ภูมิภาค (การเมือง)']]
data_re
```

เชื่อมตารางภูมิภาคและข้อมูลผู้ป่วยเข้าด้วยกันโดยใช้จังหวัดและชื่อจังหวัด

```
data_use = data_re.merge(data_compare,how='left',left_on='จังหวัด',right_on='ชื่อจังหวัด')
data_use
```

กำหนดข้อมูลในตารางโดยใช้ .iloc ให้ตัดแต่ละคอลัมน์

```
data_use2 = data_use.iloc[:,[0,1,3,4,5,6,7]]
data_use2
```

	จังหวัด	4 ภูมิภาค (การเมือง)	จำนวนประชากรชาย	จำนวนประชากรหญิง	จำนวนประชากรทั้งหมด	ผู้ติดเชื้อเพศชาย	ผู้ติดเชื้อเพศหญิง	จำนวนผู้ติดเชื้อ
0	อำนาจเจริญ	ตะวันออกเฉียงเหนือ	187122.0	189228.0	376350.0	214.0	263.0	477.0
1	บึงกาฬ	ตะวันออกเฉียงเหนือ	211494.0	210501.0	421995.0	624.0	659.0	1283.0
2	บุรีรัมย์	ตะวันออกเฉียงเหนือ	783531.0	796274.0	1579805.0	367.0	343.0	710.0
3	ชัยภูมิ	ตะวันออกเฉียงเหนือ	554299.0	567966.0	1122265.0	736.0	519.0	1255.0
4	กาฬสินธุ์	ตะวันออกเฉียงเหนือ	481988.0	493582.0	975570.0	3417.0	3714.0	7131.0
...
72	พังงา	ใต้	133785.0	134231.0	268016.0	2614.0	2626.0	5240.0
73	ภูเก็ต	ใต้	197874.0	220911.0	418785.0	7753.0	8911.0	16664.0
74	ระนอง	ใต้	98136.0	96437.0	194573.0	4538.0	5302.0	9840.0
75	สตูล	ใต้	161640.0	163195.0	324835.0	3925.0	4780.0	8705.0
76	ตรัง	ใต้	312693.0	327095.0	639788.0	7199.0	9228.0	16427.0

การสร้างคอลัมน์รวมจำนวนผู้ติดเชื้อโควิด

```
data_use2['จำนวนผู้ติดเชื้อ'] = data_use2.ผู้ติดเชื้อเพศชาย + data_use2.ผู้ติดเชื้อเพศหญิง
data_use2
```

ทำตารางเป็นไฟล์ CSV

```
data_use2.to_csv(os.path.join(path, 'data_use2.csv'))
```




DASHBOARD

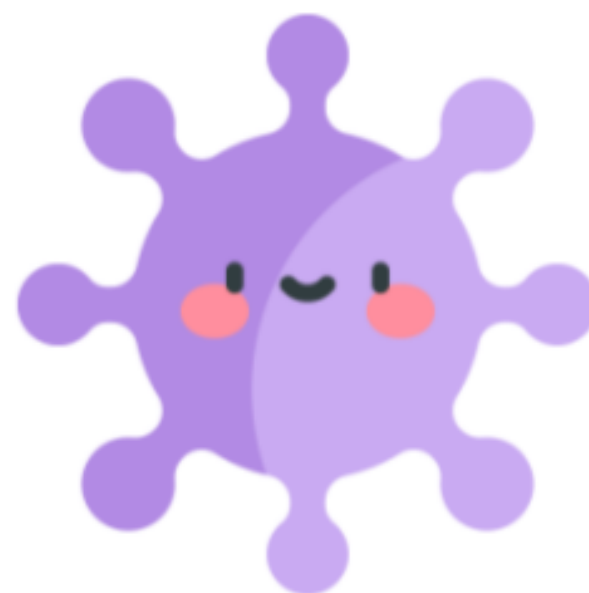
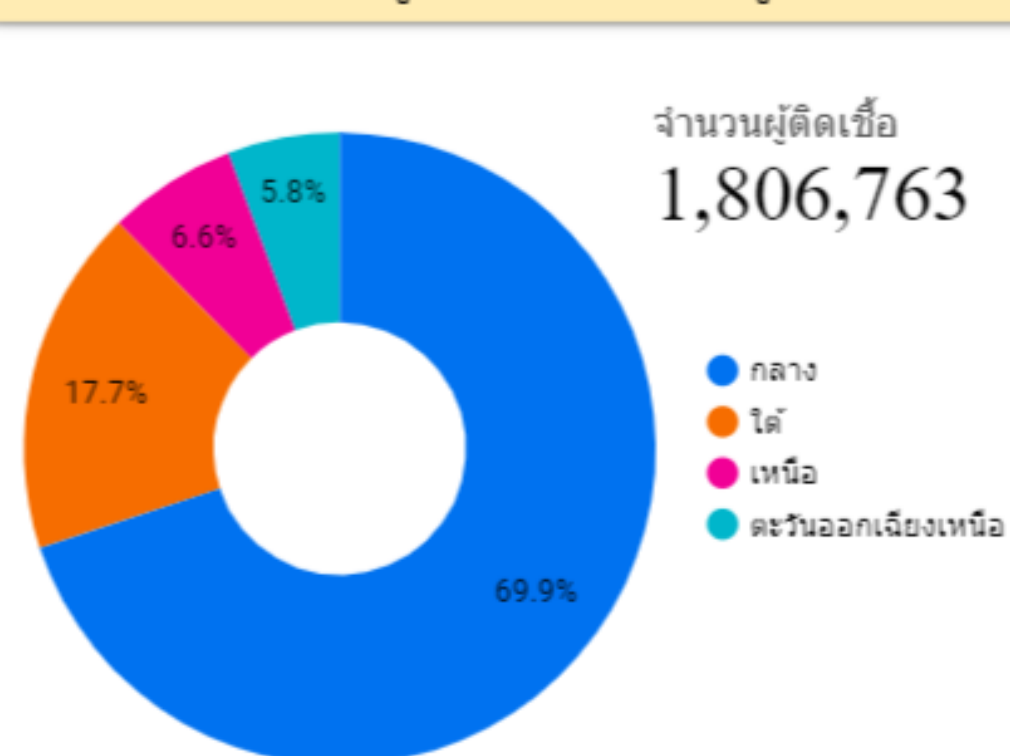
<https://datastudio.google.com/u/0/reporting/20ee2eeb-6782-4959-900d-70f3644ed291/page/PjjoC>



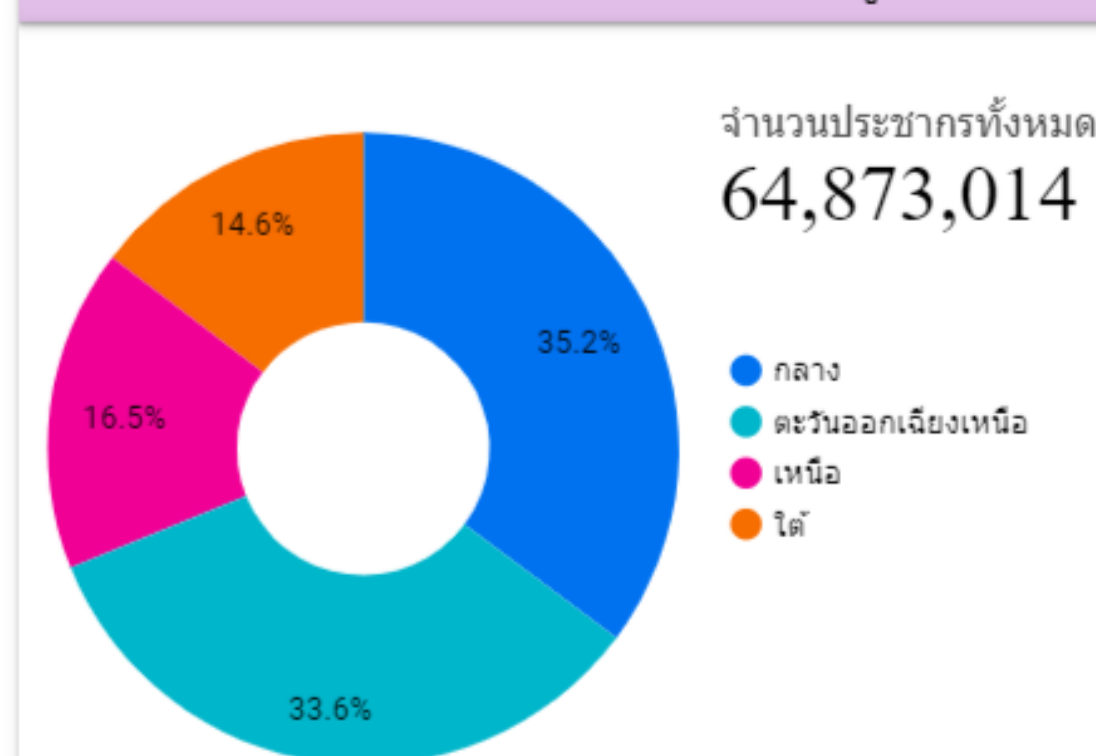
การเปรียบเทียบสัดส่วนประชากรผู้ป่วยที่ติดโควิด-19 กับประชากรทั้งหมด ปี พ.ศ.2564 โดยแบ่งตาม 4 ภูมิภาค (การเมือง)

4 ภูมิภาค (การเมือง)

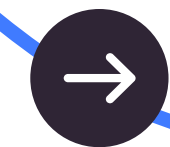
จำนวนผู้ติดเชื้อ แยกตามภูมิภาค



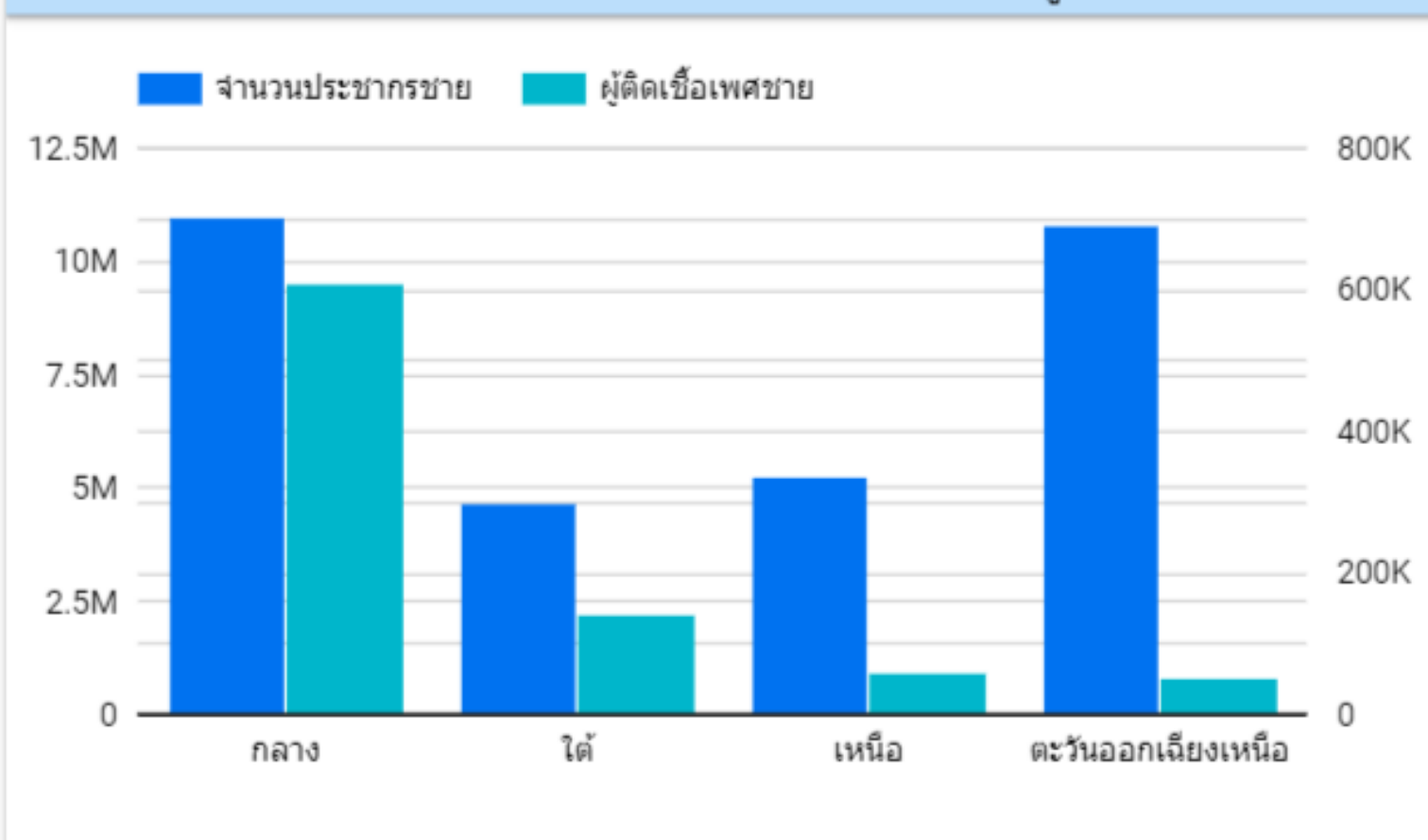
จำนวนประชากรทั้งหมด แยกตามภูมิภาค



จากแผนภาพวงกลมทั้งสอง จะพบว่าแม้จำนวนประชากรของภาคกลางมีขนาดใหญ่เคียงกับภาคตะวันออกเฉียงเหนือ แต่จำนวนผู้ที่ติดเชื้อ COVID-19 กลับมีจำนวนมากกว่าภาคตะวันออกเฉียงเหนือถึง 12 เท่า

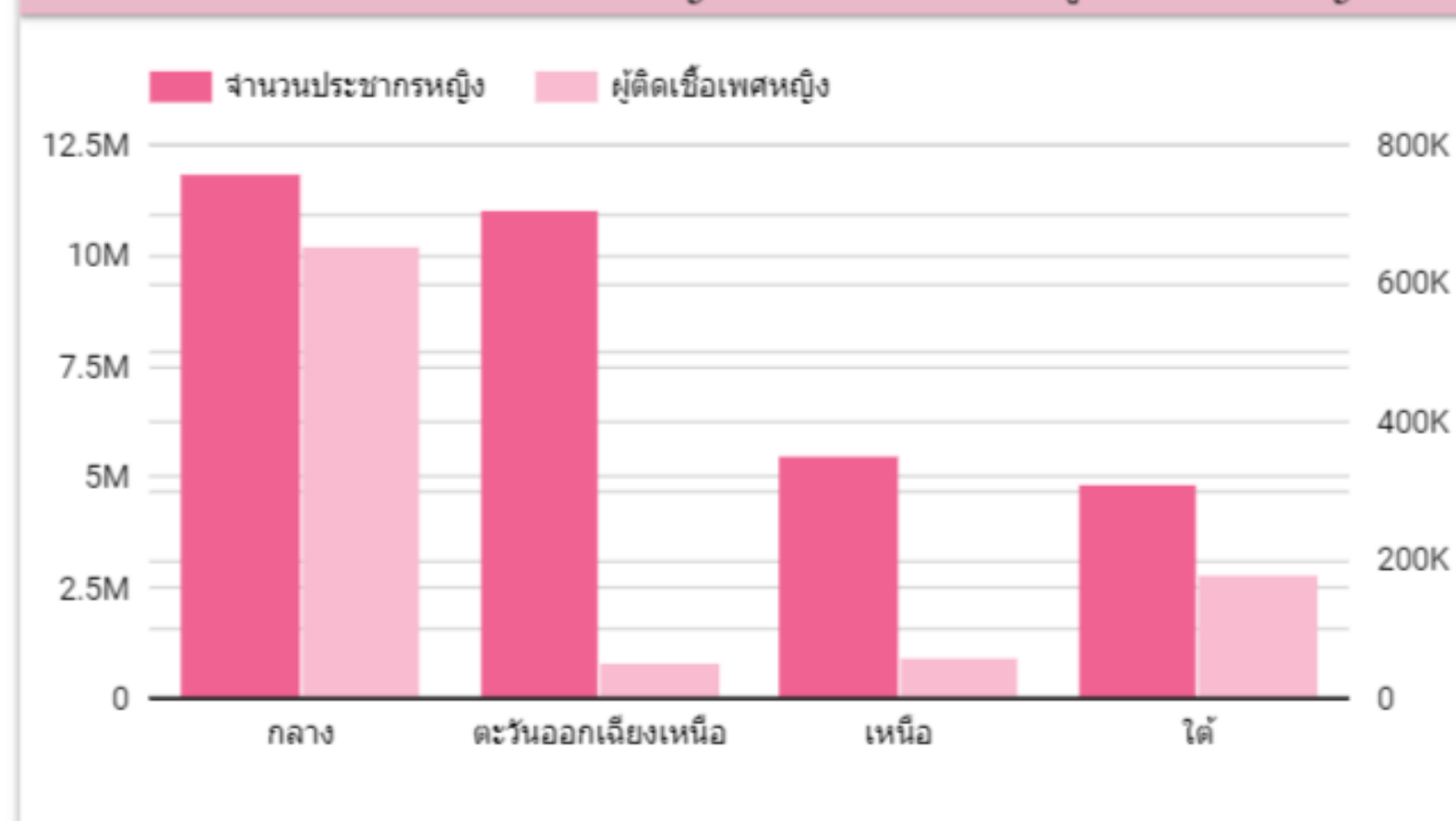


เปรียบเทียบจำนวนประชากรชายทั้งหมดกับจำนวนผู้ติดเชื้อเพศชาย

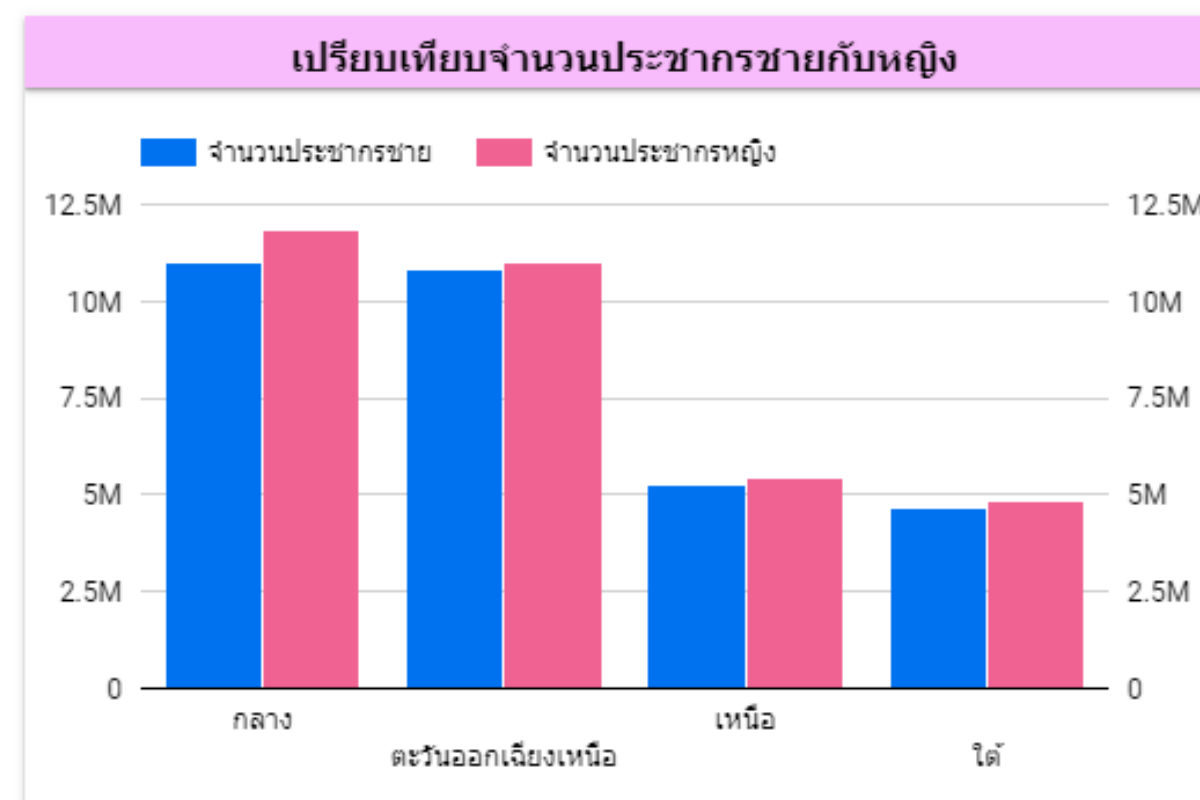
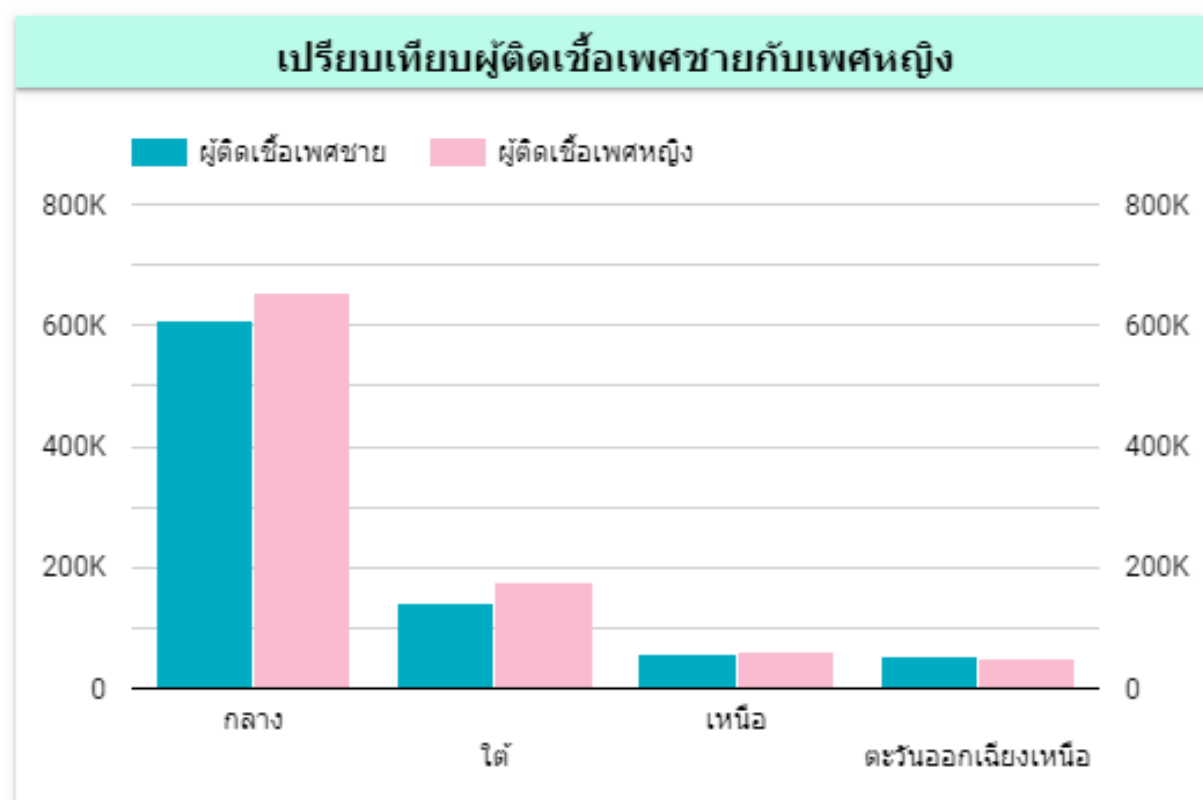


ภูมิภาคที่มีจำนวนผู้ติดเชื้อเพศชายมากที่สุดคือ
ภาคกลาง ซึ่งมีผู้ติดเชื้อจำนวน 6 แสนคน
ในจำนวนประชากรเพศชายทั้งหมดประมาณ 11 ล้านคน

เปรียบเทียบจำนวนประชากรหญิงทั้งหมดกับจำนวนผู้ติดเชื้อเพศหญิง :



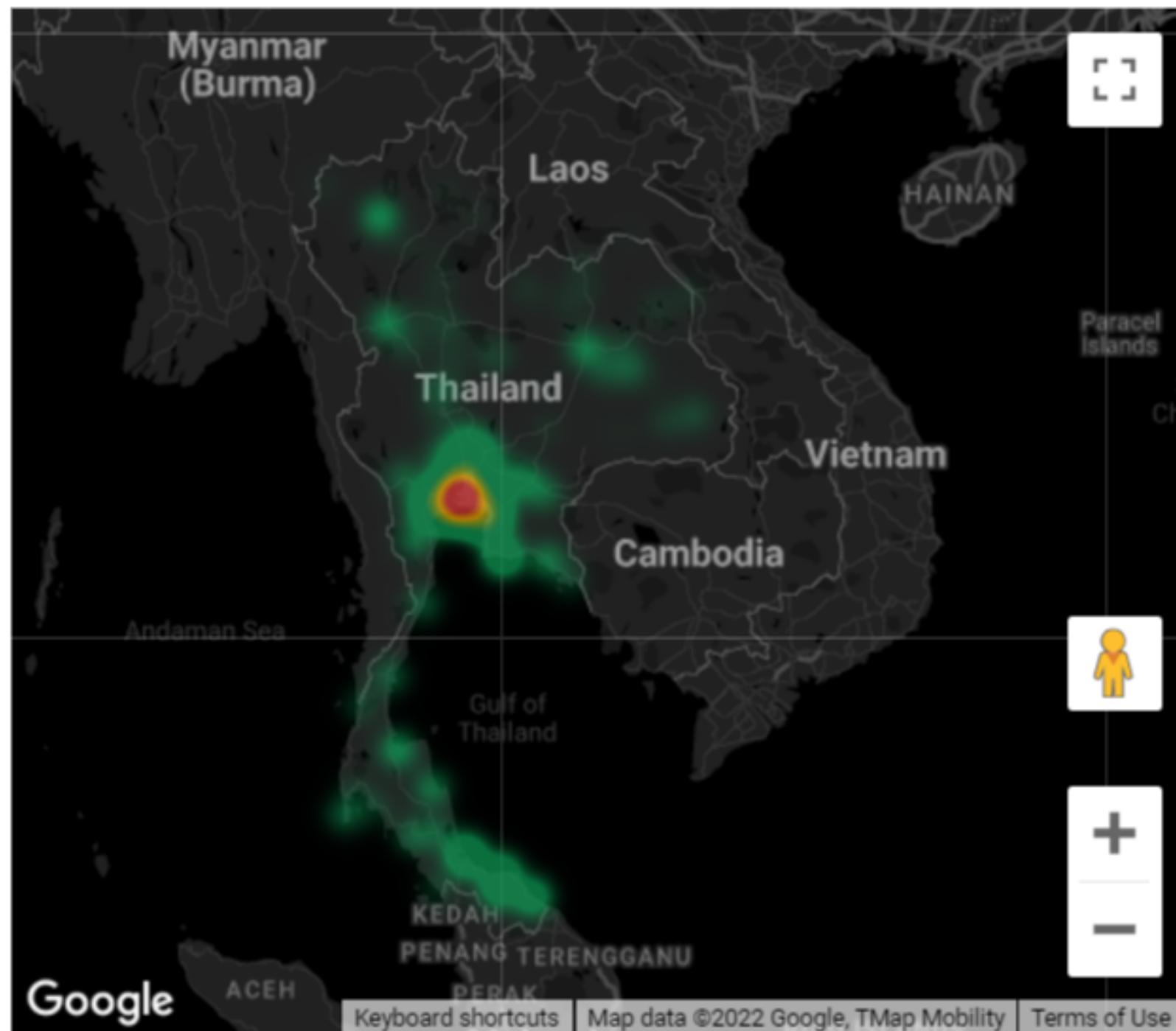
ภูมิภาคที่มีจำนวนผู้ติดเชื้อเพศหญิงมากที่สุดคือ
ภาคกลาง ซึ่งมีผู้ติดเชื้อจำนวน 6 แสน 5 หมื่นคน
ในจำนวนประชากรเพศหญิงทั้งหมดประมาณ 11 ล้านคน



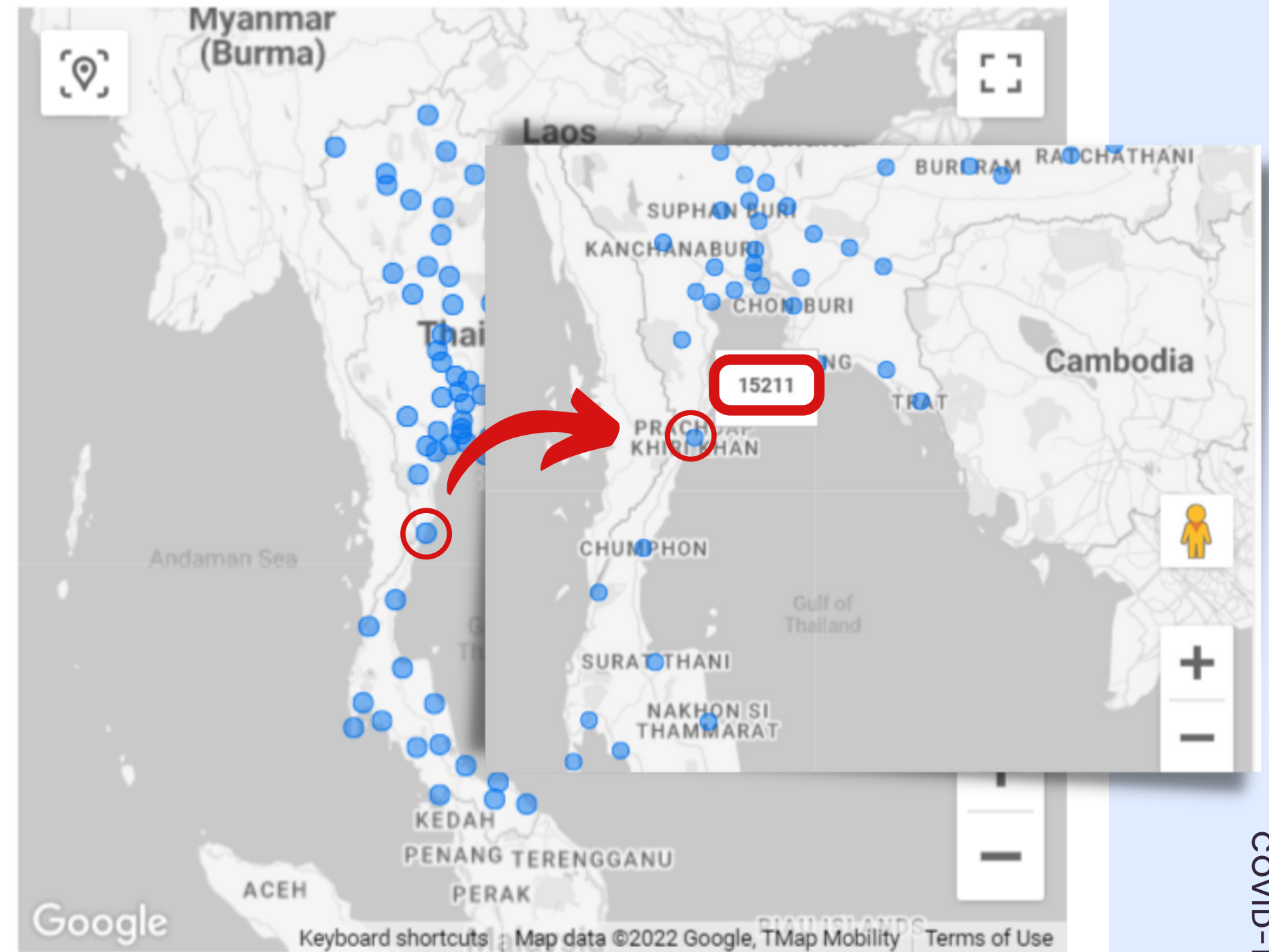
ตารางแสดงข้อมูลของจำนวนประชากรและจำนวนผู้ติดเชื้อ ในปี 2564 โดยแบ่งตาม 4 ภูมิภาค (การเมือง)						
4 ภูมิภาค (การเมือง) -	ผู้ติดเชื้อเพศชาย	ผู้ติดเชื้อเพศหญิง	จำนวนผู้ติดเชื้อ	จำนวนประชากรชาย	จำนวนประชากรหญิง	จำนวนประชากรทั้งหมด
1. ใต้	142,772	177,786	320,558	4,667,882	4,824,385	9,492,267
2. เหนือ	58,243	60,905	119,148	5,240,839	5,470,760	10,711,599
3. ตะวันออกเฉียงเหนือ	53,042	50,879	103,921	10,814,540	11,012,380	21,826,920
4. กลาง	607,789	655,347	1,263,136	10,984,989	11,857,239	22,842,228

พบว่าในส่วนของภาคตะวันออกเฉียงเหนือแม้จำนวนประชากรเพศชายจะน้อยกว่าเพศหญิง แต่กลับมีจำนวนผู้ติดเชื้อเพศชายสูงกว่าผู้ติดเชื้อเพศหญิง ซึ่งแตกต่างจากภูมิภาคอื่นๆ ที่จำนวนประชากรเพศหญิงมากกว่าเพศชาย แล้วจำนวนผู้ติดเชื้อเพศหญิงก็มากกว่าผู้ติดเชื้อเพศชายเช่นเดียวกัน

แผนที่แสดงจำนวนผู้ติดเชื้อในประเทศไทย



จำนวนผู้ติดเชื้อ 0  371,817



23

แสดงให้เห็นถึงความหนาแน่นของผู้ติดเชื้อ

แสดงให้เห็นจำนวนผู้ติดเชื้อผู้ติดเชื้อในแต่ละจังหวัด



THANK YOU

member



ศิริลักษณ์ กัณหารัตน์

623020451-3

จุฑาภาญจน์ ชิงจันทร์

623020514-5

สุพิชญา ตั้งกิจวานิชย์

623020542-0