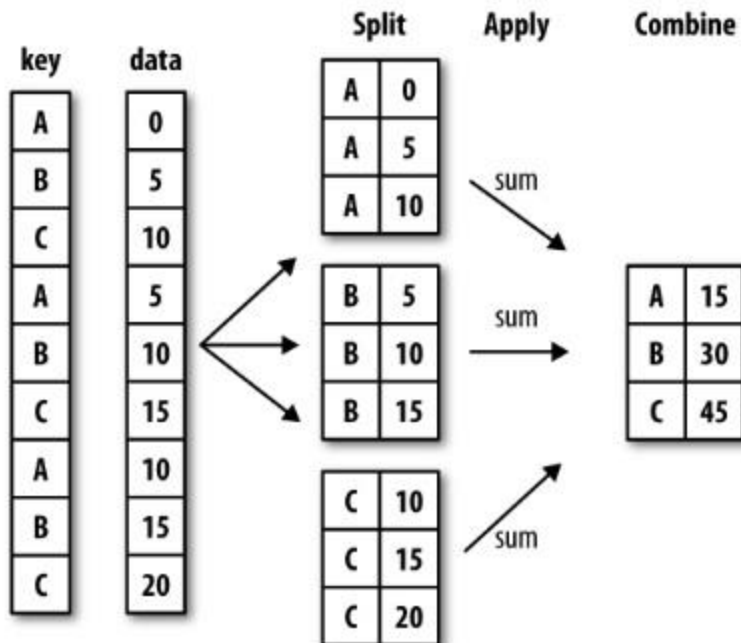


데이터 다루기 II

GroupBy

- Pandas GroupBy

- R 프로그래밍 패키지 제작자인 해들리 위캠은 분리 적용 결합이라는 그룹 연산을 통해서 처리함
- 그룹 색인의 형태는
 - 그룹으로 묶을 축과 같은 길이의 리스트나 배열
 - DataFrame의 컬럼 이름을 지칭하는 값
 - 그룹으로 묶을 값과 그룹 이름에 대응하는 사전이나 Series 개체
 - 축 색인 혹은 색인 내의 개별 이름에 대해 실행되는 함수



GroupBy

- 그룹간의 순회
- 컬럼 또는 컬럼의 일부만 선택하기
- 사전과 Series에서 묶기
- 함수로 묶기
- 색인 단계로 묶기

GroupBY

- 컬럼에 여러가지 함수 적용하기
- apply
 - aggregate와 마찬가지로 tranform을 위한 함수
- 그룹 색인 생략하기
- 그룹에 국한된 값으로 누락된 값 채우기
- 랜덤 표본과 순열
- 그룹 가중 평균과 상관관계