In [16]:
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("number-of-new-hiv-infections-per-year.csv")

df
```

Out[16]:

	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
0	Afghanistan	AFG	1990	72.77185
1	Afghanistan	AFG	1991	80.25966
2	Afghanistan	AFG	1992	100.71334
3	Afghanistan	AFG	1993	131.88900
4	Afghanistan	AFG	1994	162.07616
...	...	...	...	...
6206	Zimbabwe, Midlands	NaN	2019	4097.12350
6207	Zimbabwe, Midlands	NaN	2020	3678.44630
6208	Zimbabwe, Midlands	NaN	2021	3038.60060
6209	Zimbabwe, Midlands	NaN	2022	2826.40280
6210	Zimbabwe, Midlands	NaN	2023	2712.18260

6211 rows × 4 columns

In [42]:
```
df_THA = df[(df['Entity'] == 'Thailand')].reset_index()
df_THA
```

Out[42]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
0	5376	Thailand	THA	1990	160193.4000
1	5377	Thailand	THA	1991	174935.3900
2	5378	Thailand	THA	1992	160979.6400
3	5379	Thailand	THA	1993	133515.0000
4	5380	Thailand	THA	1994	112095.5400
5	5381	Thailand	THA	1995	94397.1300
6	5382	Thailand	THA	1996	80494.7600
7	5383	Thailand	THA	1997	70288.0100
8	5384	Thailand	THA	1998	63365.2030
9	5385	Thailand	THA	1999	57153.8240
10	5386	Thailand	THA	2000	50906.3050
11	5387	Thailand	THA	2001	45762.4060
12	5388	Thailand	THA	2002	40961.0660
13	5389	Thailand	THA	2003	36697.4800
14	5390	Thailand	THA	2004	31478.9570
15	5391	Thailand	THA	2005	27667.3980
16	5392	Thailand	THA	2006	24539.1200
17	5393	Thailand	THA	2007	22092.6910
18	5394	Thailand	THA	2008	20161.3220
19	5395	Thailand	THA	2009	18694.3400
20	5396	Thailand	THA	2010	17666.6000
21	5397	Thailand	THA	2011	16637.2460

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
22	5398	Thailand	THA	2012	15440.8950
23	5399	Thailand	THA	2013	14499.5510
24	5400	Thailand	THA	2014	13912.5890
25	5401	Thailand	THA	2015	12902.3150
26	5402	Thailand	THA	2016	11988.1320
27	5403	Thailand	THA	2017	11313.7705
28	5404	Thailand	THA	2018	10710.1750
29	5405	Thailand	THA	2019	10305.6350
30	5406	Thailand	THA	2020	9950.6610
31	5407	Thailand	THA	2021	9654.7870
32	5408	Thailand	THA	2022	9380.5160
33	5409	Thailand	THA	2023	9081.2430

In [18]:
```
df_v2 = df[(df['Entity'] == 'Philippines')].reset_index()
df_v2
```

Out[18]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
0	4528	Philippines	PHL	1990	19.03207
1	4529	Philippines	PHL	1991	23.46428
2	4530	Philippines	PHL	1992	35.85388
3	4531	Philippines	PHL	1993	53.11289
4	4532	Philippines	PHL	1994	75.46854
5	4533	Philippines	PHL	1995	94.72833
6	4534	Philippines	PHL	1996	110.61159
7	4535	Philippines	PHL	1997	131.88924
8	4536	Philippines	PHL	1998	163.44363
9	4537	Philippines	PHL	1999	215.60551
10	4538	Philippines	PHL	2000	289.71146
11	4539	Philippines	PHL	2001	386.81253
12	4540	Philippines	PHL	2002	513.78784
13	4541	Philippines	PHL	2003	683.09140
14	4542	Philippines	PHL	2004	875.46246
15	4543	Philippines	PHL	2005	1125.47810
16	4544	Philippines	PHL	2006	1448.49440
17	4545	Philippines	PHL	2007	2106.63570
18	4546	Philippines	PHL	2008	3212.19000
19	4547	Philippines	PHL	2009	3977.55130
20	4548	Philippines	PHL	2010	4444.38870
21	4549	Philippines	PHL	2011	5284.63300

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
22	4550	Philippines	PHL	2012	6344.13330
23	4551	Philippines	PHL	2013	7474.77500
24	4552	Philippines	PHL	2014	8632.85400
25	4553	Philippines	PHL	2015	9936.96800
26	4554	Philippines	PHL	2016	11413.57200
27	4555	Philippines	PHL	2017	13036.28300
28	4556	Philippines	PHL	2018	14784.57500
29	4557	Philippines	PHL	2019	16568.20900
30	4558	Philippines	PHL	2020	18693.30300
31	4559	Philippines	PHL	2021	21511.90400
32	4560	Philippines	PHL	2022	24853.17000
33	4561	Philippines	PHL	2023	28589.24400

In [11]:
```
df_SG = df[(df['Entity'] == 'Singapore')].reset_index()
df_SG
```

Out[11]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
0	4902	Singapore	SGP	1990	40.61708
1	4903	Singapore	SGP	1991	49.23315
2	4904	Singapore	SGP	1992	59.82601
3	4905	Singapore	SGP	1993	71.84572
4	4906	Singapore	SGP	1994	86.09643
5	4907	Singapore	SGP	1995	102.34135
6	4908	Singapore	SGP	1996	120.43350
7	4909	Singapore	SGP	1997	141.01138
8	4910	Singapore	SGP	1998	163.41090
9	4911	Singapore	SGP	1999	187.81378
10	4912	Singapore	SGP	2000	213.46340
11	4913	Singapore	SGP	2001	238.46875
12	4914	Singapore	SGP	2002	261.04800
13	4915	Singapore	SGP	2003	283.69165
14	4916	Singapore	SGP	2004	306.88736
15	4917	Singapore	SGP	2005	328.60434
16	4918	Singapore	SGP	2006	354.56693
17	4919	Singapore	SGP	2007	383.47217
18	4920	Singapore	SGP	2008	409.86380
19	4921	Singapore	SGP	2009	433.64035
20	4922	Singapore	SGP	2010	453.70358
21	4923	Singapore	SGP	2011	465.36210

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
22	4924	Singapore	SGP	2012	466.09930
23	4925	Singapore	SGP	2013	464.24260
24	4926	Singapore	SGP	2014	456.79250
25	4927	Singapore	SGP	2015	444.32944
26	4928	Singapore	SGP	2016	421.34567
27	4929	Singapore	SGP	2017	386.71417
28	4930	Singapore	SGP	2018	346.08807
29	4931	Singapore	SGP	2019	303.29062
30	4932	Singapore	SGP	2020	264.85820
31	4933	Singapore	SGP	2021	232.85349
32	4934	Singapore	SGP	2022	208.85173
33	4935	Singapore	SGP	2023	191.41882

In [43]:
```
df_THA = df_THA[(df_THA['Year'] >= 2014) & (df_THA['Year'] <= 2023)]
df_THA
```

Out[43]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
24	5400	Thailand	THA	2014	13912.5890
25	5401	Thailand	THA	2015	12902.3150
26	5402	Thailand	THA	2016	11988.1320
27	5403	Thailand	THA	2017	11313.7705
28	5404	Thailand	THA	2018	10710.1750
29	5405	Thailand	THA	2019	10305.6350
30	5406	Thailand	THA	2020	9950.6610
31	5407	Thailand	THA	2021	9654.7870
32	5408	Thailand	THA	2022	9380.5160
33	5409	Thailand	THA	2023	9081.2430

In [24]:
```
df_ph = df_v2[(df_v2['Year'] >= 2014) & (df['Year'] <= 2023)]
df_ph
```

```
C:\Users\tipqc\AppData\Local\Temp\ipykernel_19136\2867416515.py:1: UserWarning: Boolean Series key will be reindexed
to match DataFrame index.
  df_ph = df_v2[(df_v2['Year'] >= 2014) & (df['Year'] <= 2023)]
```

Out[24]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
24	4552	Philippines	PHL	2014	8632.854
25	4553	Philippines	PHL	2015	9936.968
26	4554	Philippines	PHL	2016	11413.572
27	4555	Philippines	PHL	2017	13036.283
28	4556	Philippines	PHL	2018	14784.575
29	4557	Philippines	PHL	2019	16568.209
30	4558	Philippines	PHL	2020	18693.303
31	4559	Philippines	PHL	2021	21511.904
32	4560	Philippines	PHL	2022	24853.170
33	4561	Philippines	PHL	2023	28589.244

In [25]:
```
df_SG2 = df_SG[(df_v2['Year'] >= 2014) & (df['Year'] <= 2023)]
df_SG2
```

```
C:\Users\tipqc\AppData\Local\Temp\ipykernel_19136\3830868204.py:1: UserWarning: Boolean Series key will be reindexed
to match DataFrame index.
  df_SG2 = df_SG[(df_v2['Year'] >= 2014) & (df['Year'] <= 2023)]
```

Out[25]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
24	4926	Singapore	SGP	2014	456.79250
25	4927	Singapore	SGP	2015	444.32944
26	4928	Singapore	SGP	2016	421.34567
27	4929	Singapore	SGP	2017	386.71417
28	4930	Singapore	SGP	2018	346.08807
29	4931	Singapore	SGP	2019	303.29062
30	4932	Singapore	SGP	2020	264.85820
31	4933	Singapore	SGP	2021	232.85349
32	4934	Singapore	SGP	2022	208.85173
33	4935	Singapore	SGP	2023	191.41882

In [44]:

```
df_new = pd.concat([df_SG2, df_ph, df_THA])
df_new
```

Out[44]:

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
24	4926	Singapore	SGP	2014	456.79250
25	4927	Singapore	SGP	2015	444.32944
26	4928	Singapore	SGP	2016	421.34567
27	4929	Singapore	SGP	2017	386.71417
28	4930	Singapore	SGP	2018	346.08807
29	4931	Singapore	SGP	2019	303.29062
30	4932	Singapore	SGP	2020	264.85820
31	4933	Singapore	SGP	2021	232.85349
32	4934	Singapore	SGP	2022	208.85173
33	4935	Singapore	SGP	2023	191.41882
24	4552	Philippines	PHL	2014	8632.85400
25	4553	Philippines	PHL	2015	9936.96800
26	4554	Philippines	PHL	2016	11413.57200
27	4555	Philippines	PHL	2017	13036.28300
28	4556	Philippines	PHL	2018	14784.57500
29	4557	Philippines	PHL	2019	16568.20900
30	4558	Philippines	PHL	2020	18693.30300
31	4559	Philippines	PHL	2021	21511.90400
32	4560	Philippines	PHL	2022	24853.17000
33	4561	Philippines	PHL	2023	28589.24400
24	5400	Thailand	THA	2014	13912.58900
25	5401	Thailand	THA	2015	12902.31500

	index	Entity	Code	Year	New HIV Infections - Age: total - Sex: total - Central estimate
26	5402	Thailand	THA	2016	11988.13200
27	5403	Thailand	THA	2017	11313.77050
28	5404	Thailand	THA	2018	10710.17500
29	5405	Thailand	THA	2019	10305.63500
30	5406	Thailand	THA	2020	9950.66100
31	5407	Thailand	THA	2021	9654.78700
32	5408	Thailand	THA	2022	9380.51600
33	5409	Thailand	THA	2023	9081.24300

In [45]:
```
df_new = df_new.rename(columns={
    'New HIV Infections - Age: total - Sex: total - Central estimate' : 'New HIV Infections Estimate',
    'Entity' : 'Country'})
```

In [46]:
```
df_new.reset_index()
```

Out[46]:

	level_0	index	Country	Code	Year	New HIV Infections Estimate
0	24	4926	Singapore	SGP	2014	456.79250
1	25	4927	Singapore	SGP	2015	444.32944
2	26	4928	Singapore	SGP	2016	421.34567
3	27	4929	Singapore	SGP	2017	386.71417
4	28	4930	Singapore	SGP	2018	346.08807
5	29	4931	Singapore	SGP	2019	303.29062
6	30	4932	Singapore	SGP	2020	264.85820
7	31	4933	Singapore	SGP	2021	232.85349
8	32	4934	Singapore	SGP	2022	208.85173
9	33	4935	Singapore	SGP	2023	191.41882
10	24	4552	Philippines	PHL	2014	8632.85400
11	25	4553	Philippines	PHL	2015	9936.96800
12	26	4554	Philippines	PHL	2016	11413.57200
13	27	4555	Philippines	PHL	2017	13036.28300
14	28	4556	Philippines	PHL	2018	14784.57500
15	29	4557	Philippines	PHL	2019	16568.20900
16	30	4558	Philippines	PHL	2020	18693.30300
17	31	4559	Philippines	PHL	2021	21511.90400
18	32	4560	Philippines	PHL	2022	24853.17000
19	33	4561	Philippines	PHL	2023	28589.24400
20	24	5400	Thailand	THA	2014	13912.58900
21	25	5401	Thailand	THA	2015	12902.31500

	level_0	index	Country	Code	Year	New HIV Infections Estimate
22	26	5402	Thailand	THA	2016	11988.13200
23	27	5403	Thailand	THA	2017	11313.77050
24	28	5404	Thailand	THA	2018	10710.17500
25	29	5405	Thailand	THA	2019	10305.63500
26	30	5406	Thailand	THA	2020	9950.66100
27	31	5407	Thailand	THA	2021	9654.78700
28	32	5408	Thailand	THA	2022	9380.51600
29	33	5409	Thailand	THA	2023	9081.24300

In [47]:
```
df_new = df_new.drop('index', axis=1)
df_new
```

Out[47]:

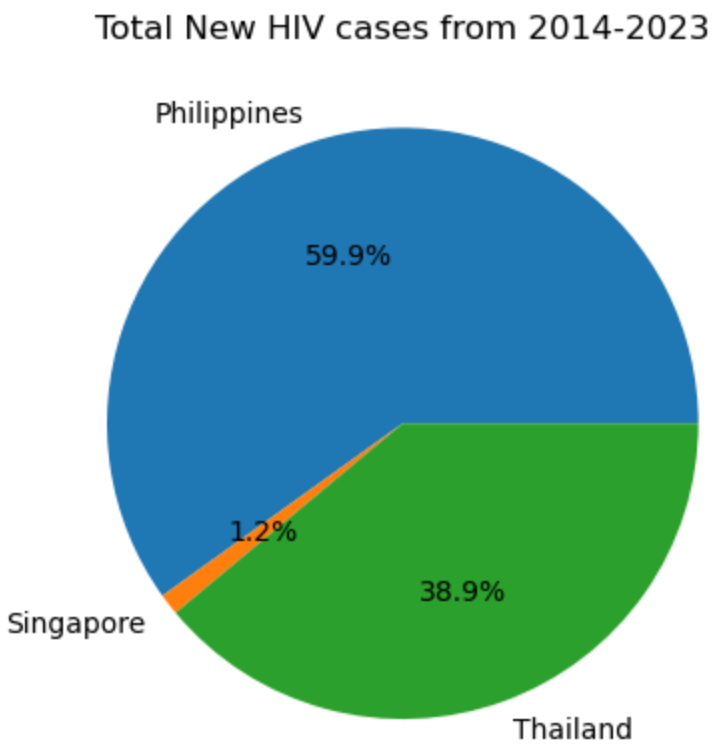
	Country	Code	Year	New HIV Infections Estimate
24	Singapore	SGP	2014	456.79250
25	Singapore	SGP	2015	444.32944
26	Singapore	SGP	2016	421.34567
27	Singapore	SGP	2017	386.71417
28	Singapore	SGP	2018	346.08807
29	Singapore	SGP	2019	303.29062
30	Singapore	SGP	2020	264.85820
31	Singapore	SGP	2021	232.85349
32	Singapore	SGP	2022	208.85173
33	Singapore	SGP	2023	191.41882
24	Philippines	PHL	2014	8632.85400
25	Philippines	PHL	2015	9936.96800
26	Philippines	PHL	2016	11413.57200
27	Philippines	PHL	2017	13036.28300
28	Philippines	PHL	2018	14784.57500
29	Philippines	PHL	2019	16568.20900
30	Philippines	PHL	2020	18693.30300
31	Philippines	PHL	2021	21511.90400
32	Philippines	PHL	2022	24853.17000
33	Philippines	PHL	2023	28589.24400
24	Thailand	THA	2014	13912.58900
25	Thailand	THA	2015	12902.31500

	Country	Code	Year	New HIV Infections Estimate
26	Thailand	THA	2016	11988.13200
27	Thailand	THA	2017	11313.77050
28	Thailand	THA	2018	10710.17500
29	Thailand	THA	2019	10305.63500
30	Thailand	THA	2020	9950.66100
31	Thailand	THA	2021	9654.78700
32	Thailand	THA	2022	9380.51600
33	Thailand	THA	2023	9081.24300

In [65]:
```
totalHIV = df_new.groupby(['Country'])['New HIV Infections Estimate'].sum()
country_names = ['Philippines','Singapore','Thailand']
```

In [84]:
```
plt.pie(totalHIV,labels = country_names, autopct ='%1.1f%%')
plt.title('Total New HIV cases from 2014-2023')
```

Out[84]:
```
Text(0.5, 1.0, 'Total New HIV cases from 2014-2023')
```



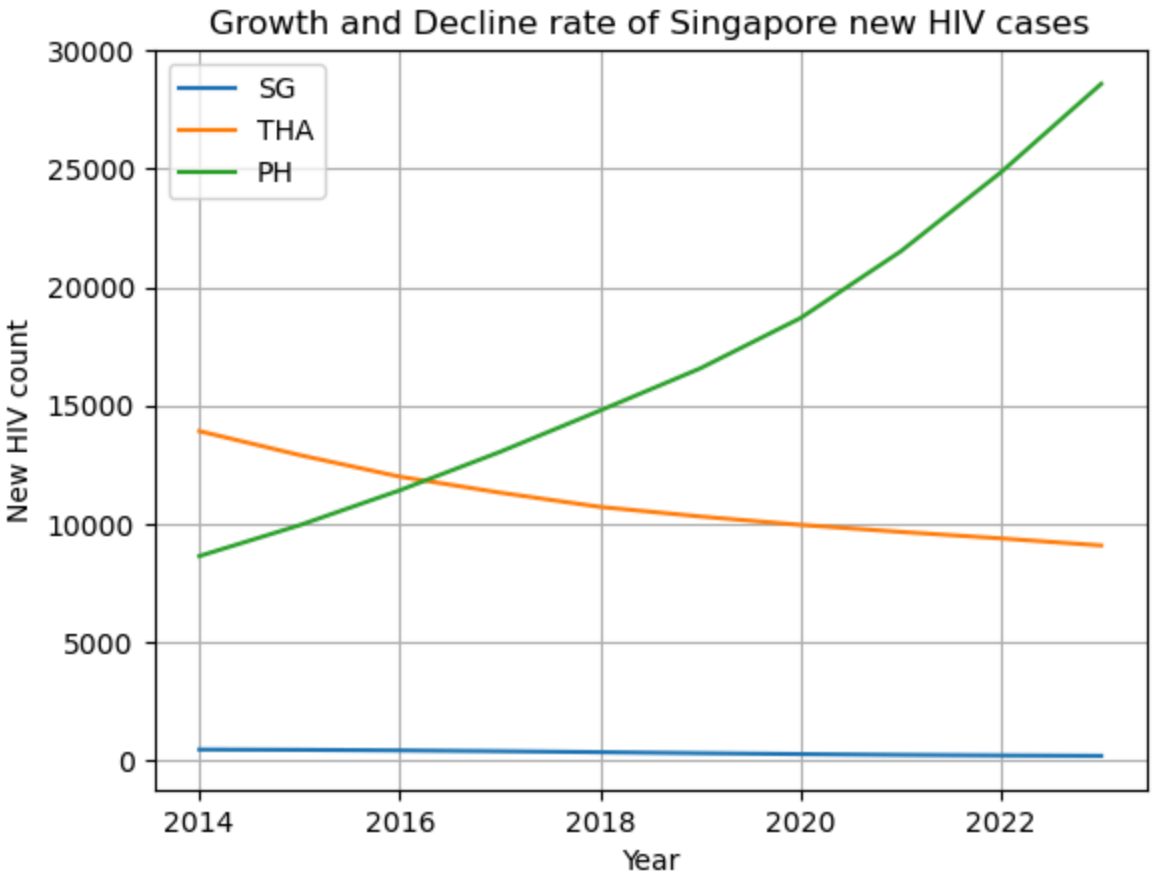
The pie chart shows us that the Philippines has the largest count of New HIV cases(59.9%) summed from the year 2014-2023, Thailand follows suit having 38.9%, and lastly Singapore having only 1.2% which means Singapore has a way to control the cases of HIV rather than PH and Thailand

In [83]:
```
plt.plot(df_SG2['Year'],df_SG2['New HIV Infections - Age: total - Sex: total - Central estimate'],label = 'SG')
plt.plot(df_THA['Year'],df_THA['New HIV Infections - Age: total - Sex: total - Central estimate'],label = 'THA')
plt.plot(df_ph['Year'],df_ph['New HIV Infections - Age: total - Sex: total - Central estimate'],label = 'PH')


plt.title('Growth and Decline rate of Singapore new HIV cases')
plt.xlabel('Year')
plt.ylabel('New HIV count')
plt.legend()
plt.grid()
```

```
plt.show
```

Out[83]: <function matplotlib.pyplot.show(close=None, block=None)>



This plot shows us that the Singapore already had a low count of new HIV cases during 2014, and even had it decline. While Thailand started with a high count of people getting HIV but had it on a decline. On the other hand the Philippines started with a lower count than Thailand but managed to grow the new HIV ammount cases per year.

What we did:

Cleaned the data Sorted the data Filtered the data Combined data frames Renamed columns

Made data visualization to better interpret data(Pie chart, Plot)

In conclusion, managing data like this and cleaning and using, its quite easy now when we have learned a lot.