근로자 휴가조사에 대한

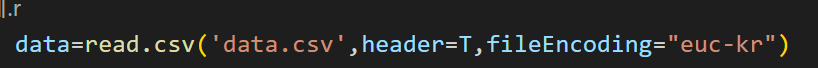
간단한 분석

학번 : 20195718

학과 : 산업보안학과

이름 : 이동길

1. csv 파일 읽어오기



현재 디렉터리에 .r파일과 data.csv 파일이 같이 있으므로 이렇게 작성해준다.

2. 데이터 파악

Str(data)를 통해 구조를 파악한다

화살이(가) 표시된 사진

자동 생성된 설명

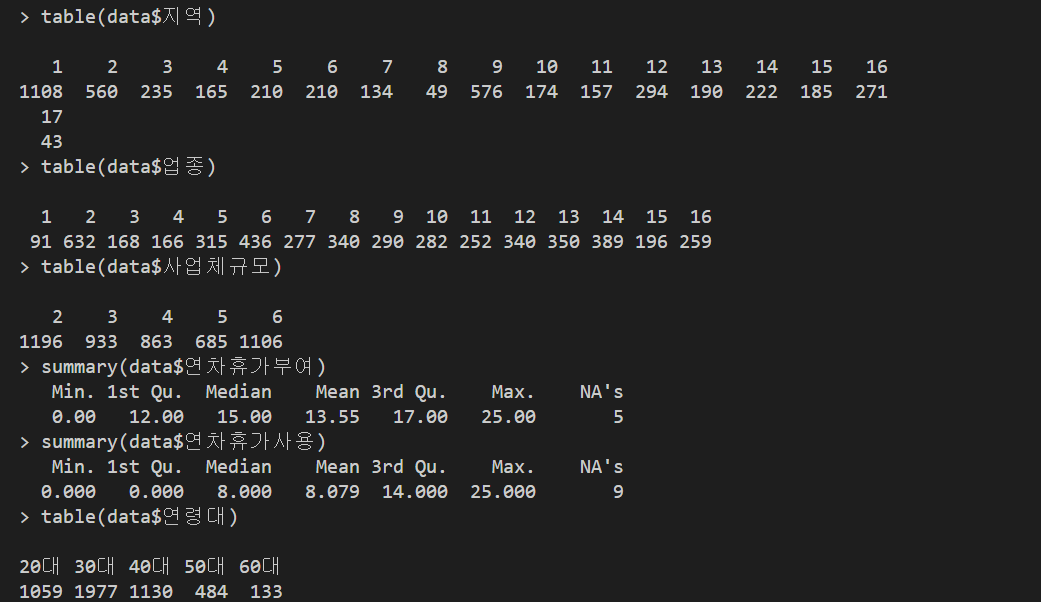
이를 통해서 4773개의 요소가 있고, 10개의 컬럼(지역, 업종, 사업체규모, 연차휴가부여, 연차휴가사용, 연령대, 성별, 학력, 직업, 월평균수득구간)이 있는 것을 확인 가능하다.

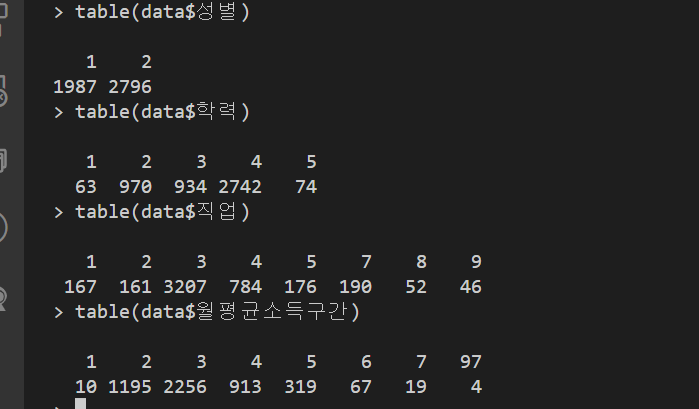
또한 지역, 업종, 사업체규모, 연차휴가부여, 성별, 학력, 직업, 월평균 소득구간의 데이터타입은 int이고, 연령대는 chr형태임을 확인할 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

위의 코드를 통해 각 컬럼 별 데이터가 범주형인지, 양인지 파악한다

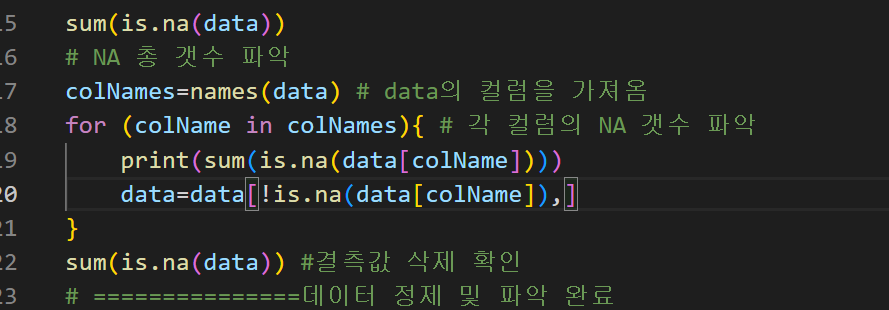




이를 통해 각 컬럼별 데이터의 종류, 평균값들을 알 수 있다.

지역, 업종, 사업체 규모, 연령대, 성별, 학력, 직업, 월평균소득구간은 특정 종류의 데이터로 한정되어있고, 연차휴가부여와 연차휴가사용은 연속된 데이터이다.

3. 데이터 정제 (결측값 제거)



Sum(is.na(data))를 통해서 총 결측값이 몇 개인지 파악한다.

텍스트이(가) 표시된 사진

자동 생성된 설명

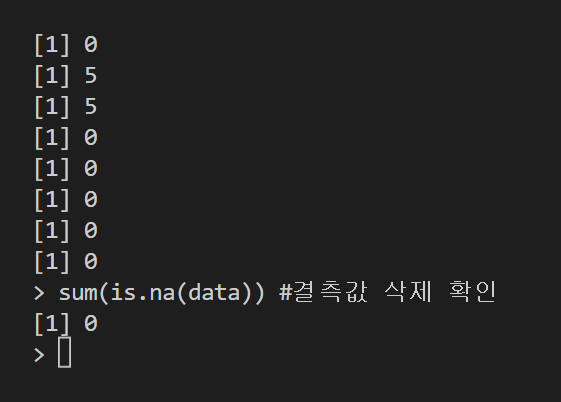
총 14개인 것을 확인했으니 어느 컬럼에 있는지도 파악해준다.

텍스트이(가) 표시된 사진

자동 생성된 설명

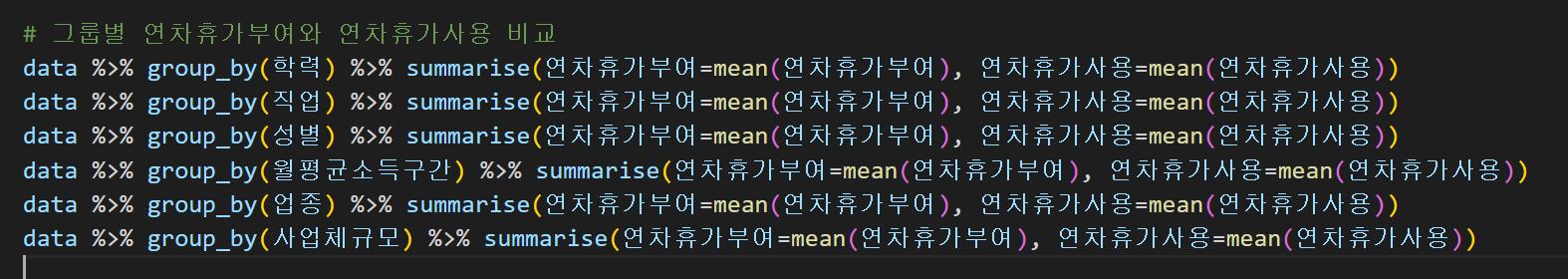
For 문을 통해 어느 컬럼에 NA가 있는지 확인이 가능하고, 동시에 제거해주었다.

Is.na(data)를 통해 제대로 제거되었는지 확인한다.



4. 데이터 분석

1) 정제된 데이터로부터 범주변수별로 연차휴가부여 평균과 연차휴가사용평균을 비교



데이터 구조 파악을 통해 얻은 범주형 변수들(학력, 직업, 성별, 월평균소득구간, 업종, 사업체규모) 에 대해서 group\_by를 실행하여 비교해준다. 아래는 6개 항목이 실행된 모습이다.

텍스트이(가) 표시된 사진

자동 생성된 설명

학력에 따른 연차휴가부여 대비 연차 휴가사용은 5) 대학원 이상의 학력을 가진 사람들은 67.3%의 휴가를 사용하여 연차부여대비 사용량이 가장 커보인다. 부여 일수도 가장 많고, 사용량도 가장 많다.

텍스트, 전자기기, 검은색이(가) 표시된 사진

자동 생성된 설명

직업에 따른 연차부여대비 사용은 9) 단순 노무종사자의 경우 가장 적게 사용하는 것으로 확인된다. 부여일수도 가장 적고, 실제 사용도 가장 어려움을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

성별에 따른 휴가부여대비 사용은 남성이 휴가 부여도 많고, 연차 휴가 사용량도 많으나 해당 비율은 여성이 더 높은 것으로 확인된다.

텍스트이(가) 표시된 사진

자동 생성된 설명

월평균소득별 휴가부여 대비 사용은 1) 100만원 이하의 소득을 가진 사람들이 부여휴가대비 가장 많은 휴가를 사용했다. 그러나 부여량은 가장 적고, 사용량은 두번째로 적다.

텍스트이(가) 표시된 사진

자동 생성된 설명

사업체 규모별로는 5) 50-99인의 부여대비 휴가 사용량이 가장 많았다. 비율도 가장 높다.

텍스트이(가) 표시된 사진

자동 생성된 설명

업종별로는 12) 전문, 과학 및 기술 서비스업의 휴가 비율이 가장 높았고, 1) 광업의 비율이 가장 적었다. 부여휴가 일수 자체와 사용량자체는 3)전기, 가스, 증기 및 수도사업이 가장 높았다.

2) 20~30대, 50~60대의 학력별 분포

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

20~30대의 학력별 분포, 중졸 이하는 매우 적고, 대졸 이상이 대부분임을 알 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

50~60대의 학력별 분포, 대졸과 고졸의 비율이 거의 비슷한 것을 확인할 수 있다. 대학원 이상의 학력인 사람보다 중졸이하인 학력을 가진 사람이 더 많다.

두 가지 데이터를 비교해보았을 때, 고등학교까지의 교육이 더 활성화되고, 대학의 진학률이 높아 진 것을 추론해볼 수 있다. 동시에, 해당 조사가 근로자임을 고려했을 때, 근로를 위한 조건이 더욱 높아졌다고도 추론해 볼 수 있을 것 같다.

3) 사업체 규모별 성별 월평균소득구간 비교

텍스트이(가) 표시된 사진

자동 생성된 설명

사업체 규모의 범주가 1~3(월평균 100만원~300만원미만)인 데이터의 성별별 월평균소득구간 비교

남성이 여성에 비해 평균적으로 소득구간이 더 높은 것을 확인할 수 있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

사업제 구모의 범조가 4~6(월평균 300만원~600만원 미만)인 데이터의 성별별 월평균소득구간 비교

남성과 여성의 평균 소득구간이 거의 비슷함을 알 수 있다.

해당 데이터의 비교를 통해서 사업체 규모가 커지면 커질수록 임금 격차가 해소된다는 점을 추론해볼 수 있다.