

## 5.2 피벗테이블과 크로스탭 활용

피벗테이블은 `groupby()`를 사용하여 레벨이 있는 인덱스 구조 기반의 데이터 재구성 연산

### 5.2.1 Pivot-table 기법

기본적인 연산 형태는 index로 원하는 컬럼 지정

```
penguins.pivot_table(index=['species', 'sex'])
```

연산 방식 미지정시 index에서 지정하는 컬럼을 제외한 전체 컬럼의 평균값 구함

Pivot 연산 기본: `groupby()`

연산될 값: values에 입력, index로 저장 가능

index / columns 선택하여 연산 가능

함수 포함: `margin = True`로 지정

Pivot의 기본 집계 함수: `mean()` ↔ 다른 연산 방식의 새로운 dimension 지정하여 사용 가능  
즉, 여러 dimension을 통해 연산 하여 다양한 그룹을 기반으로 통계를 구할 수 있음

### 5.2.2 crosstab 기법

`crosstab()` 함수: 특정 그룹의 빈도를 계산하기 위한 피벗 테이블의 특수한 경우

↑위 연산이 가능한 경우 → 2개 이상의 범주형 변수 간의 빈도를 계산하는 경우

↑작동원리: 2개 이상의 변수에 대해 다른 변수를 계산하여 테이블로 반환 하는 것

`crosstab` 함수의 margin 활용 가능

```
pd.crosstab(index=[penguins.species, penguins.sex],
            columns=penguins.island,
            margins=True)
```

2개 이상의 변수가 포함된 index에 index로 쓰기  
결측값은 `dropna = False` 이용하여 사용.

### [7] 데이터 시각화

7.1 맷플롯립 데이터 시각화

7.1.1 기법: 개인 그래프 출력과 라인 플롯 사용 가능

`plot()` 함수: 개인 그래프 그리기

\* 그래프 함수에 데이터 입력하기

`plt.plot()` ← y값 입력하여

`plt.show`

→ x값, y값 값 입력을 2개 생성

• 데이터가 4000 데이터를 많았을 때 그래프 출력

• 다른 컬럼을 입력용 data frame 입력하여 그래프 출력

\* x, y를 레이블 설정

`xlabel()`, `ylabel()`

\* 컬러 지정하기

`plot()` 함수의 dimension color

\* 라벨 표시

\* x, y를 범위 지정하기

함수: `legend()`

`xlim()`, `ylim()`