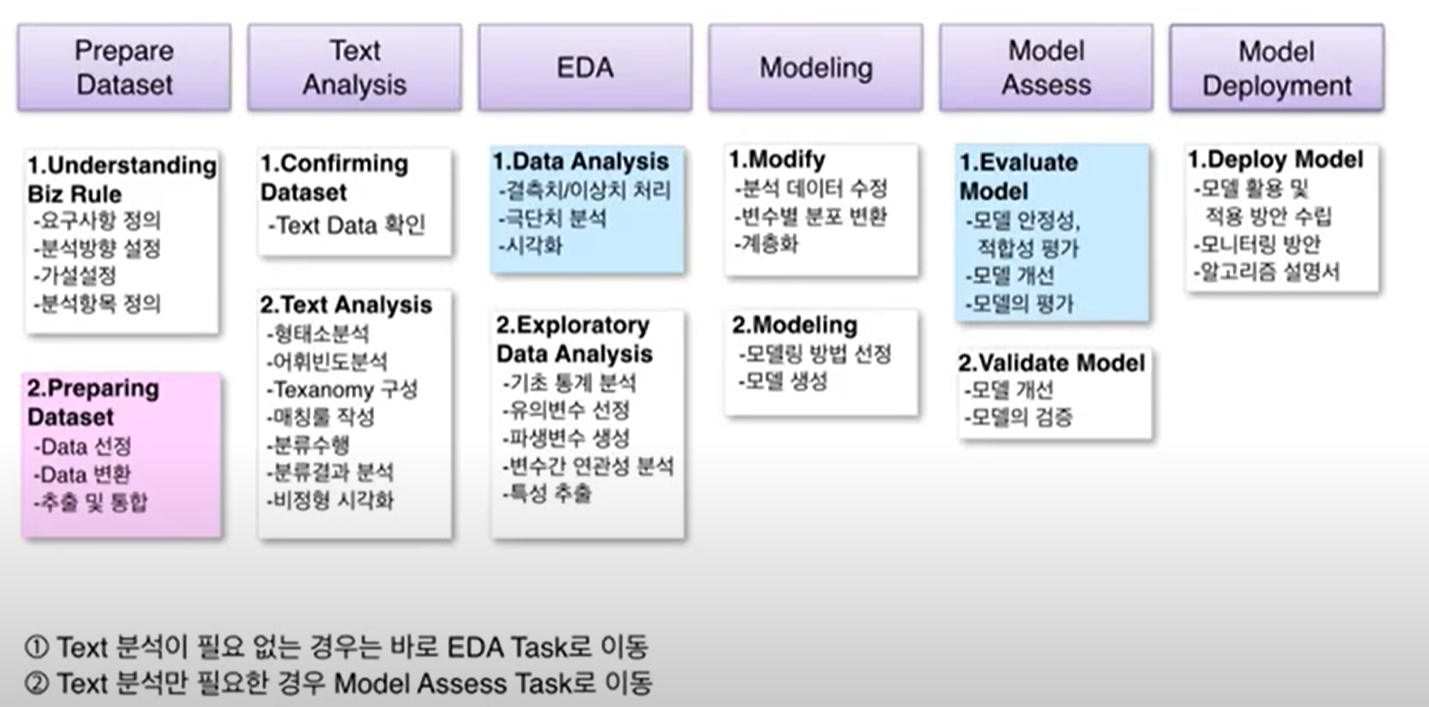
[ 자료의 형태 ]

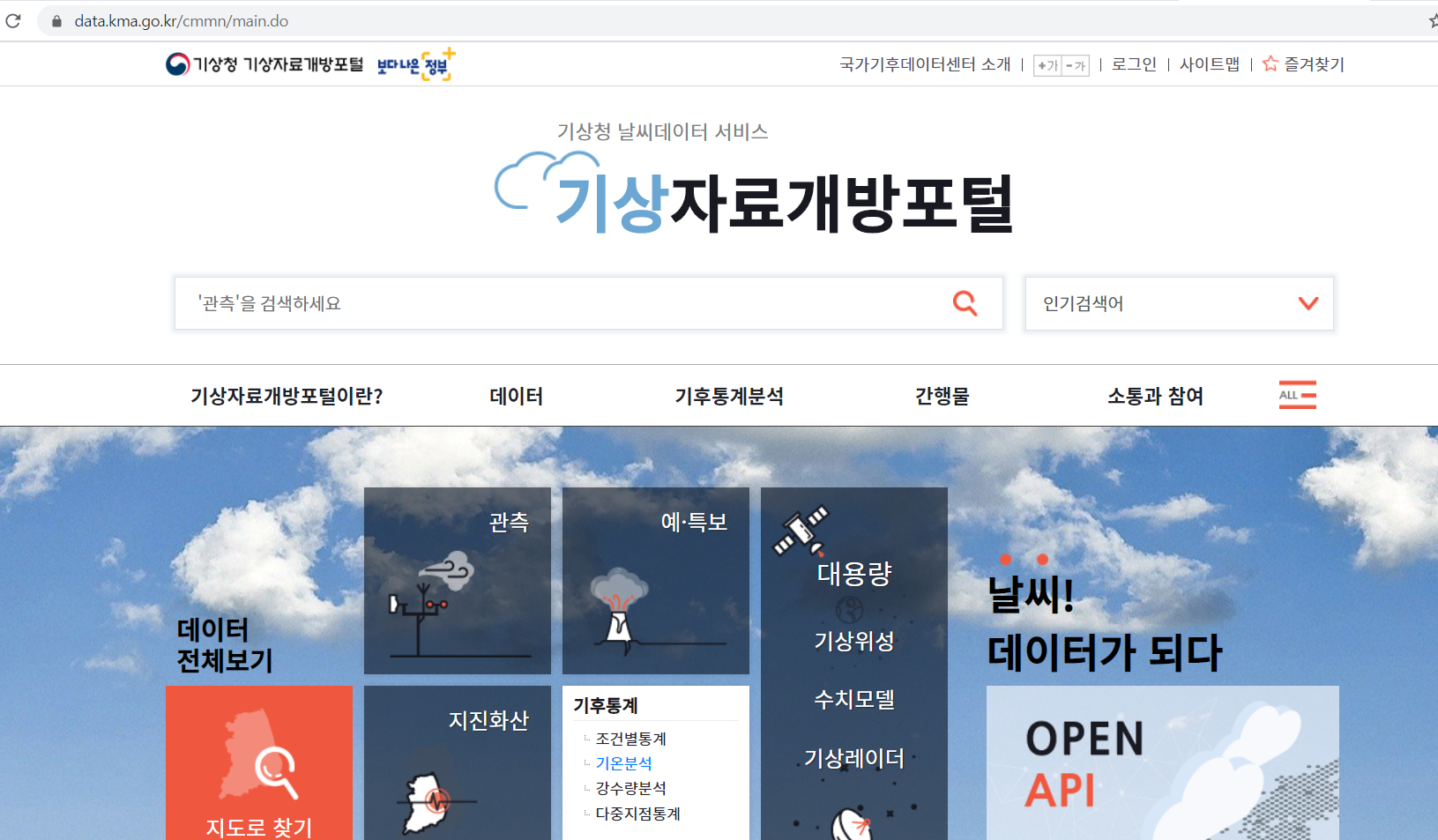
방대한 자료를 그래프나 표 또는 몇 개의 수치로 요약하여 주어진 자료의 전반적인 내용을 쉽고 빠르게 파악할 수 있도록 하는 기법을 다루는 분야를 기술(descriptive) 통계학이라 한다. 기술 통계학의 목적은 자료가 가지고 있는 특성과 모습을 잘 정리하고 요약하는 것이다.

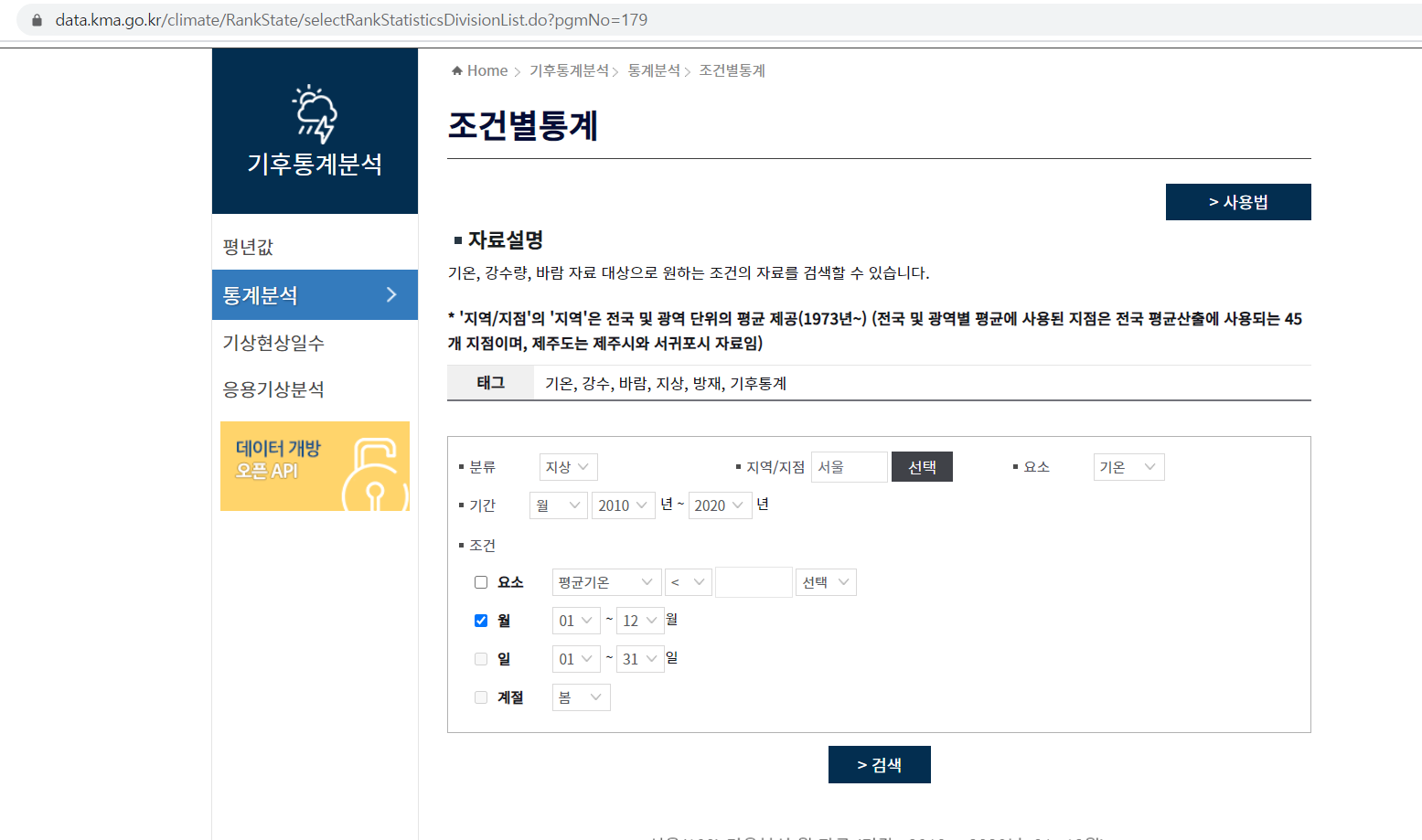
자료의 형태는 크게 질적자료(정성적 자료)와 양적자료(정량적 자료)로 구분된다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 종류 | 특징 | 질적 자료의 종류 | 특징 | 자료 요약 방법 |
| 질적자료 | 성별, 혈액형, 순위 등과 같이 어떤 속성을 나타낸 자료  사칙연산 적용 불가 | 명목형 자료 | 성, 지역, 직업, 학과와 같이 자료의 크기나 순서에 대한 의미가 없으며 단지 자료값 자체의 이름만 의미를 부여한다. | 범주별 빈도  막대그래프  원그래프  분할표 |
| 순서형 자료 | 어떤 기준에 따라 자료값들의 순서에 의미를 부여할 수 있는 자료로서 자료가 숫자로 구성되더라도 순서를 나타낼 뿐이지 비율에는 의미가 없다. 학년, 순위 선호도 등이다. |
| 양적자료 | 길이, 무게, 자동차 생산대수 등과 같은 많고 적음을 나타내는 수치로 된 자료를 말하며 양적자료에 대해서는 사칙연산이 가능하다. | 계수(count)형 자료 | 자료가 셀 수 있는 정수 형태로서 형제 수, 자동차 보유대수, 입사 지원자수, 보험 해약건수 등이 속한다. | 기술통계량(최대, 최소, 평균, 중앙값, 사분위 값 등)  히스토그램  박스플롯  산점도 |
| 연속형 자료 | 자료의 측정이 셀 수 없는 소수점을 포함하는 경우이다. 사람의 키, 몸무게, 물체의 무게, 길이, 온도, 시간, 연령, 성적 등이다.  비율형과 구간형으로 나뉘며 비율형은 수치값들 사이의 차이뿐만 아니라 비율에도 의미를 부여할 수 있는 경우이며 절대영점을 가진다. 사칙연산이 모두 가능하다. 구간형은 수치값들 사이의 차이(구간)에는 의미를 부여할 수 있지만 비율에는 의미를 부여할 수 없는 경우이다. 값의 순서와 간격에만 의미가 있으며 덧셈과 뺄셈 연산만 의미가 있다. 성적, 온도 등이 속한다. |

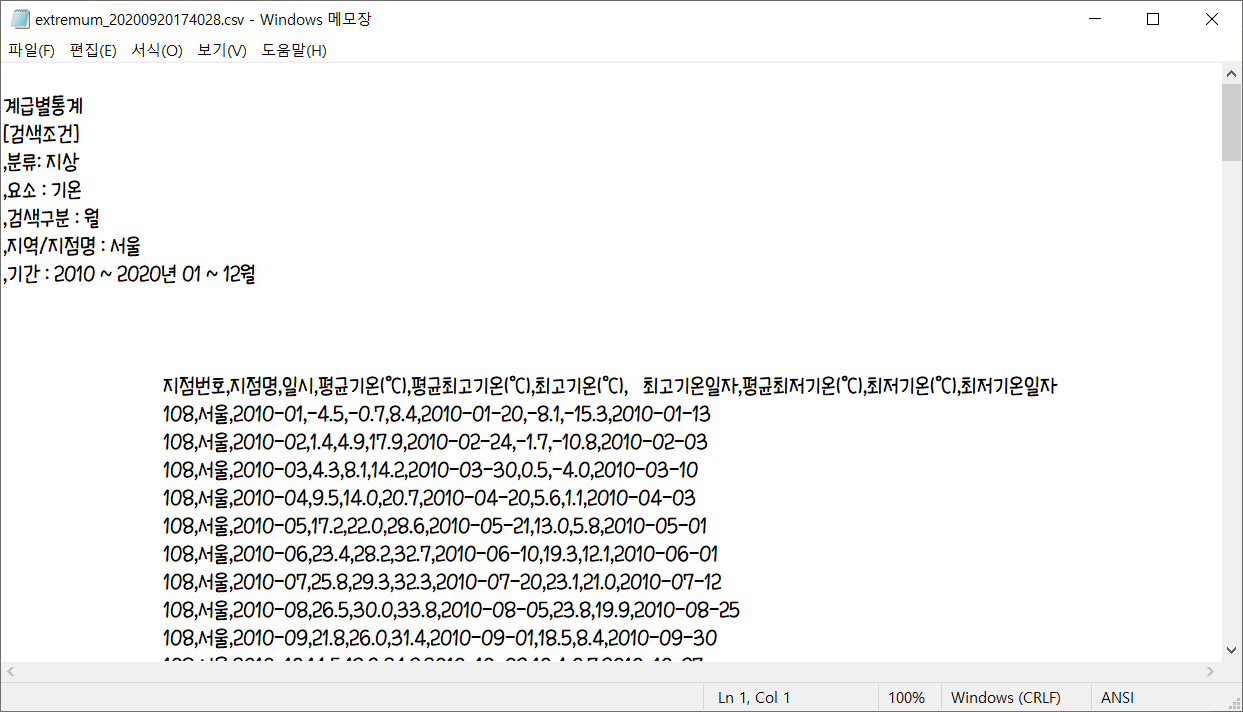


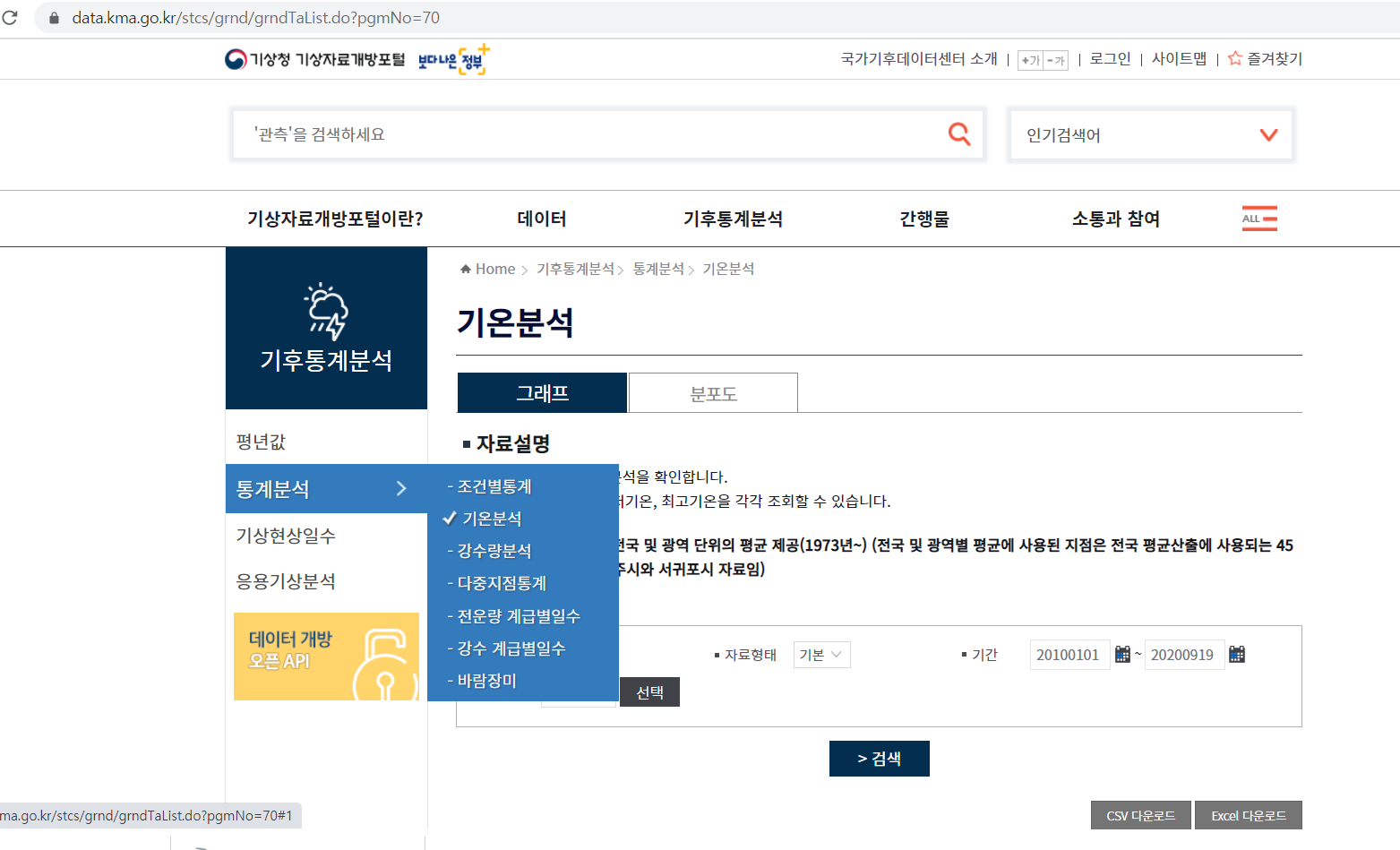
[ 기상자료로 데이터 분석하기 ]

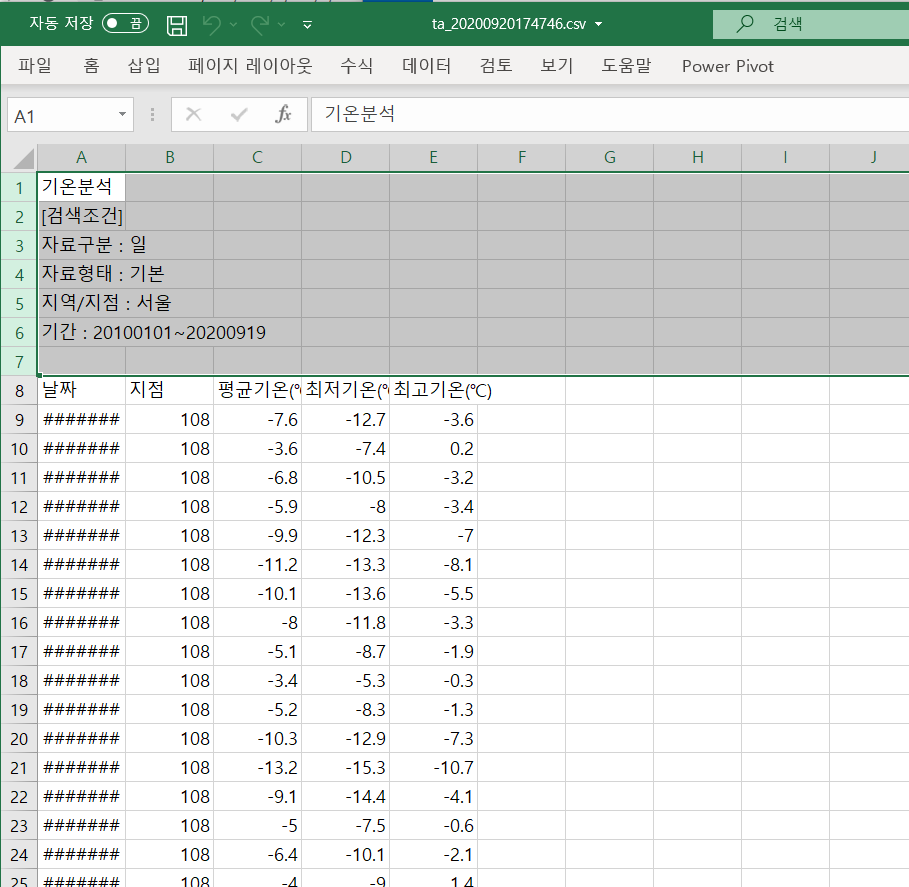




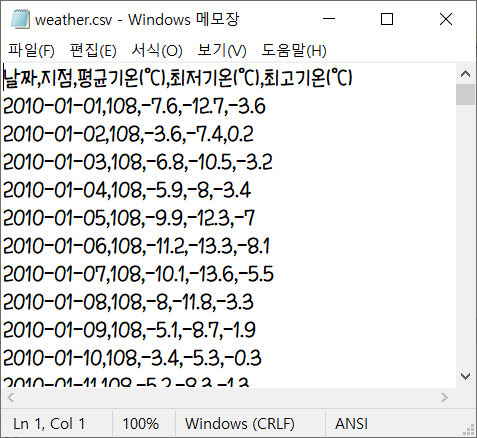


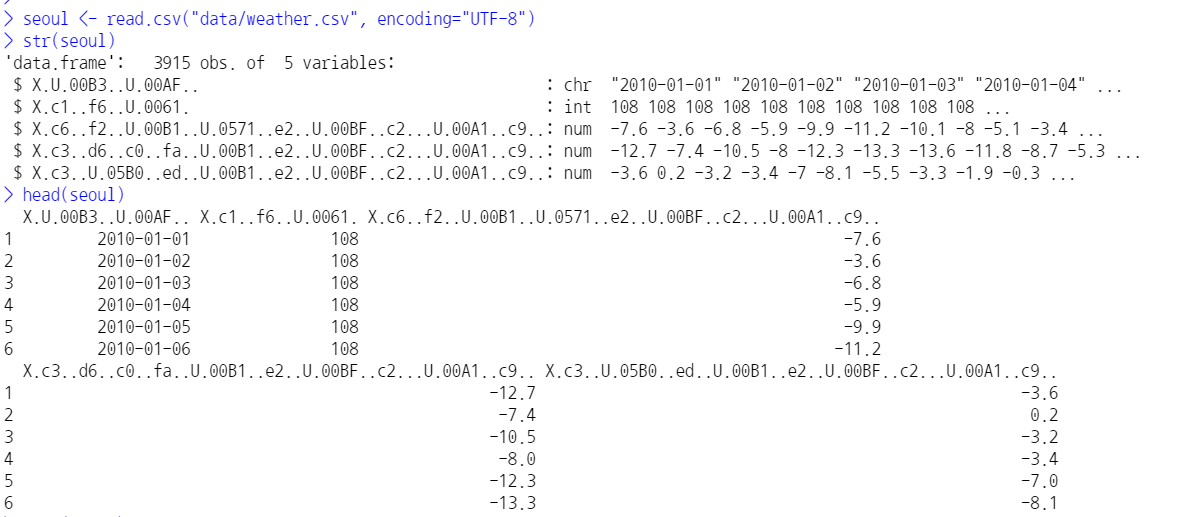


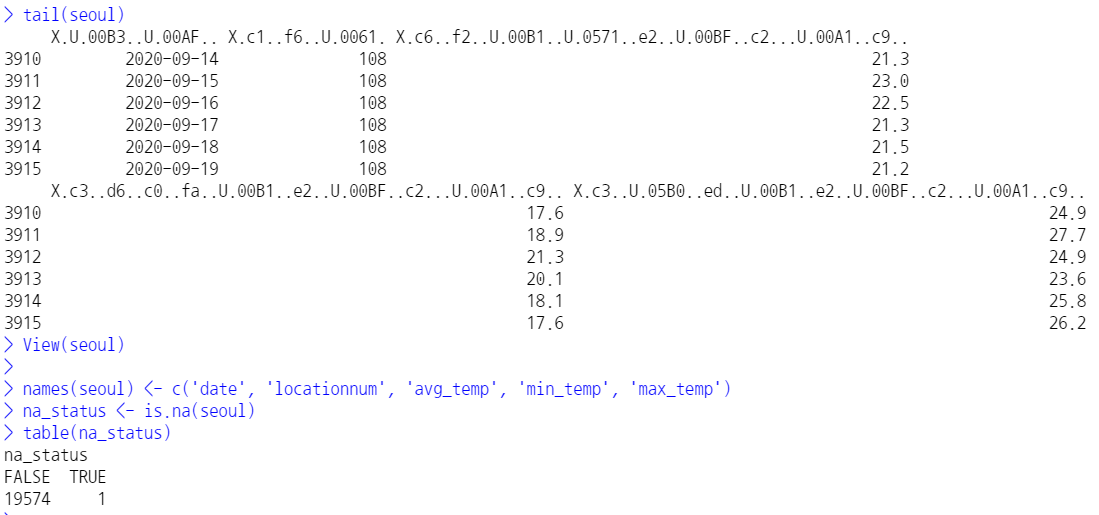
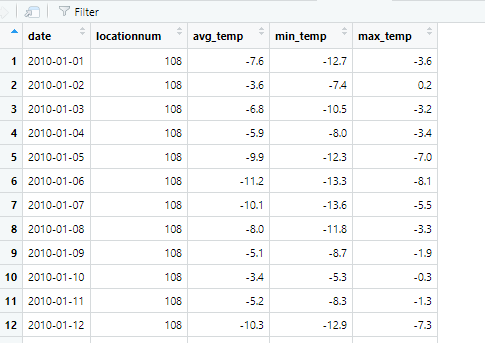


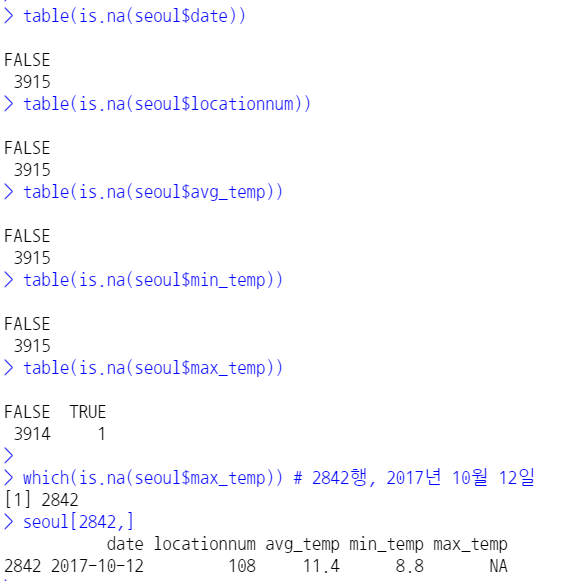


삭제해서 weather.csv 로 저장

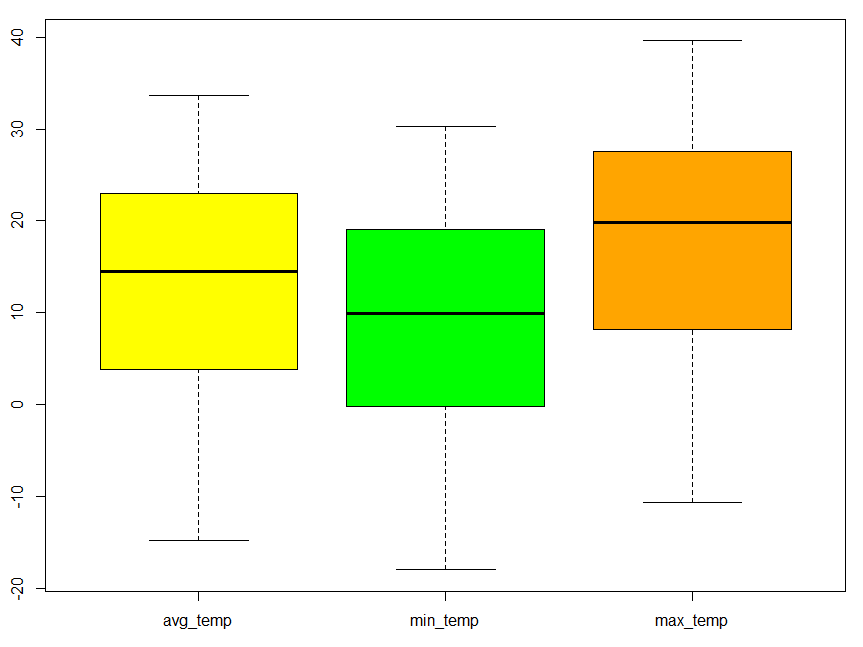








- 이상치 채크 🡪 boxplot(seoul[c(3,4,5)], col=c('yellow', 'green', 'orange'))



[ 미션 1 ]

서울이 가장 더웠던 날은 언제였을까? 얼마나 더웠을까?

일교차가 가장 큰 시기는 1년 중 언제쯤일까?

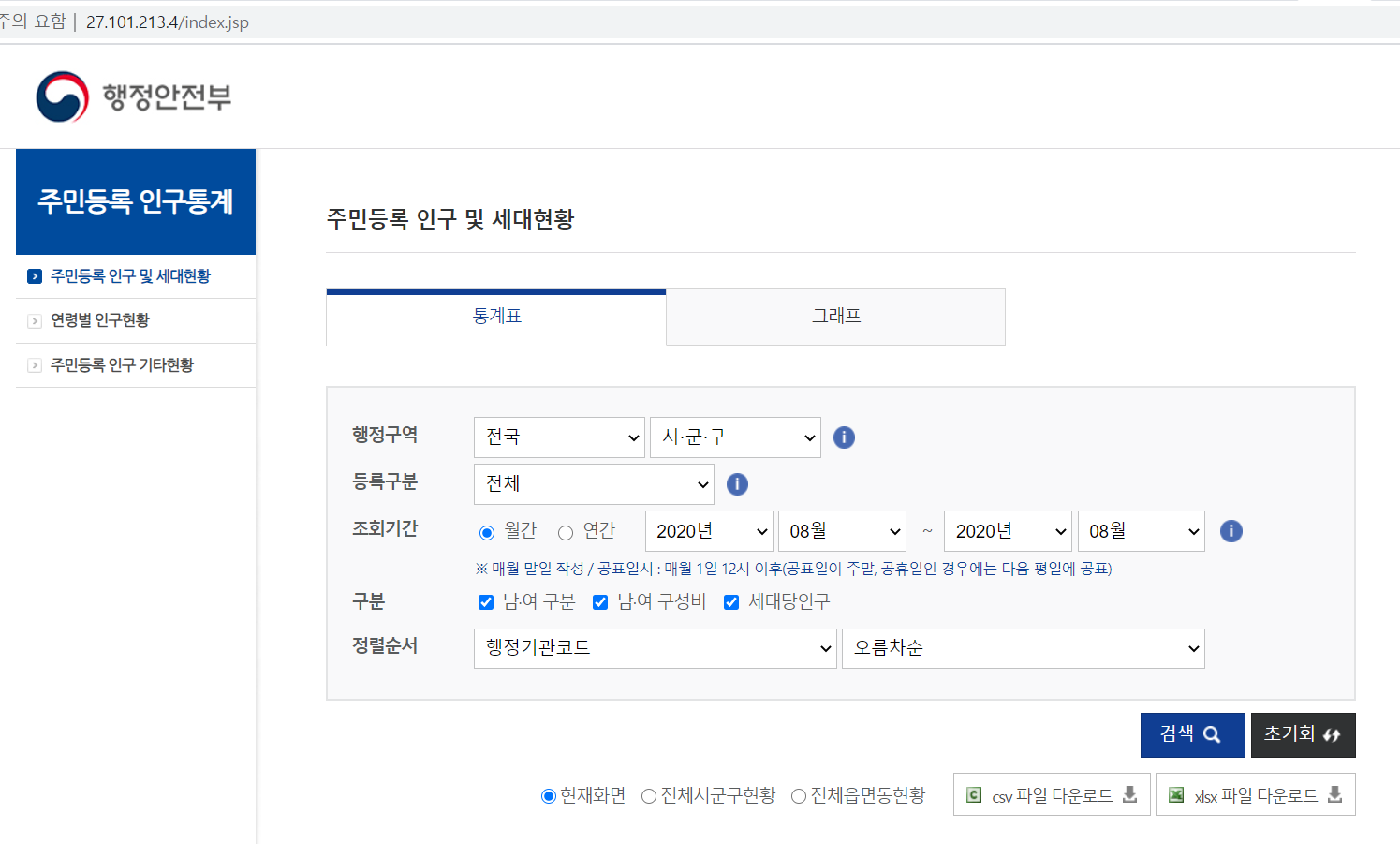
겨울에는 언제 가장 추울까? 12월? 1월? 2월?

년도별 5월 5일 기온변화를 그래프로 그리기

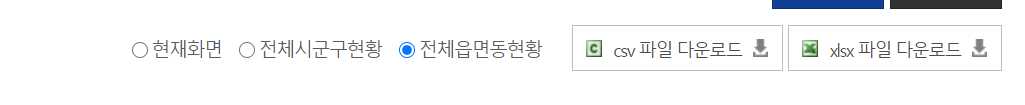
2010년부터 2019년도까지 월별 1월 1일의 기온 그래프를 그리기

[ 주민등록 인구 통계로 데이터 분석하기 ]









[ 미션 2 ]

- 영유아가 가장 많은 동네는?

- 우리 지역에 가장 많은 연령대는?