

ALL\_DATA(범주형\_시간추가\_결측치열모두제거) (1).CSV

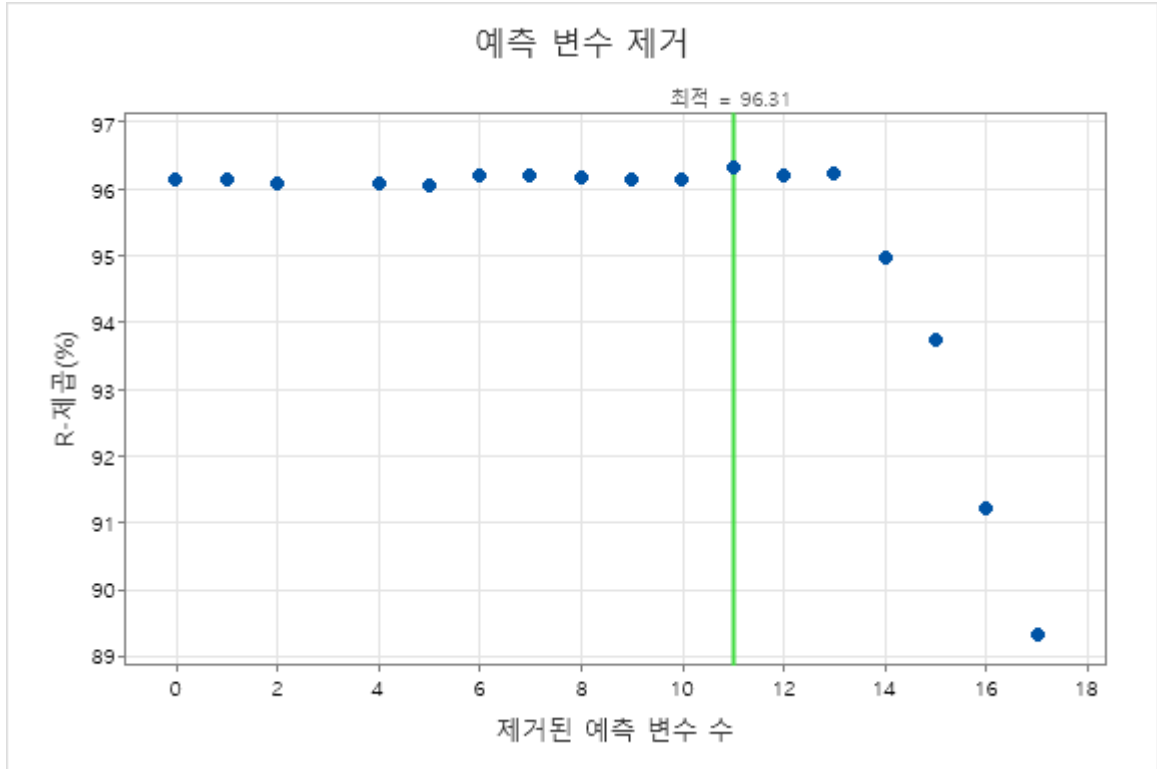
**TreeNet® 회귀 - 주요 예측 변수 검색: 전력거래량(MWh) vs**  
**기온(°C), 풍속(m/s), 풍향(16방위), 습도(%), 증기압(hPa),**  
**이슬점온도(°C), 현지기압(hPa), 해면기압(hPa), 전운량(10분위),**  
**중하층운량(10분위), 시정(10m), 지면온도(°C), 지역, 년도, 월, 일,**  
**요일, 시간**

## 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 = 18
사용된 행	974890
사용되지 않은 행	3926

## 반응 정보

데이터 집합	N	N의 비율(%)	평균	표준 편차	최소값	Q1	중위수	Q3	최대값
교육	682420	70.00	3799.16	3953.52	0	333.25	2798	6220	18838
검정	292470	30.00	3798.31	3951.49	0	334.00	2805	6224	18880



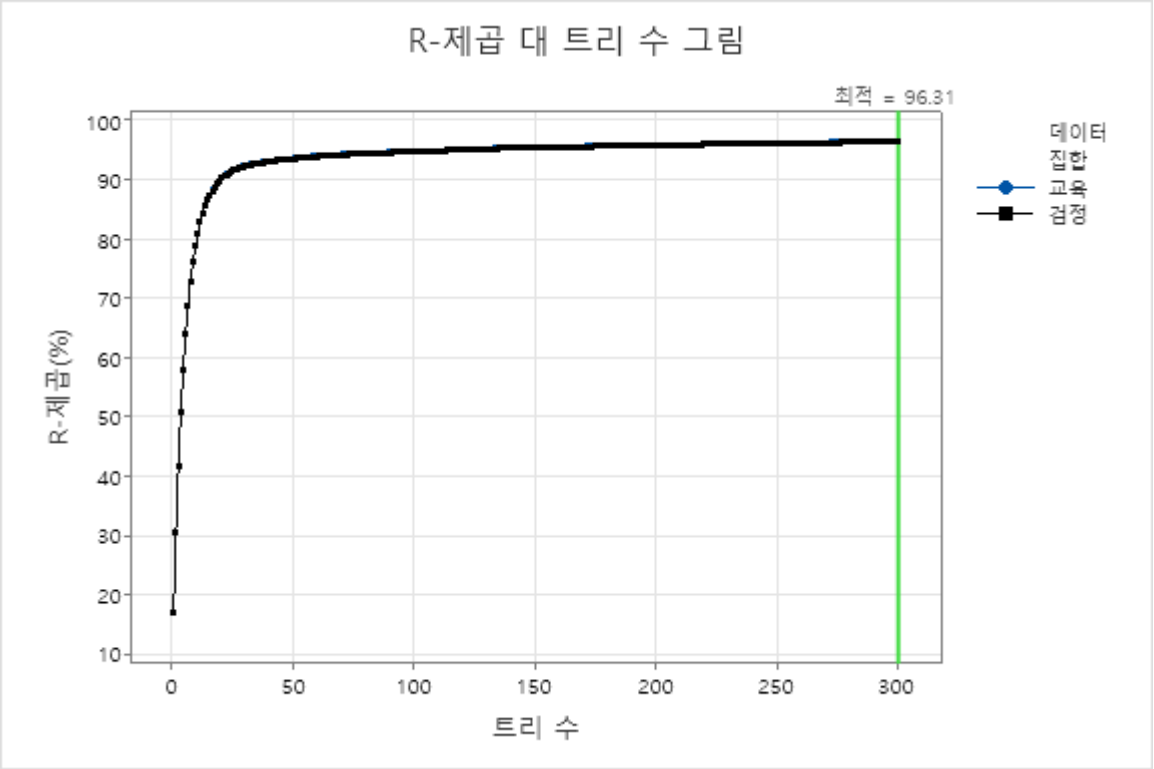
## 중요하지 않은 예측 변수를 제거하여 모형 선택

검정

모형	최적의 트리 수	R-제곱(%)	예측 변수 수	제거된 예측 변수
1	300	96.12	18	없음
2	300	96.13	17	풍향(16방위)
3	300	96.07	16	풍속(m/s)
4	300	96.08	14	해면기압(hPa), 중하층운량(10분위)
5	300	96.05	13	전운량(10분위)
6	300	96.21	12	시정(10m)
7	300	96.20	11	습도(%)
8	300	96.18	10	현지기압(hPa)
9	300	96.14	9	지면온도(°C)
10	300	96.12	8	이슬점온도(°C)
11*	300	96.31	7	일
12	300	96.20	6	증기압(hPa)
13	300	96.21	5	기온(°C)
14	300	94.95	4	요일
15	300	93.75	3	시간
16	300	91.20	2	년도
17	228	89.32	1	월

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

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

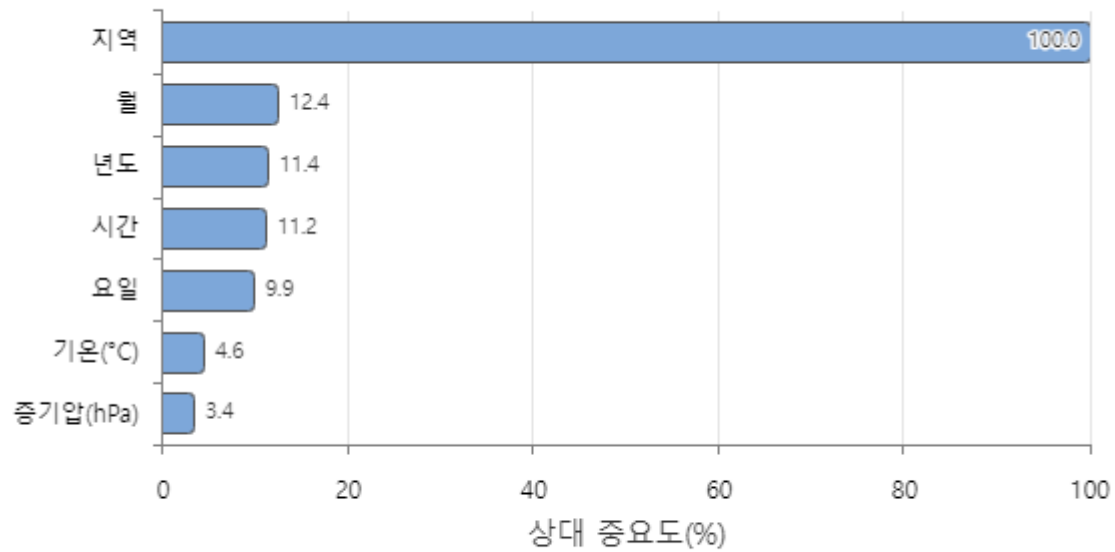


모형 요약

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

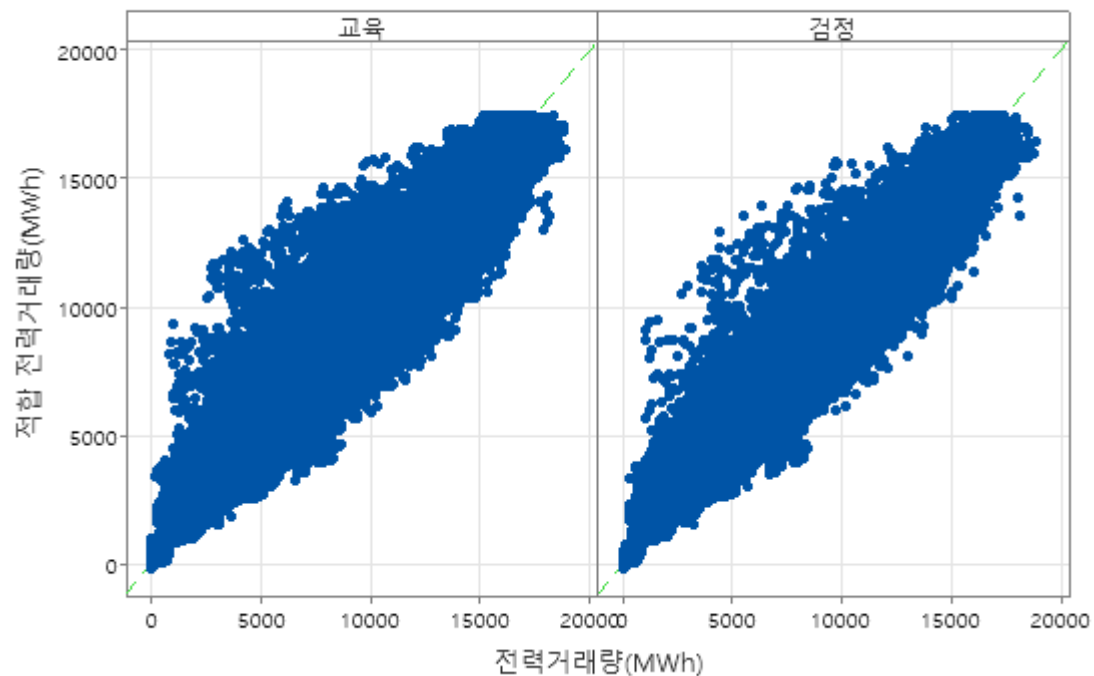
통계량	교육	검정
R-제공	96.33%	96.31%
루트 평균 제곱 오차(RMSE)	757.8773	758.8694
평균 제곱 오차(MSE)	574377.9396	575882.7714
평균 절대 편차(MAD)	479.2243	480.2689
평균 절대 백분율 오차(MAPE)	1.5741	1.5887

### 상대 변수 중요도



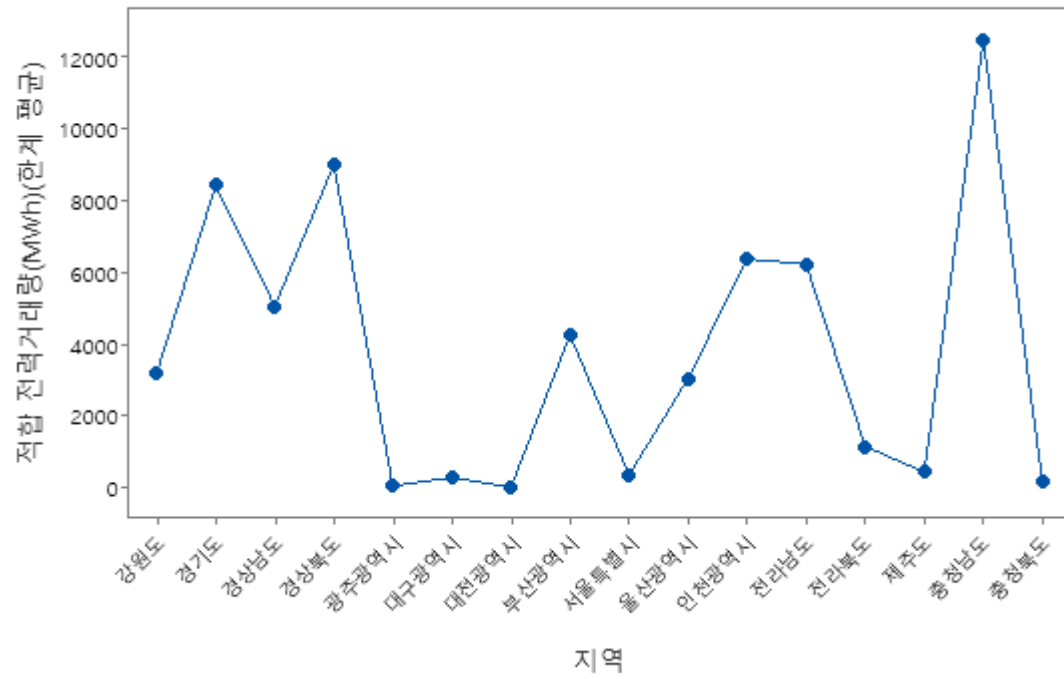
변수 중요도는 예측 변수에 분할을 적용할 때 모형 개선을 측정합니다. 상대 중요도는 상위 예측 변수를 기준으로 한 % 개선으로 정의됩니다.

### 반응 적합치 대 실제 값의 산점도

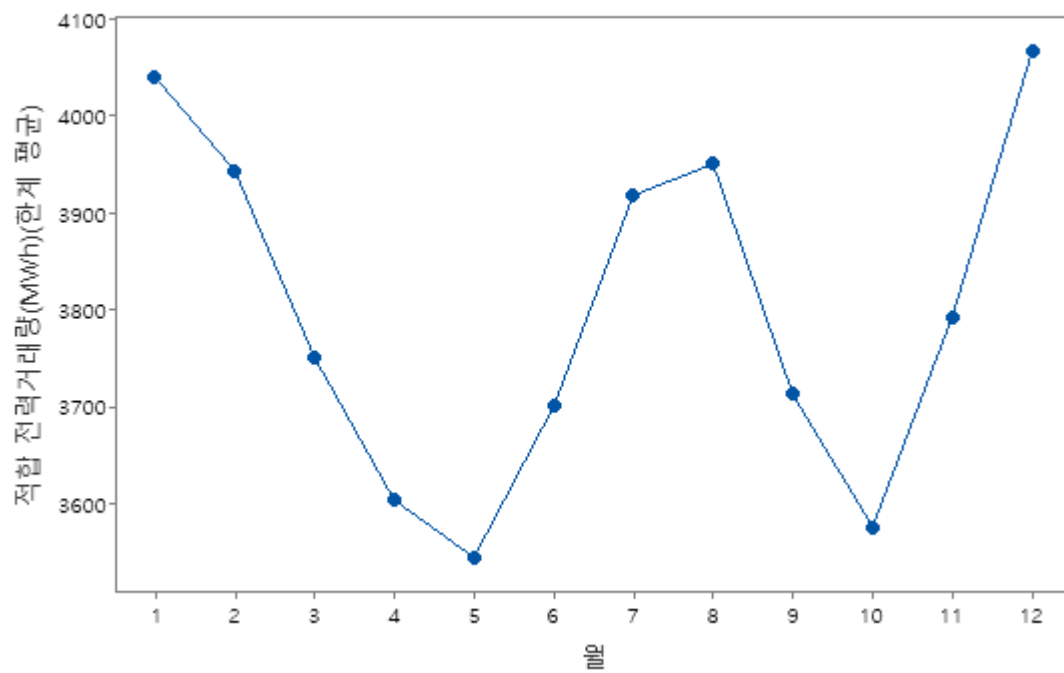


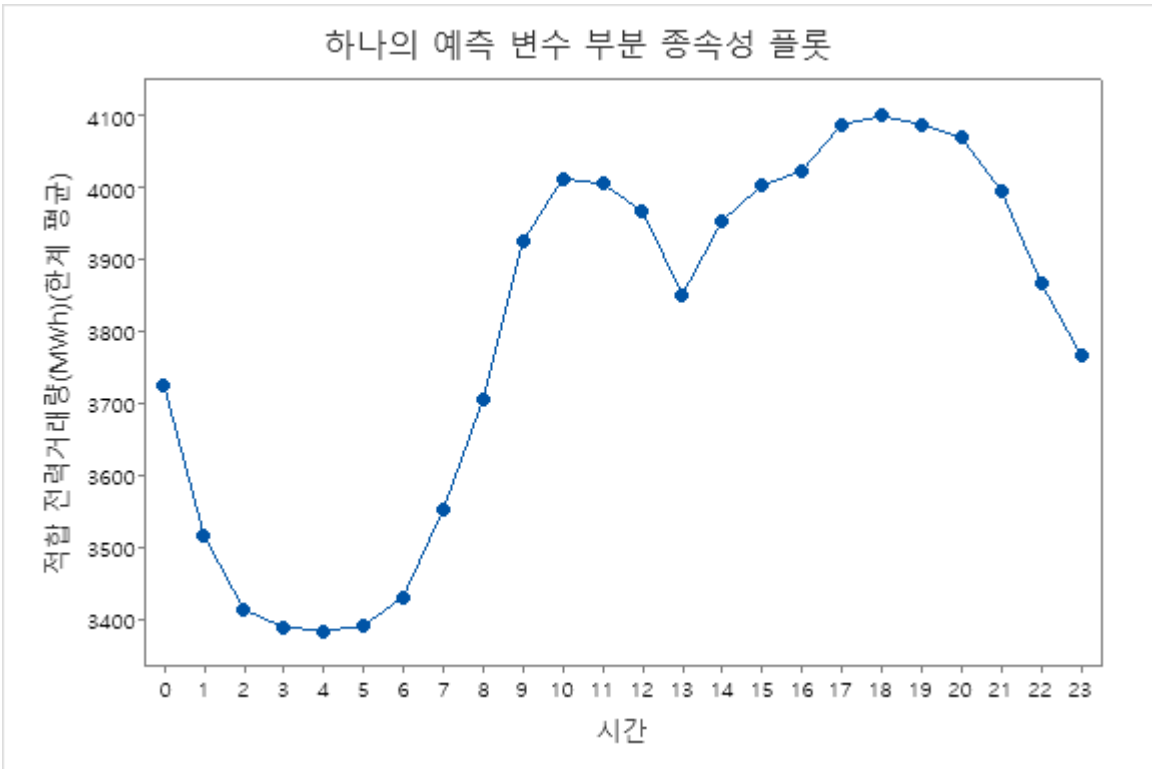
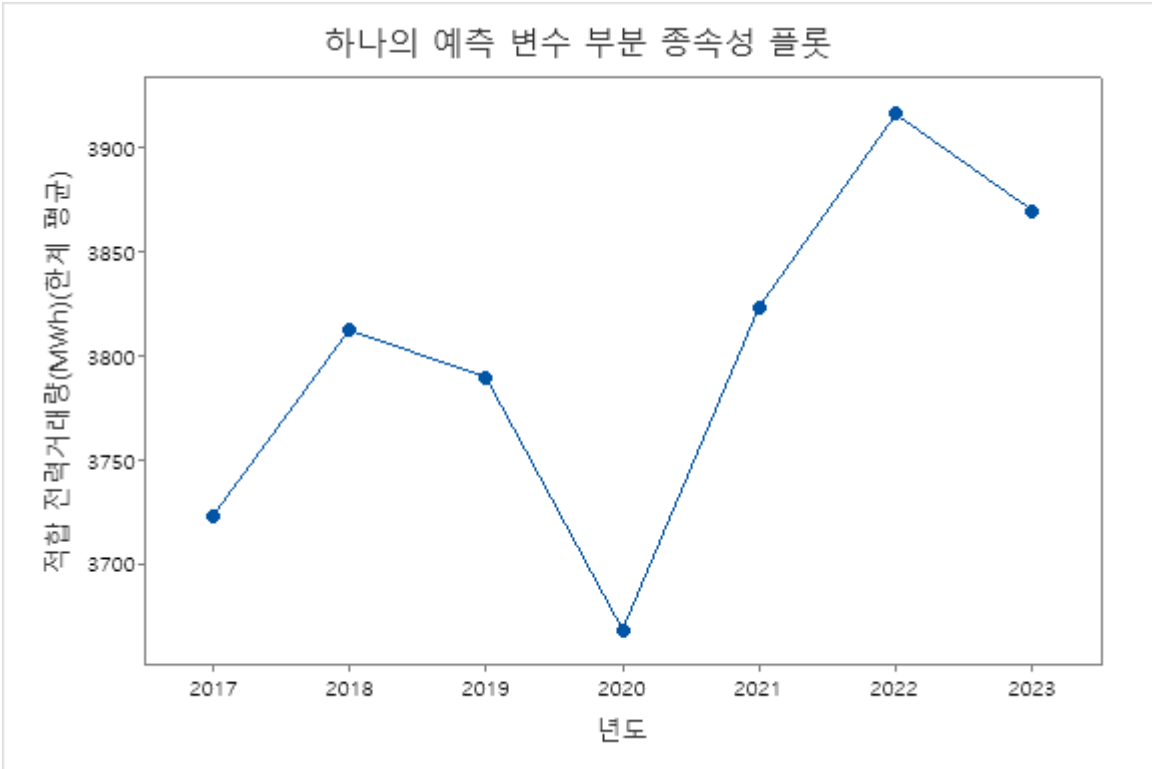
하나의 예측 변수 부분 종속성 플롯

하나의 예측 변수 부분 종속성 플롯



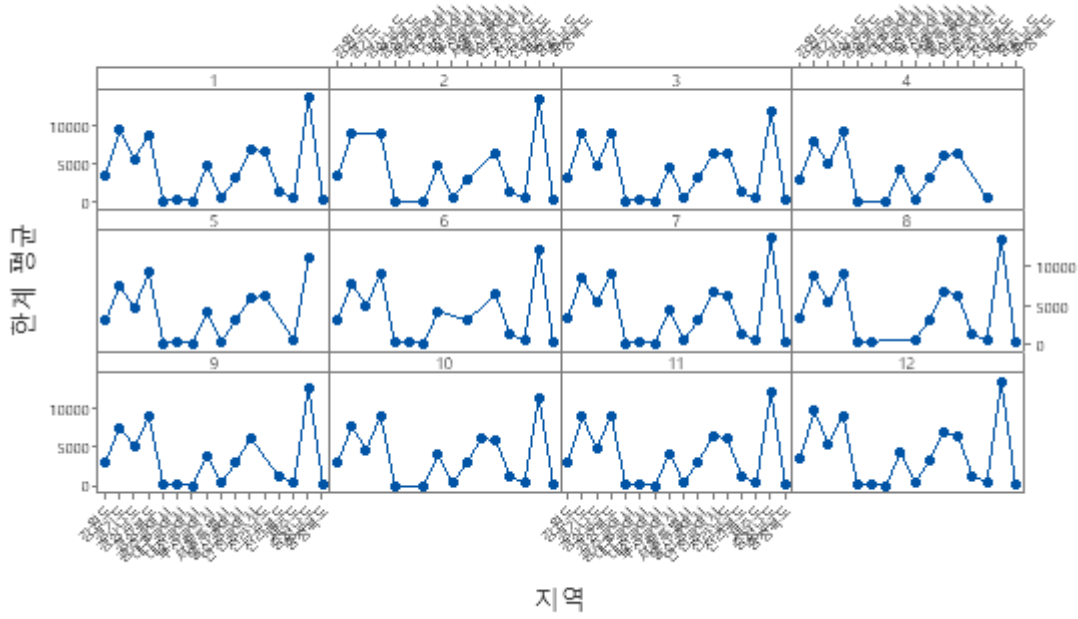
하나의 예측 변수 부분 종속성 플롯





두 개의 예측 변수 부분 종속성 플롯

장착된 분산형 플롯 전력거래량(MWh)



패널 변수: 월