

ALL\_DATA(범주형\_시간추가\_결측치제거) (1).CSV

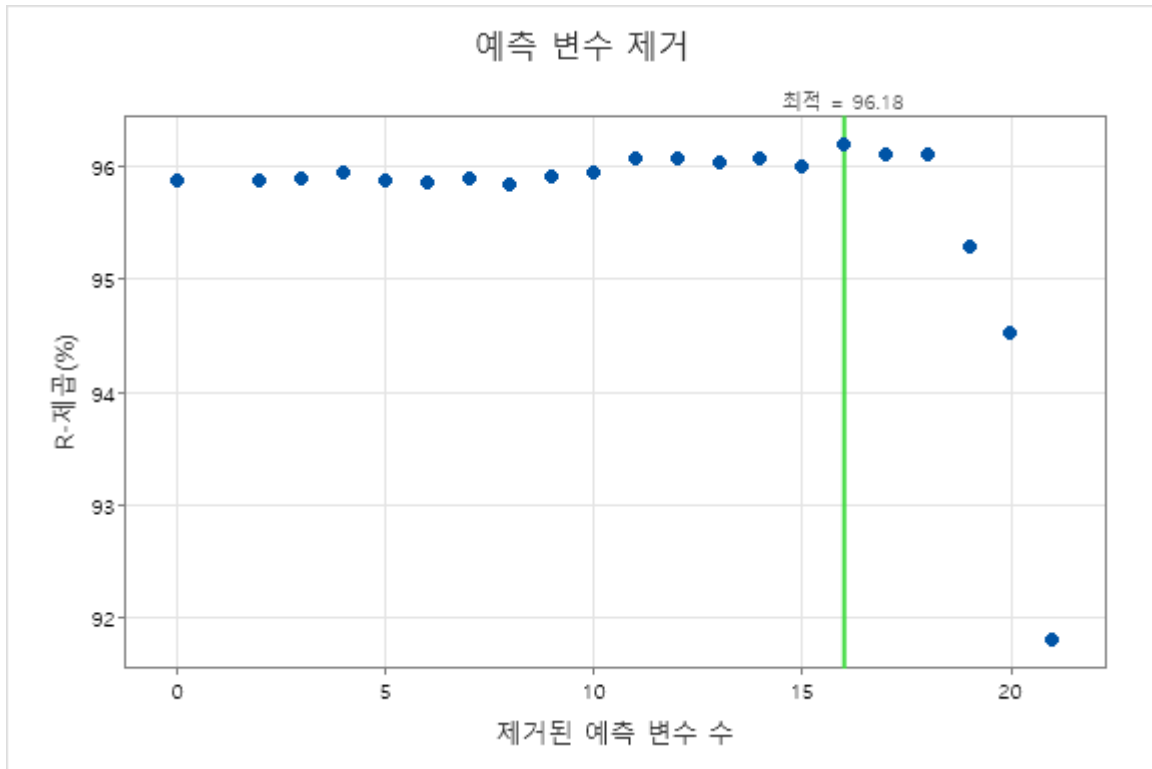
**TreeNet® 회귀 - 주요 예측 변수 검색: 전력거래량(MWh) vs**  
**기온(°C), 풍속(m/s), 풍향(16방위), 습도(%), 증기압(hPa),**  
**이슬점온도(°C), 현지기압(hPa), 해면기압(hPa), 전운량(10분위),**  
**중하층운량(10분위), 최저운고(100m), 시정(10m), 지면온도(°C),**  
**5cm 지중온도(°C), 10cm 지중온도(°C), 20cm 지중온도(°C), 30cm**  
**지중온도(°C), 지역, 년도, 월, 일, 요일, 시간**

## 방법

손실 함수	제곱 오차
최적 트리 수 선택 기준	최대 R-제곱
모형 검증	70/30% 교육/검정 집합
학습률	0.1
하위 표본 부분	0.5
트리당 최대 터미널 노드 수	6
최소 단말 노드 크기	3
노드 분할을 위해 선택된 예측 변수 수	총 예측 변수 수 = 23
사용된 행	446804

## 반응 정보

데이터 집합	N	N의 비율(%)	평균	표준 편차	최소값	Q1	중위수	Q3	최대값
교육	312809	70.01	4680.91	3930.97	1	833	4350	7149	18880
검정	133995	29.99	4680.02	3929.43	1	836	4353	7136	18560



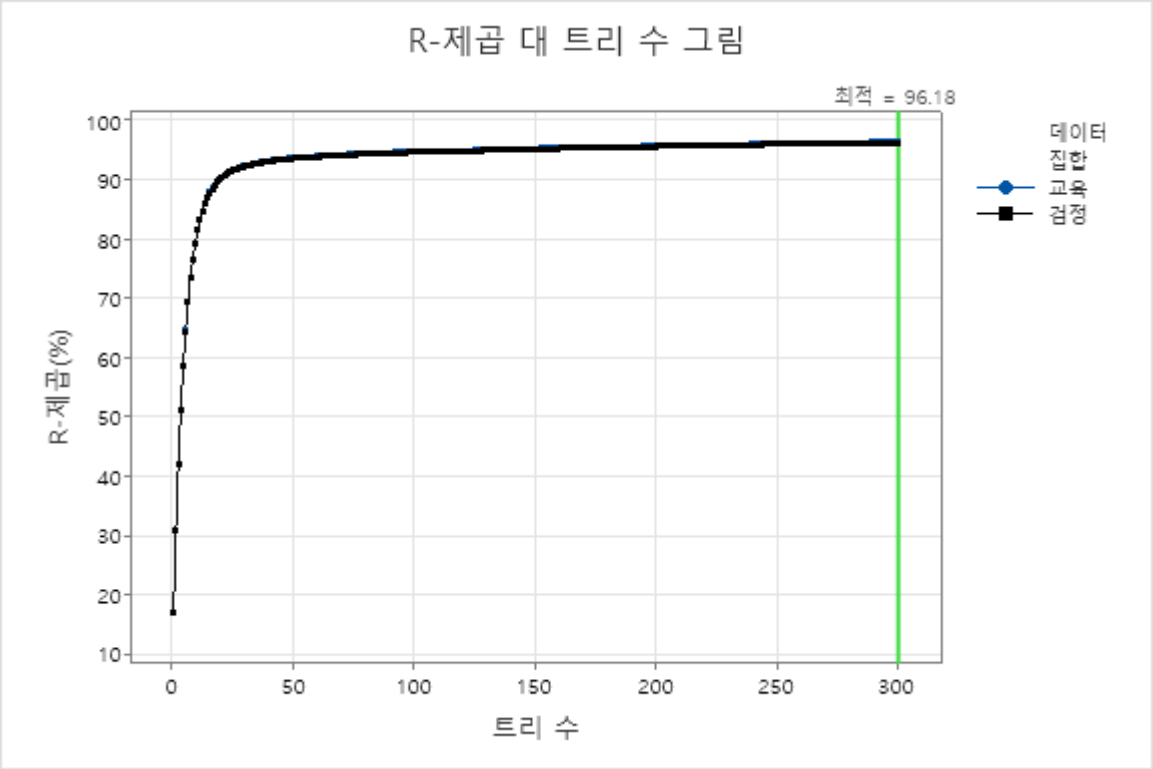
## 중요하지 않은 예측 변수를 제거하여 모형 선택

검정

모형	최적의 트리 수	R-제곱(%)	예측 변수 수 제거된 예측 변수
1	300	95.88	23 없음
2	300	95.87	21 풍향(16방위), 해면기압(hPa)
3	300	95.88	20 풍속(m/s)
4	300	95.93	19 전운량(10분위)
5	300	95.86	18 10cm 지중온도(°C)
6	300	95.85	17 습도(%)
7	300	95.88	16 최저운고(100m )
8	300	95.83	15 시정(10m)
9	300	95.91	14 중하층운량(10분위)
10	300	95.95	13 5cm 지중온도(°C)
11	300	96.07	12 지면온도(°C)
12	300	96.06	11 현지기압(hPa)
13	300	96.02	10 이슬점온도(°C)
14	300	96.06	9 20cm 지중온도(°C)
15	300	95.99	8 일
16*	300	96.18	7 증기압(hPa)
17	300	96.10	6 30cm 지중온도(°C)
18	300	96.11	5 기온(°C)
19	300	95.28	4 시간
20	300	94.52	3 요일
21	283	91.82	2 년도

알고리즘은 각 단계에서 하나의 예측 변수와 중요도가 0인 임의의 예측 변수를 제거했습니다.

\* 선택한 모형에는 최대 R-제곱이 있습니다. 선택한 모형의 출력이 뒤에 나옵니다.

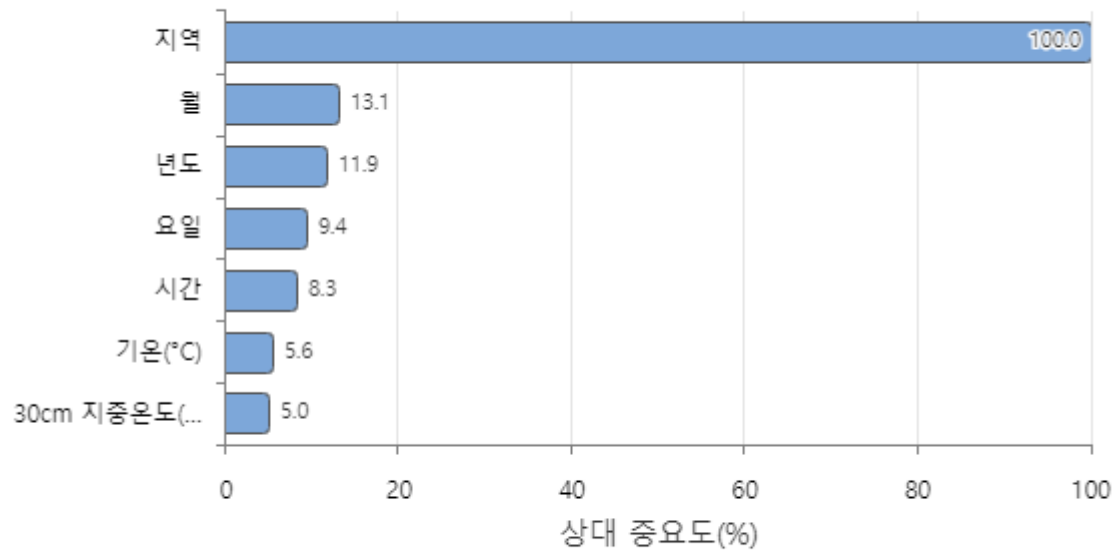


### 모형 요약

전체 예측 변수 7  
 중요 예측 변수 7  
 성장한 트리 수 300  
 최적의 트리 수 300

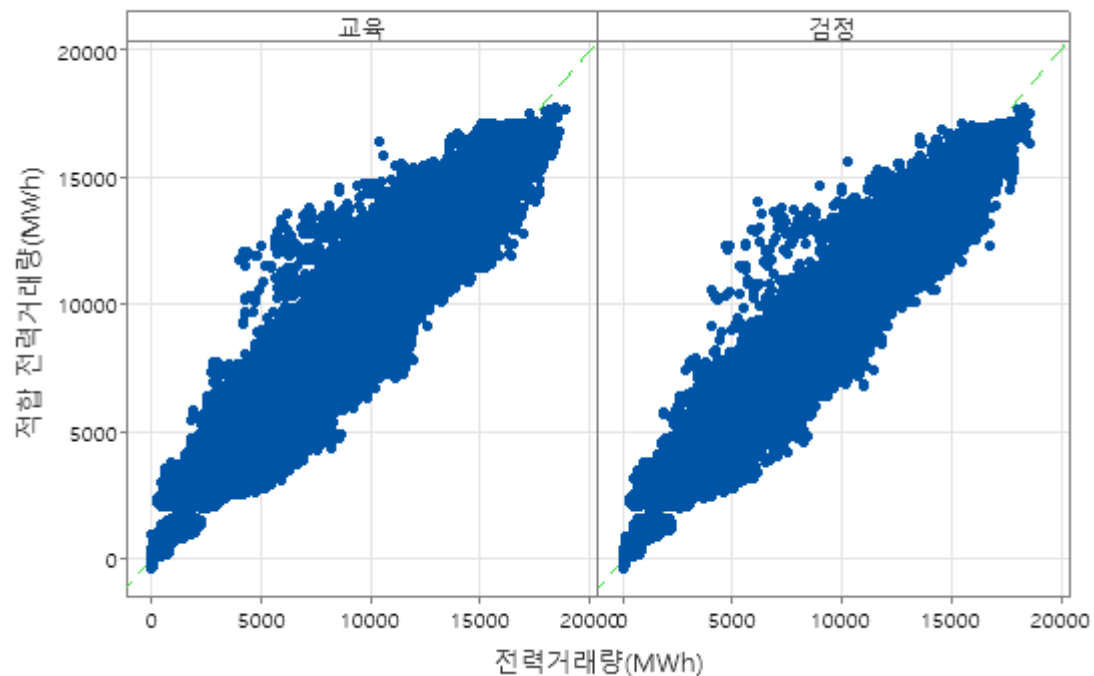
통계량	교육	검정
R-제공	96.24%	96.18%
루트 평균 제곱 오차(RMSE)	761.9778	767.8046
평균 제곱 오차(MSE)	580610.1602	589523.8316
평균 절대 편차(MAD)	521.7512	524.0640
평균 절대 백분율 오차(MAPE)	1.4262	1.4256

### 상대 변수 중요도



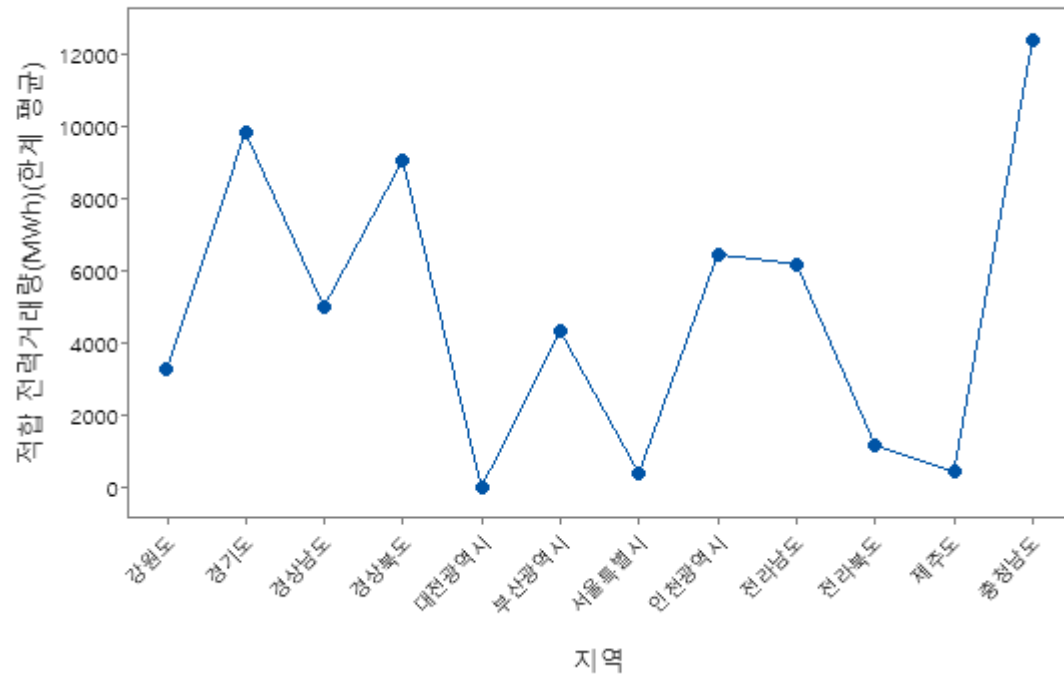
변수 중요도는 예측 변수에 분할을 적용할 때 모형 개선을 측정합니다. 상대 중요도는 상위 예측 변수를 기준으로 한 % 개선으로 정의됩니다.

### 반응 적합치 대 실제 값의 산점도

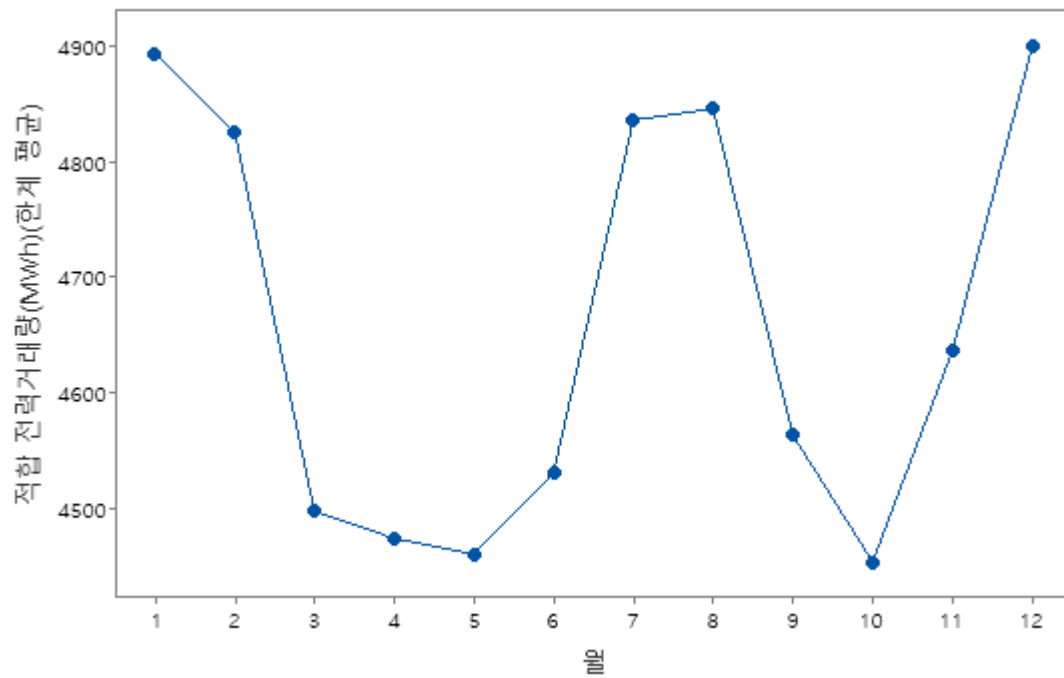


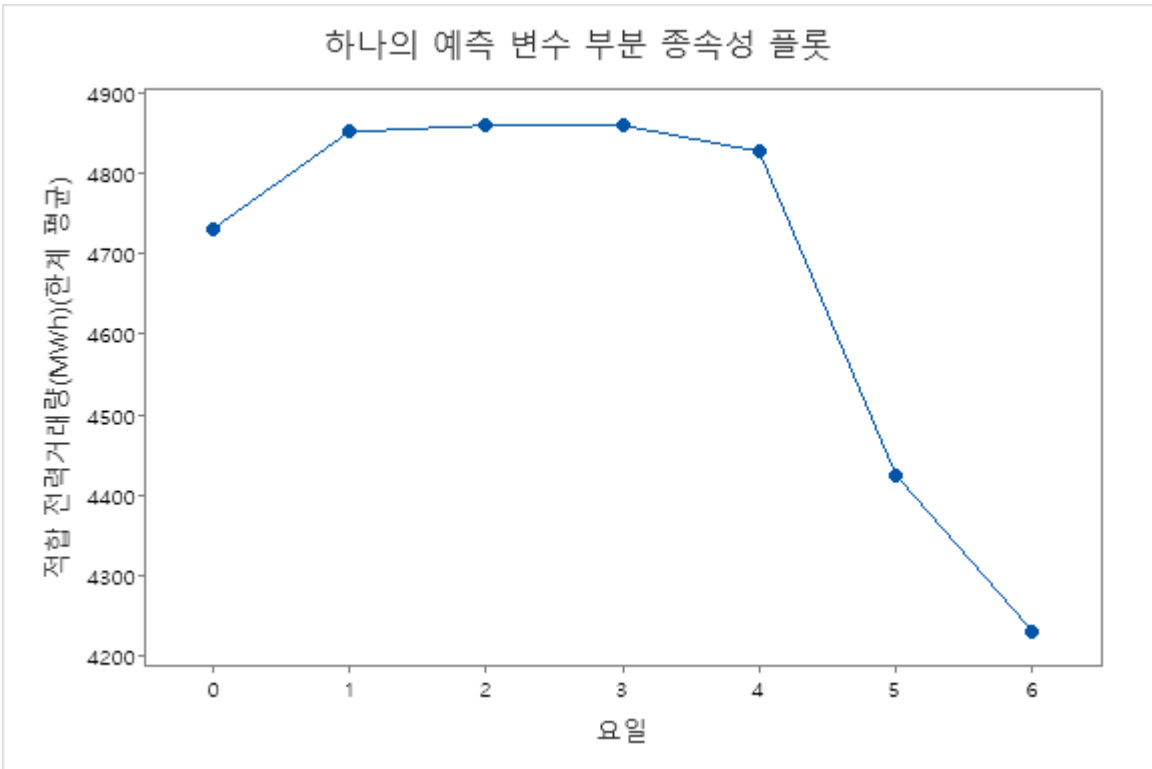
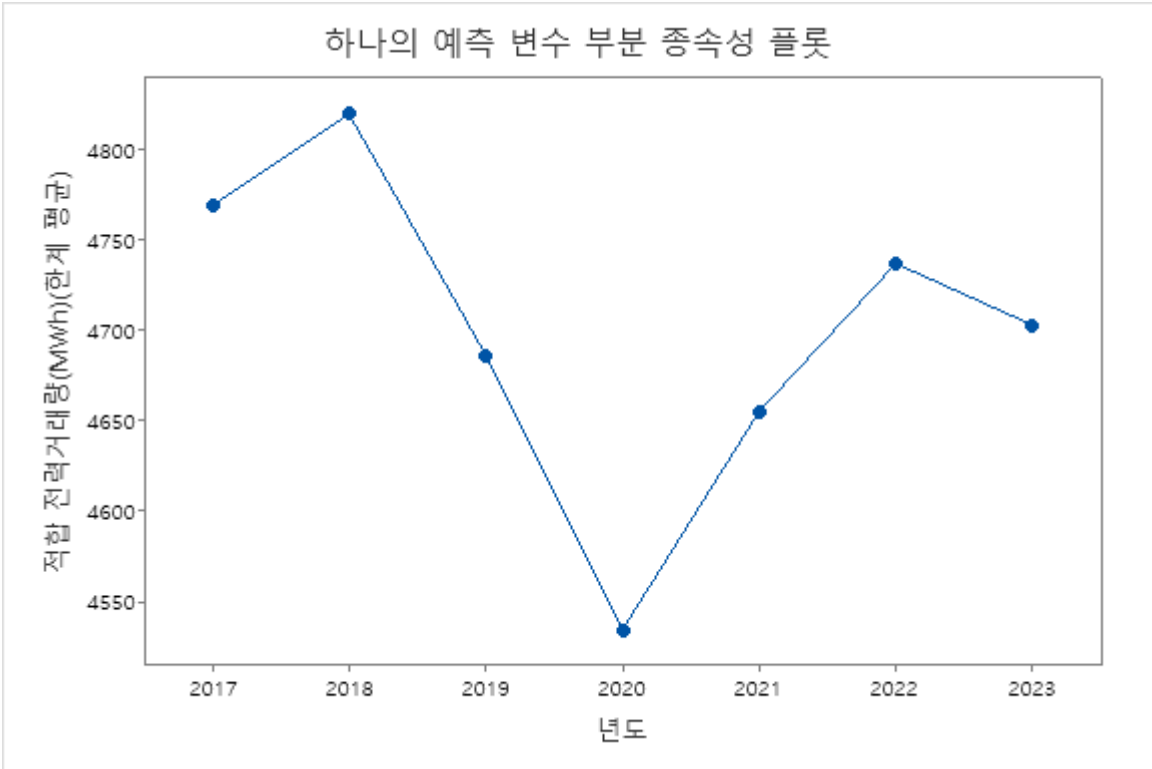
하나의 예측 변수 부분 종속성 플롯

하나의 예측 변수 부분 종속성 플롯



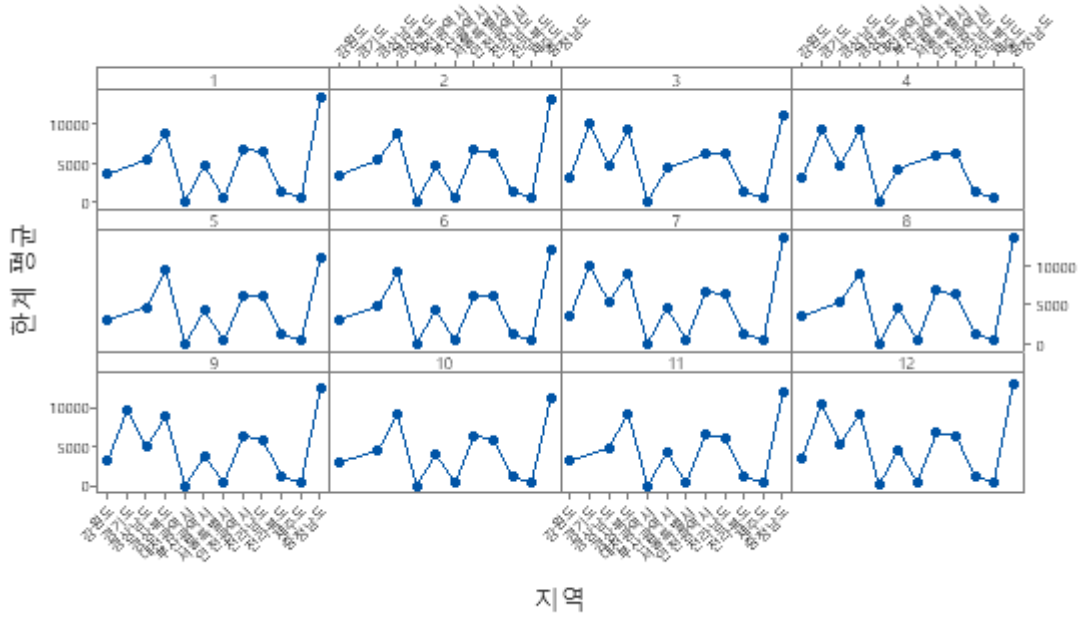
하나의 예측 변수 부분 종속성 플롯





두 개의 예측 변수 부분 종속성 플롯

장착된 분산형 플롯 전력거래량(MWh)



패널 변수: 월