

## 필요한 라이브러리 불러오기

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline
```

## 인구수예제.xlsx 파일 불러오기

```
In [2]: data = pd.read_excel('data/인구수예제.xlsx')
data.head()
```

Out[2]:

	도시	자치구	연도	남자인구	여자인구	총인구
0	서울	강남구	2013	73	92	165
1	서울	강남구	2014	139	55	194
2	서울	강남구	2015	123	83	206
3	서울	강남구	2016	147	150	297
4	서울	강남구	2017	57	133	190

jupyter notebook에서는 한글이 깨져서 나와서 설정해줘야합니다

다음 수업시간에 설명할 해드릴 예정이니 여기서는 영어로 바꾸고 풀면 될 것 같습니다

1. 각 column을 아래와 같이 바꿔주세요

2. '서울' -> 'Seoul', '부산' -> 'Busan'으로 바꿔주세요

```
In [2]: # 아래 셀에서 작성해주세요
```

Out[2]:

	city	province	year	male	female	total
0	Seoul	강남구	2013	73	92	165
1	Seoul	강남구	2014	139	55	194
2	Seoul	강남구	2015	123	83	206
3	Seoul	강남구	2016	147	150	297
4	Seoul	강남구	2017	57	133	190
5	Seoul	서대문구	2013	95	111	206
6	Seoul	서대문구	2014	149	150	299
7	Seoul	서대문구	2015	106	77	183
8	Seoul	서대문구	2016	56	109	165
9	Seoul	서대문구	2017	82	96	178
10	Seoul	종로구	2013	121	68	189

```
In [24]: data.columns = ['city', 'province', 'year', 'male', 'female', 'total']
data = data.replace(['서울', '부산'], ['Seoul', 'Busan'])
data.head()
```

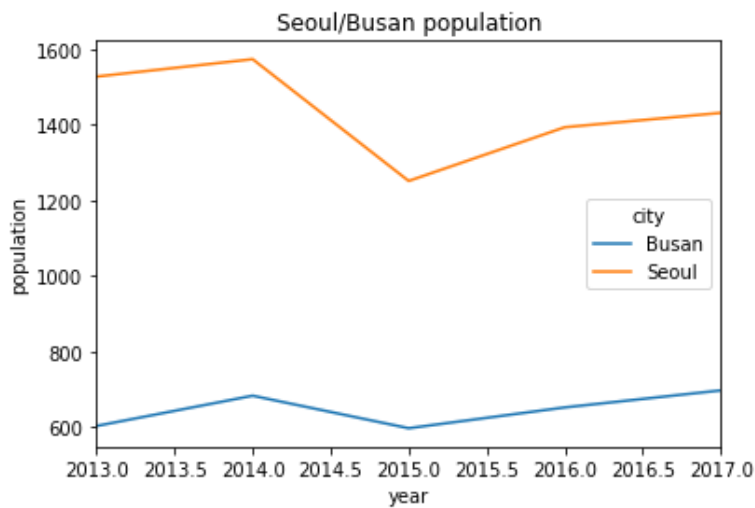
Out[24]:

	city	province	year	male	female	total
0	Seoul	강남구	2013	73	92	165
1	Seoul	강남구	2014	139	55	194
2	Seoul	강남구	2015	123	83	206
3	Seoul	강남구	2016	147	150	297
4	Seoul	강남구	2017	57	133	190

## 2. Seoul과 Busan의 2013~2017년도의 총인구 합 선그래프를 그려주세요

```
In [3]: # 아래 셀에서 작성해주세요
```

```
Out[3]: Text(0.5, 1.0, 'Seoul/Busan population')
```



```
In [25]: df = pd.DataFrame(data.groupby(['city', 'year'])['total'].sum())  
df
```

Out[25]:

total		
city	year	
Busan	2013	603
	2014	683
	2015	597
	2016	652
	2017	697
Seoul	2013	1527
	2014	1573
	2015	1251
	2016	1393
	2017	1431

```
In [26]: df = pd.DataFrame(data.groupby(['city', 'year'])['total'].sum().unstack('city'))  
df
```

Out[26]:

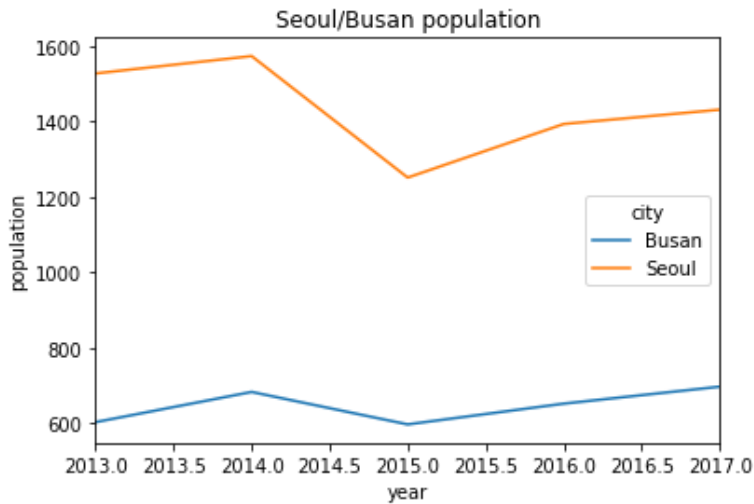
city	Busan	Seoul
year		
2013	603	1527
2014	683	1573
2015	597	1251
2016	652	1393
2017	697	1431

```
In [30]: axes = df.plot()

axes.set_xlabel('year')
axes.set_ylabel('population')
axes.set_title('Seoul/Busan population')
axes.legend(['Busan', 'Seoul'], loc = "center right", title = "city")

axes
```

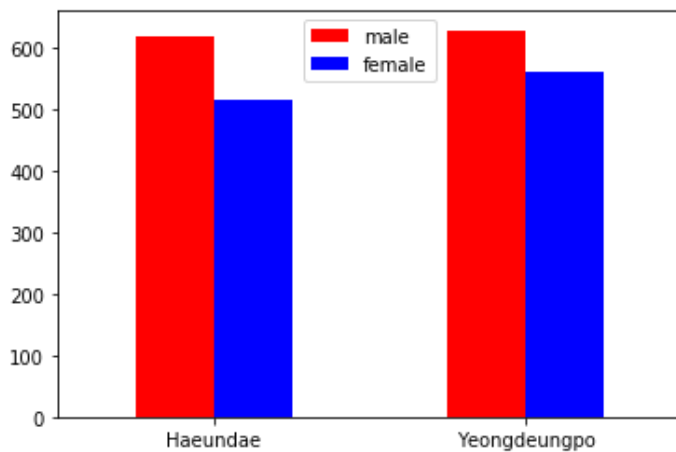
Out[30]: <matplotlib.axes.\_subplots.AxesSubplot at 0x122d655d0>



### 3. Seoul과 Busan에서 각각 총인구가 가장 많은 구의 남자/여자 인구 막대 그래프를 그려주세요

```
In [4]: # 아래 셀에서 작성해주세요
```

Out[4]: <matplotlib.legend.Legend at 0x15896826688>



In [31]: data

Out[31]:

	city	province	year	male	female	total
0	Seoul	강남구	2013	73	92	165
1	Seoul	강남구	2014	139	55	194
2	Seoul	강남구	2015	123	83	206
3	Seoul	강남구	2016	147	150	297
4	Seoul	강남구	2017	57	133	190
5	Seoul	서대문구	2013	95	111	206
6	Seoul	서대문구	2014	149	150	299
7	Seoul	서대문구	2015	106	77	183
8	Seoul	서대문구	2016	56	109	165
9	Seoul	서대문구	2017	82	96	178
10	Seoul	종로구	2013	121	68	189

In [34]: pd.DataFrame(data.groupby(['city', 'province'])['total'].sum().sort\_values(ascending=True))

Out[34]:

		total
city	province	
Seoul	영등포구	1191
Busan	해운대구	1135
	수영구	1061
Seoul	강남구	1052
	동작구	1036
Busan	동래구	1036
Seoul	도봉구	1035
	서대문구	1031
	송파구	974
	종로구	856

In [50]: data = data.replace(['영등포구', '해운대구'], ['Yeongdeungpo', 'Haeundae'])

```
In [51]: df2 = data[(data.province == "Haeundae") | (data.province == "Yeongdeungpo")]
df2
```

Out[51]:

	city	province	year	male	female	total
15	Seoul	Haeundae	2013	146	113	259
16	Seoul	Haeundae	2014	127	117	244
17	Seoul	Haeundae	2015	70	72	142
18	Seoul	Haeundae	2016	141	136	277
19	Seoul	Haeundae	2017	145	124	269
35	Busan	Yeongdeungpo	2013	124	103	227
36	Busan	Yeongdeungpo	2014	101	144	245
37	Busan	Yeongdeungpo	2015	115	70	185
38	Busan	Yeongdeungpo	2016	134	126	260
39	Busan	Yeongdeungpo	2017	146	72	218

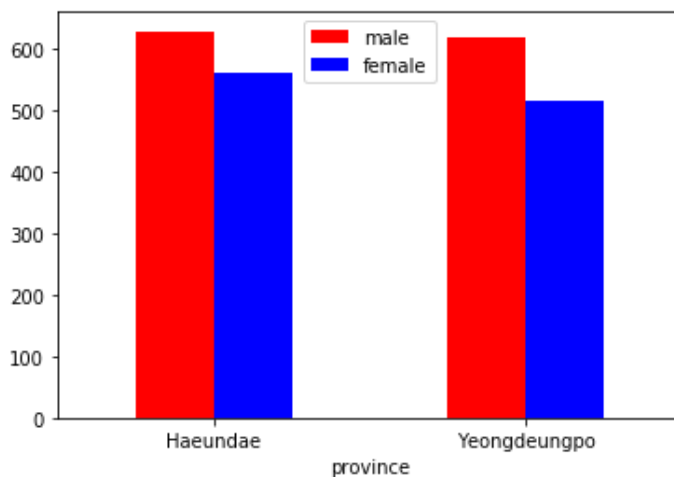
```
In [52]: data2 = df2.groupby(["province"])[ 'male', 'female' ].sum()
data2
```

Out[52]:

	male	female
province		
Haeundae	629	562
Yeongdeungpo	620	515

```
In [57]: axes = data2.plot.bar(rot = 0, color = ["red", "blue"])
axes.legend(loc = "upper center")
```

Out[57]: <matplotlib.legend.Legend at 0x123253e10>



**NC Dinos.xlsx파일 중 2013년도 파일만 이용해서**

**아래와 같이 scatter를 그리는 draw 함수를 만들어주세요**

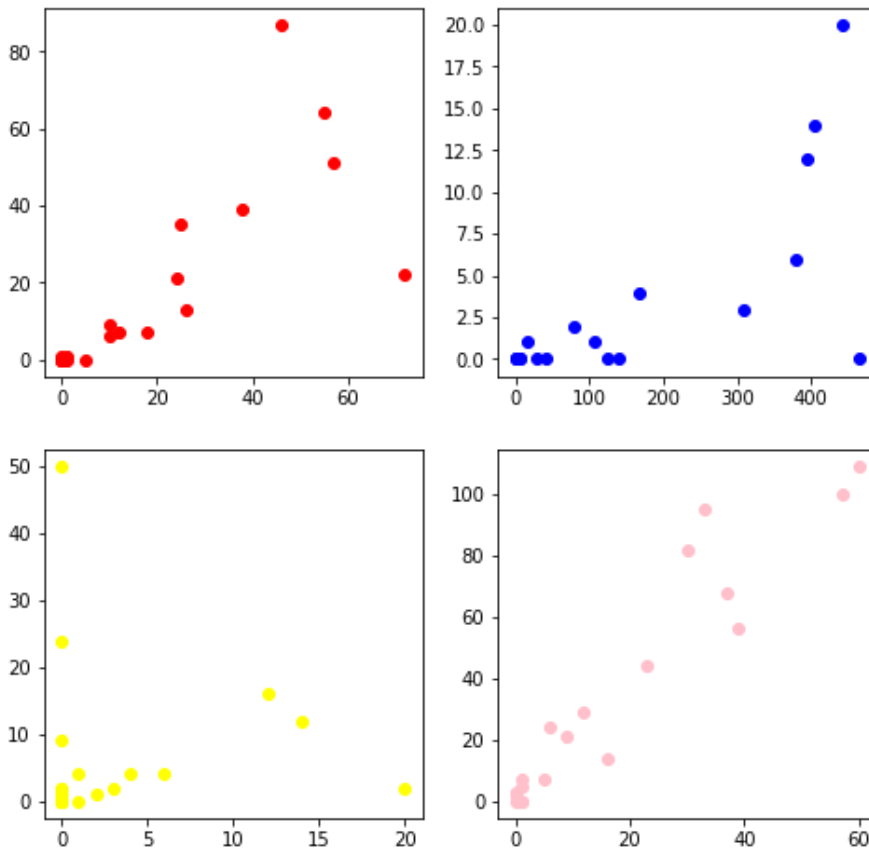
```

In [5]: # 아래 셀에서 작성해주세요
'''
    1 : 첫번째 도화지 선택
    2 : 색깔 입력
    3 : x축에 들어갈 column
    4 : y축에 들어갈 column
'''

# figsize 지정
fig = plt.figure(figsize=(8, 8))

draw(1, 'red', '득점', '타점')
draw(2, 'blue', '타수', '홈런')
draw(3, 'yellow', '홈런', '도루')
draw(4, 'pink', '볼넷', '삼진')

```



```
In [59]: baseball = pd.read_excel("data/NC Dinos.xlsx", sheet_name = "2013")
baseball
```

Out[59]:

	Unnamed: 0	선수명	팀명	경기	타석	타수	안타	홈런	득점	타점	볼넷	삼진	도루	BABIP	타율	출루율	장타율	OPS
0	0	모창민	NC	108	436	395	109	12	57	51	37	68	16	0.307	0.276	0.339	0.443	0.782
1	1	이호준	NC	126	508	442	123	20	46	87	60	109	2	0.324	0.278	0.362	0.475	0.837
2	2	김종호	NC	128	546	465	129	0	72	22	57	100	50	0.352	0.277	0.376	0.333	0.709
3	3	나성범	NC	104	458	404	98	14	55	64	33	95	12	0.279	0.243	0.319	0.416	0.735
4	4	조영	NC	120	426	380	107	6	38	39	39	56	4	0.316	0.282	0.350	0.413	0.763

```
In [62]: def draw(num, col, x, y):
ax = fig.add_subplot(2, 2, num)

ax.scatter(baseball[x], baseball[y], color = col)
```

```
In [64]: fig = plt.figure(figsize=(8, 8))
```

```
draw(1, 'red', '득점', '타점')
draw(2, 'blue', '타수', '홈런')
draw(3, 'yellow', '홈런', '도루')
draw(4, 'pink', '볼넷', '삼진')
```

