

필요한 라이브러리 불러오기

```
In [28]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

인구수예제.xlsx 파일 불러오기

```
In [164]: data = pd.read_excel("data/인구수예제.xlsx")
```

jupyter notebook에서는 한글이 깨져서 나와서 설정해줘야합니다

다음 수업시간에 설명할 헤드릴 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

```
In [165]: data = data.replace(["서울", "부산"], ["Seoul", "Busan"])
```

```
In [166]: data.columns = ["city", "province", "year", "male", "female", "total"]
```

```
In [167]: data.head()
```

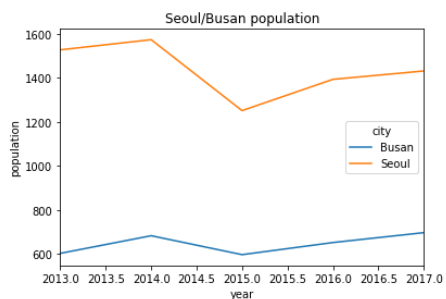
```
Out [167]:
```

	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
```

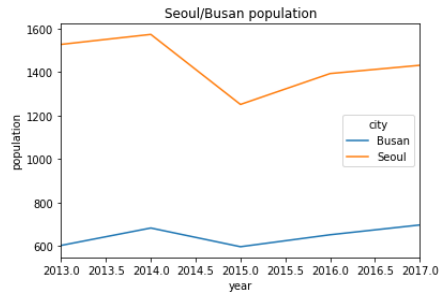
```
Out [3]: Text(0.5, 1.0, 'Seoul/Busan population')
```



```
In [147]: sr = pd.DataFrame(data.groupby(["city", "year"])["total"].sum().unstack("city"))
```

```
In [150]: axes=sr.plot()
axes.set_xlabel("year")
axes.set_ylabel("population")
axes.set_title("Seoul/Busan population")
axes.legend(["Busan", "Seoul"], loc="center right", title="city")
```

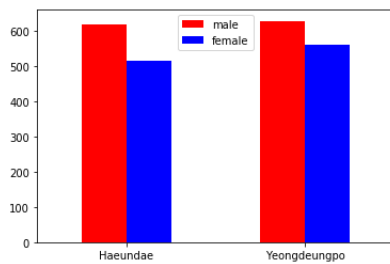
```
Out [150]: <matplotlib.legend.Legend at 0x2c1d82c65f8>
```



3. Seoul과 Busan에서 각각 총인구가 가장 많은 구의 남자/여자 인구 막대그래프를 그려주세요

```
In [4]: # 아래 셀에서 작성해주세요
```

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



```
In [175]: data = data.replace(["강남구", "서대문구", "종로구", "영등포구", "송파구", "도봉구", "동작구", "해운대구", "수영구", "동래구"], ["Gangnam", "Seodaemun", "Jongno", "Yeongdeungpo", "Songpa", "Dobong", "Dongjak", "Haeundae", "Suyeong", "Dongrae"])
```

```
In [176]: pd.DataFrame(data.groupby(["city", "province"])["total"].sum().sort_values(ascending=False))
```

```
Out [176]:
```

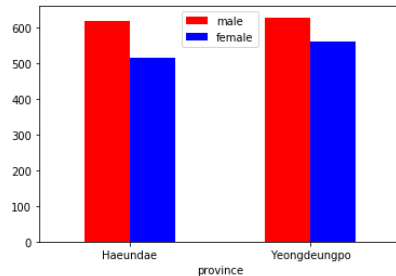
		total
city	province	
Seoul	Yeongdeungpo	1191
Busan	Haeundae	1135
	Suyeong	1061
Seoul	Gangnam	1052
	Dongjak	1036
Busan	Dongnae	1036
Seoul	Dobong	1035

```
In [193]: data2 = data[(data["province"]=="Yeongdeungpo") | (data.province=="Haeundae")]
```

```
In [199]: sr = data2.groupby(["province"])[["male", "female"]].sum()
```

```
In [205]: axes=sr.plot(kind="bar", rot=0, color=["red", "blue"])
axes.legend(loc="upper center")
```

```
Out [205]: <matplotlib.legend.Legend at 0x2c1d85a2eb8>
```



NC Dinos.xlsx파일 중 2013년도 파일만 이용해서

아래와 같이 scatter를 그리는 draw 함수를 만들어주세요

```
In [220]: df = pd.read_excel("data/NC Dinos.xlsx", sheet_name="2013")
```

```
In [231]: def draw(num, color, x, y):
ax = fig.add_subplot(2, 2, num)
ax.scatter(df[x], df[y], color = color)
```

```
In [232]: fig = plt.figure(figsize=(8, 8))
```

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

