

# 기상청 날씨 빅데이터 콘테스트 보고서

주제: 기상 데이터를 활용한 시청률 예측

팀 : D&A

김동규, 박예리, 신민용, 조단비

# 목차

1. 공모 배경

2. 활용 데이터 정의

3. 데이터 처리 방안 및 활용 분석 기법

4. 분석 결과

5. 서비스 활용 방안

6. 서비스 기대효과

# 1. 공모 배경

날씨와 시청률. 어떤 관계가 있을까.

날씨 정보는 우리의 일상생활을 좌우하는 변수 중 하나이다. 출근을 하거나 학교를 가기 전, 휴가 계획을 잡거나 쇼핑을 할 때에도 우리는 날씨를 고려하게 된다. '폭염 주의보'에 오늘 어떤 옷을 입어야 할지 고민하고 '다음 주 비 소식'에 휴가 계획을 조정하기도 한다.

기상체계는 이렇듯 사회 전반적으로 직·간접적인 영향력을 행사한다. 날씨 정보의 영향력이 소비자의 행동을 결정하는 주 요인임을 일찌감치 알아챈 기업들은 '날씨 마케팅'을 하나의 경영 전략에 이용하고 있다.

세계기상기구(WMO)에 따르면 전 세계 날씨정보의 활용가치는 연간 3조 5000억에서 6조 5000억에 달한다고 한다. 지난 2017년 1월 제주도에 32년 만의 폭설이 내렸고 지난해 여름에는 올해 못지않은 폭염이 150년 만에 찾아왔다. 기사로 흔히 접할 수 있게 된 요즈음, 이러한 기상이변에 의한 세계적 경제 손실은 연간 약 2800조원에서 3351조원에 이른다. 날씨 정보의 영향력이 날로 커질 수밖에 없는 이유이다.

지난 5월 컨슈머인사이트와 세종대학교 관광산업연구소는 '기상 문제가 발생한다는 일기 예보에 여행을 취소할 것인가' 는 질문에 54% 응답자가 '여행 계획을 취소할 것'으로 답했다고 발표했다. 위 응답자에는 젊은 층보다 나이가 많을수록, 미혼보다는 기혼, 남성보다 여성이 기상 상황에 더 민감하게 반응했다.

조사 결과를 바탕으로 질문을 넓혀 보려고 한다. '나이가 많은 기혼 여성'이 악천후를 예보하는 일기를 듣고 일정을 취소했을 시 보편적으로 취할 수 있는 행동은 무엇일까. 더 나아가 폭우 주의보와 열대야에 대부분의 사람들은 어떤 장소로 향할까. 그 장소가 '실내'일 것이라는 예상에 기반하여 '날씨가 좋지 않을 때 집안에서 tv 시청을 할 것이다.' 는 가설을 착안했다.

날씨 정보를 이용해 시청률을 예측한다면 어떨까. 또한 예측된 시청률로 창출 되는 경제 효용 가치는 얼마나 될까. 날씨에 따라 시청률을 분석 및 예측 할 수 있다면 이를 가장 원하는 타겟은 누구 일까. 바로 방송사와 광고주라고 생각 할 수 있다. 방송사의 경우, 기상 및 특보 변수에 의해 영향 받는 시청률을 알게 된다면 프로그램 유치를 더욱 효과적으로 할 수 있을 것이다. 2016년 기준 지상파 방송사업자의 TV 방송광고매출 총규모는 1조 3999억 원으로 큰 시장을 보여주고 있다. 방송광고매출에 가장 큰 영향은 시청률이다. 시청률을 높이는 방법도 중요하지만 시청률을 예측 할 수 있다면 광고매출 규모를 키울 수 있을 것이다. 날씨가 좋지 않을 때 시청률이 높아질 것이라는 객관적인 분석 결과가 있다면 시청자들이 TV앞을 지키도록 재밌는 프로그램 방영을 고려해 볼 수 있을 것이다.

방송사는 시청자의 니즈를 만족시키는 것이라면 광고주는 시청률 자체의 수치를 통해 광고 런칭을 고려 할 수 있다. 무조건 높은 시청률에 큰 비용을 들여서 광고를 하는 것이 아닌, 날씨를 통해 시청률을 예측하고 같은 비용이지만 최대 효과를 내도록 광고를 런칭 할 수 있을 것이다. 날씨와 시청률의 관계를 분석하고 이를 서비스화 한다면 광고 런칭시 날씨를 바꾸거나 시간대를 조정하여 같은 비용으로 더 좋은 광고 효과를 누릴 수 있을 것이다.

이를 토대로 날씨정보와 시청률 간의 관계를 분석하고, 이를 예측하며 실제 서비스 방안을 살펴보고자 한다.

## 2. 활용 데이터 정의

1. 기상 데이터
- a. 기상청 빅데이터 플랫폼 '날씨누리'에서 시청률과 연관 있을 것 같은 16개의 변수 데이터 수집 (2009-01-01 ~ 2018-06-30)

b. 수집한 데이터를 활용하여 기상특보 발표 기준에 따라 특보 변수 7개 추가 생성

c. 특보일 경우, 사람들이 실내에 있을 확률이 높아질 것이고 이에 따라 시청률이 높아질 것이라 기대

<기상 특보 발표 기준>

종류	주의보 기준
호우	강수량이 110mm이상일 때
강풍	최대순간풍속이 14m/s 이상일 때
대설	최심적설이 5cm이상일 때
건조	습도가 35% 이하일 때
한파	최저기온이 영하12도 이하일 때
폭염	최고기온이 33도 이상일 때
황사	미세먼지 농도가 30미만일 때 좋음 30이상, 80미만일 때 보통 80이상, 150미만일 때 나쁨 150이상일 때 매우 나쁨

	Date	평균기온	최저기온	최고기온	강수량	최대순간풍속	풍속	습도	일조시간	일사량	최심적설	X3 시간 신적설	전날	지면온도	불쾌지수	체감온도	미세먼지	요일	호우	강풍	대설	건조	한파	폭염	황사
1	2009-01-01	-5.8	-9.5	-2.5	0.0	7.1	2.3	45.8	8.8	9.49	0.0	0.0	0.0	-6.2	32.42276	-6.1030603	43.5	목요일	0	0	0	0	0	0	보통
2	2009-01-02	-2.8	-6.9	1.2	0.0	5.9	1.6	51.5	8.8	8.87	0.0	0.0	0.0	-4.9	35.23992	-2.0751083	45.8	금요일	0	0	0	0	0	0	보통
3	2009-01-03	-1.0	-5.6	4.0	0.0	6.0	1.8	48.3	8.8	8.73	0.0	0.0	0.9	-3.2	38.10493	-0.4282961	46.8	토요일	0	0	0	0	0	0	보통
4	2009-01-04	0.4	-2.8	4.5	0.0	6.4	1.9	44.3	8.4	9.02	0.0	0.0	2.0	-1.2	40.46453	0.9446292	40.1	일요일	0	0	0	0	0	0	보통
5	2009-01-05	-1.6	-3.4	1.2	0.0	7.8	2.2	53.1	4.9	6.32	0.0	0.0	1.6	-3.5	36.56960	-1.4928627	58.5	월요일	0	0	0	0	0	0	보통
6	2009-01-06	-2.0	-5.4	1.4	0.0	4.6	1.7	50.0	5.0	6.19	0.0	0.0	3.6	-3.6	36.54000	-1.3637484	96.0	화요일	0	0	0	0	0	0	나쁨
7	2009-01-07	-0.5	-4.2	4.2	0.0	5.7	1.8	42.0	8.6	8.21	0.0	0.0	0.5	-2.8	39.68110	0.1002533	83.1	수요일	0	0	0	0	0	0	나쁨
8	2009-01-08	-0.9	-4.0	2.5	0.0	6.2	1.8	42.0	6.2	7.07	0.0	0.0	3.0	-3.6	39.19078	-0.3225863	72.2	목요일	0	0	0	0	0	0	보통
9	2009-01-09	-3.5	-6.8	-0.5	0.0	9.6	3.0	43.0	6.9	6.89	0.0	0.0	0.5	-4.3	35.82605	-4.2646993	65.3	금요일	0	0	0	0	0	0	보통
10	2009-01-10	-7.5	-9.9	-4.3	0.0	9.0	3.5	42.0	8.8	10.00	0.0	0.0	0.4	-7.2	31.10050	-9.0684925	54.1	토요일	0	0	0	0	0	0	보통
11	2009-01-11	-8.0	-9.6	-4.2	0.0	13.8	3.3	39.5	8.9	10.21	0.0	0.0	0.4	-8.4	31.04310	-9.4549871	41.5	일요일	0	0	0	0	0	0	보통
12	2009-01-12	-7.8	-11.0	-3.2	0.0	11.8	4.1	47.5	8.1	9.30	0.0	0.0	0.5	-8.0	29.52155	-9.8534003	52.9	월요일	0	0	0	0	0	0	보통
13	2009-01-13	-5.5	-9.0	-1.5	0.0	6.5	2.1	45.3	8.9	10.21	0.0	0.0	1.4	-7.1	32.90051	-5.5569811	42.5	화요일	0	0	0	0	0	0	보통
14	2009-01-14	-6.6	-9.4	-3.2	0.0	10.9	3.5	39.6	9.0	10.32	0.0	0.0	0.4	-6.6	32.70374	-8.0730922	62.4	수요일	0	0	0	0	0	0	보통
15	2009-01-15	-6.3	-11.2	-2.0	0.0	6.6	2.1	34.4	8.7	10.41	0.0	0.0	2.5	-7.5	34.13227	-6.4113620	38.6	목요일	0	0	0	1	0	0	보통
16	2009-01-16	-2.8	-4.8	0.4	3.3	5.5	2.0	80.8	1.8	4.16	5.1	5.1	6.0	-3.1	30.23782	-2.5641669	53.8	금요일	0	0	1	0	0	0	보통
17	2009-01-17	-1.4	-5.9	3.1	0.0	5.8	2.3	76.8	7.5	9.10	1.5	0.0	0.9	-3.3	33.11915	-1.3751619	94.1	토요일	0	0	0	0	0	0	나쁨
18	2009-01-18	2.6	0.1	6.1	0.0	9.9	2.4	71.8	0.3	3.34	0.0	0.0	7.8	0.3	39.98673	2.8422122	108.5	일요일	0	0	0	0	0	0	나쁨
19	2009-01-19	0.6	-2.7	4.7	0.0	9.7	2.6	48.0	7.8	9.85	0.0	0.0	2.3	-0.6	40.20712	0.5219187	81.8	월요일	0	0	0	0	0	0	나쁨

## < 날씨 데이터 셋 >

## 2. 시청률 데이터

### a. '닐슨코리아'에서 일별 시청률 데이터 TOP20 크롤링

#### < 시청률 정의 >



출처 : [http://www.nielsenkorea.co.kr/research\\_methodology.asp?menu=Tit\\_4](http://www.nielsenkorea.co.kr/research_methodology.asp?menu=Tit_4)

- b. 시청자수 데이터는 2017년부터 제공 되어 데이터수 부족, 따라서 시청률 데이터 수집  
 시청률일 지라도 해당 날짜의 합이 100이 아니라 그시간대에 해당하는 시청률이므로 비교 및 분석  
 데이터로 생각해도 무방할 것이라 판단
- c. 장르와 방송 시간대 정보는 포털사이트에서 크롤링하여 수집
- d. 제약조건 :
  - 모든 프로그램에 대한 시청률이 아닌 해당 날짜 top20만 얻을 수 있음
  - top20에 속한 장르의 범위가 예능과 드라마에 지나치게 치중되어 시사교양 데이터를 얻는데  
 어려움이 있음
  - 따라서, 예능과 드라마에 중점을 두고 분석 실시
  - 결방, 시간 지연, 변경 등에 대한 데이터 전처리 수작업 필요

< 시청률 데이터 셋 >

Date	Name	Viewrate	Month	Genre	Time
2009-01-01	수목미니시리즈(종합병원2)	20.3	목요일	드라마	22
2009-01-01	아침연속극(순결한당신)	11.3	목요일	드라마	8
2009-01-01	일일시트콤(그분이오신다)	12.6	목요일	드라마	19
2009-01-01	일일연속극(너는내운명)	41.0	목요일	드라마	20
2009-01-01	특집놀라운대회스타킹왕중왕전 <재>	11.5	목요일	예능	19
2009-01-01	해피투게더	15.7	목요일	예능	23
2009-01-02	MBC아침드라마(하얀거짓말)	14.5	금요일	드라마	8
2009-01-02	W	12.5	금요일	드라마	22
2009-01-02	남희석최은경의여유만만	10.9	금요일	예능	10

### 3. 기타 데이터

- a. 분석 및 예측에 도움이 될 것같은 변수 추가 수집
- b. 'SKT 데이터 허브'에서 일별 배달 데이터 수집
- c. 일별 구글 트렌드 수집 (주단위라서 7일마다 같은값으로 처리)

Date	Name	Trend
2013-07-22	아침연속극(당신의여자)	91
2013-07-23	아침연속극(당신의여자)	91
2013-07-24	아침연속극(당신의여자)	91
2013-07-25	아침연속극(당신의여자)	91
2013-07-26	아침연속극(당신의여자)	91
2013-07-29	아침연속극(당신의여자)	100
2013-07-30	아침연속극(당신의여자)	100
2013-07-31	아침연속극(당신의여자)	100
2013-08-01	아침연속극(당신의여자)	100
2013-08-02	아침연속극(당신의여자)	100
2013-08-05	아침연속극(당신의여자)	26

< 구글 트렌드 데이터 셋 >

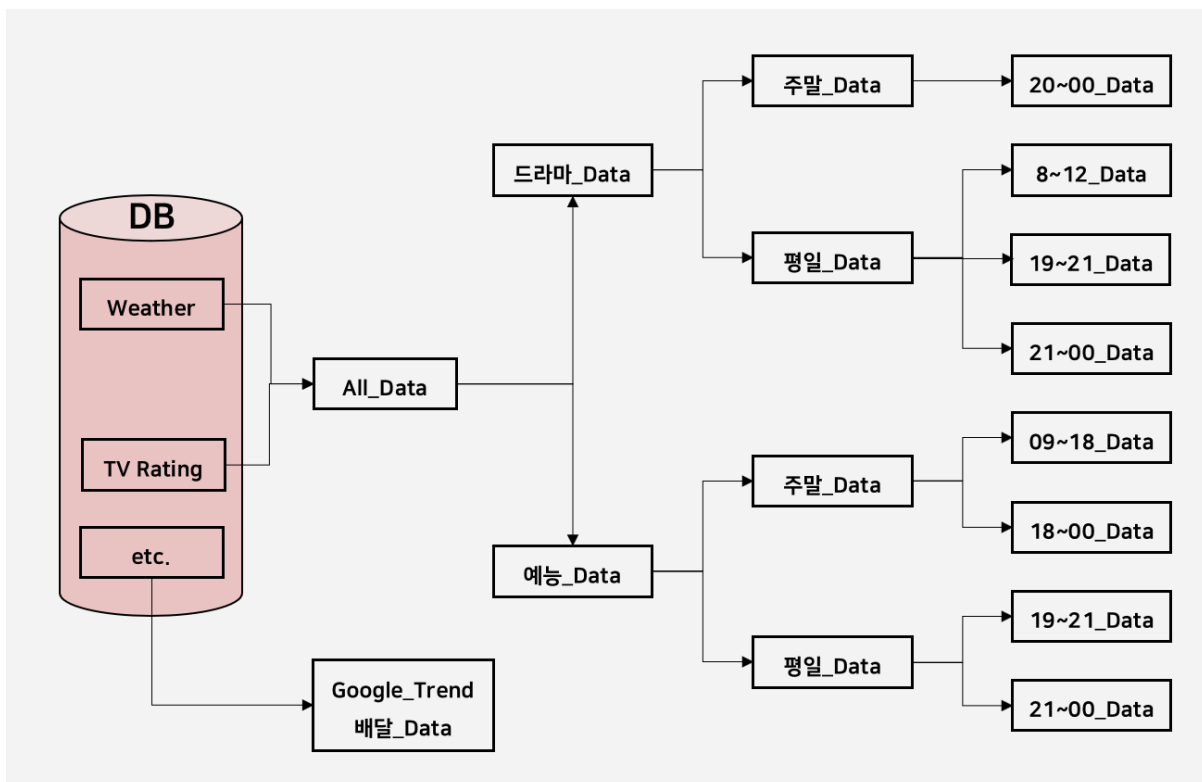
Date	Call
2013-08-01	60829
2013-08-02	69136
2013-08-03	82485
2013-08-04	82276
2013-08-05	56185
2013-08-06	60502
2013-08-07	58637
2013-08-08	61436
2013-08-09	72126
2013-08-10	86000
2013-08-11	89180

< 배달 데이터 셋 >

### 3. 데이터 처리 방안 및 활용 기법

#### ㄱ. 데이터 처리 방안

< 분석용 데이터셋 >



##### 1. All\_Data :

- 날씨와 시청률 데이터를 날짜별로 시청률 집계
- Top20만 있는 시청률 데이터를 고려해서 집계시 20개의 시청률을 더함.

- C. 행 3468개 (2009-01-01 ~ 2018-06-30)  
열 20개 (날짜, 시청률, 평균기온, 강수량, 최대순간풍속, 풍속, 습도, 일사량,  
최심적설, 3시간\_신적설, 전운량, 미세먼지, 요일, 호우, 강풍, 대설,  
건조, 한파, 폭염, 황사)

## 2. 예능, 드라마 \_Data :

- 날씨와 시청률 데이터를 각 장르별로 나누고 날짜별로 시청률 집계
- 장르별로 데이터 셋을 나눌 경우, 날짜마다 top20에 속한 해당 장르의 수가 다르기 때문에 집계방식은 평균 사용
- 행과 열 1번과 동일

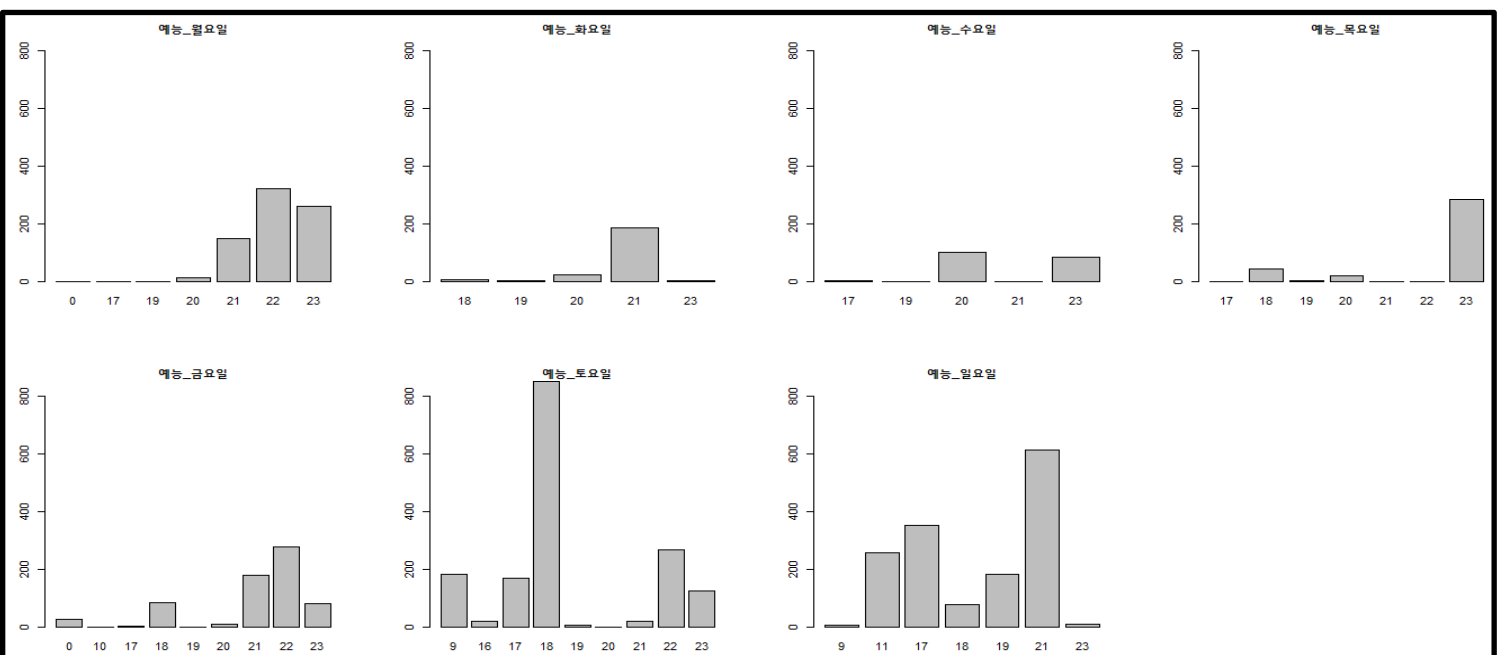
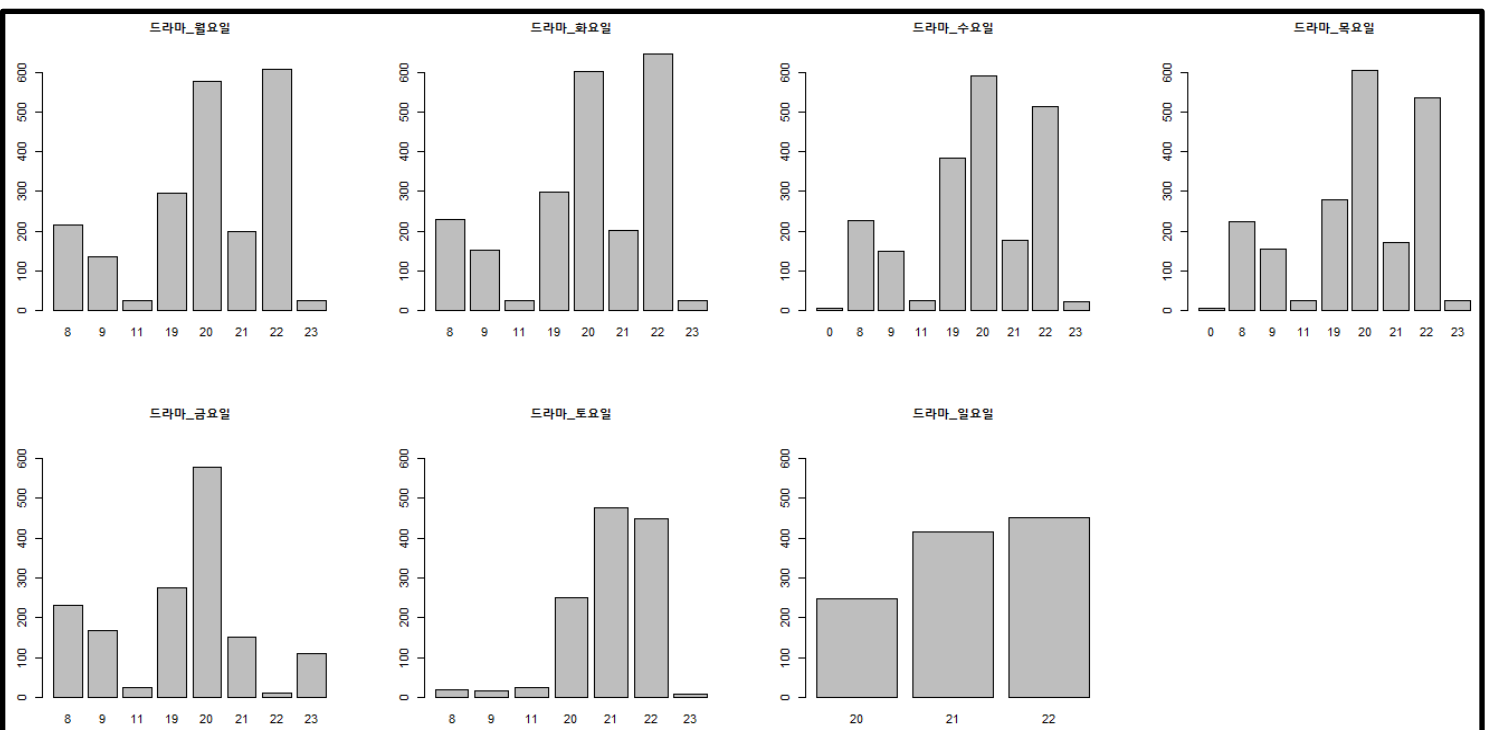
### 3. 주말, 평일 \_Data :

- 요일의 특성을 고려하여 장르별로 나눈 데이터 셋을 주말과 평일로 분할
- 2번과 마찬가지로 집계방식은 평균 사용
- 행 -> 주말 드라마 (963개), 평일 드라마 (2467개),  
주말 예능 (986개), 평일 예능 (1543개)  
열 -> 1번과 동일

#### 4. 시간대별 \_Data :

- TV 특성상 요일 뿐만 아니라 방송 시간에 따라 차이가 존재하기 때문에 시간대별로 분할
- 시각화를 통해 시간대별 데이터셋 수를 편향 되지 않게 분할
- 시각화 결과 평일, 주말에 따라 다른 시간대 군집 형성  
 드라마\_평일(8~12, 19~21, 21~00) / 드라마\_주말(20~00) /  
 예능\_평일(19~21, 21~00) / 예능\_주말(09~18, 18~00)





< 장르별 시간대 분포 >

## 5. ect. data :

- 분석 후 예측시 성능을 높이기 위한 변수 데이터
- 날짜별 배달데이터와 프로그램명 구글 트렌드 사용
- 배달데이터의 경우 배달주문 건수가 높아질수록 집에서 TV를 시청 할 것을 가정
- 구글 트렌드의 경우 해당 프로그램의 검색량이 많을수록 시청률이 높을 것으로 가정
- 2013년 7월 1일 ~ 2018년 6월 30일 총 5년 데이터 적용

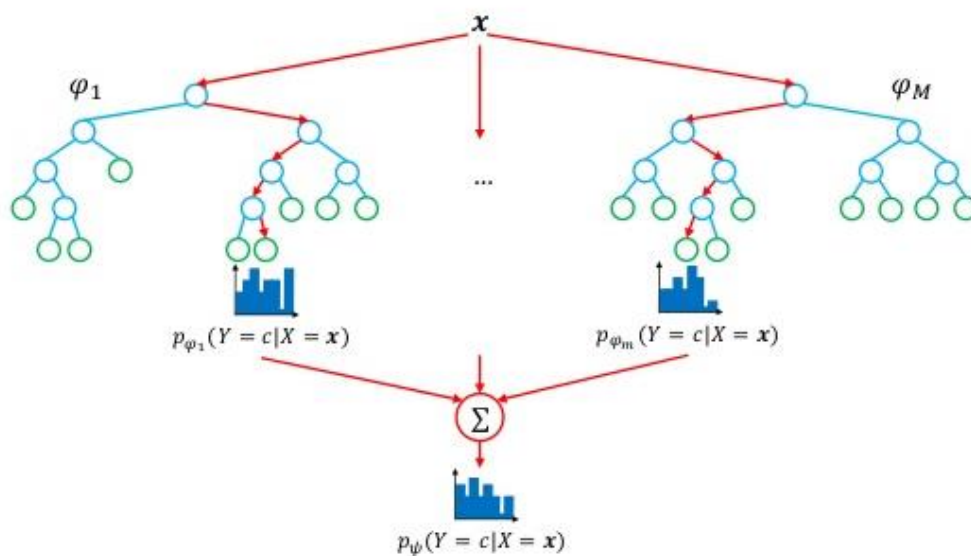
## L. 활용 기법

### 1. 시각화를 통한 분석

- 상관계수행렬 plot을 통해 날씨 데이터의 다중공선성 제거, 변수 선택
- ggplot을 통해 각 데이터셋과 수치형 날씨변수 관계 파악
- boxplot을 통해 각 데이터셋과 범주형 날씨(특보)변수 관계 파악
- 그래프 해석시 해당 변수 회귀 적합을 통해 정확한 의미 파악
- 시각화를 통해 얻은 분석결과를 통해 변수 재 선택
- 시계열 그래프를 통해 시계열 자료임을 확인

### 2. 예측 모델 구현

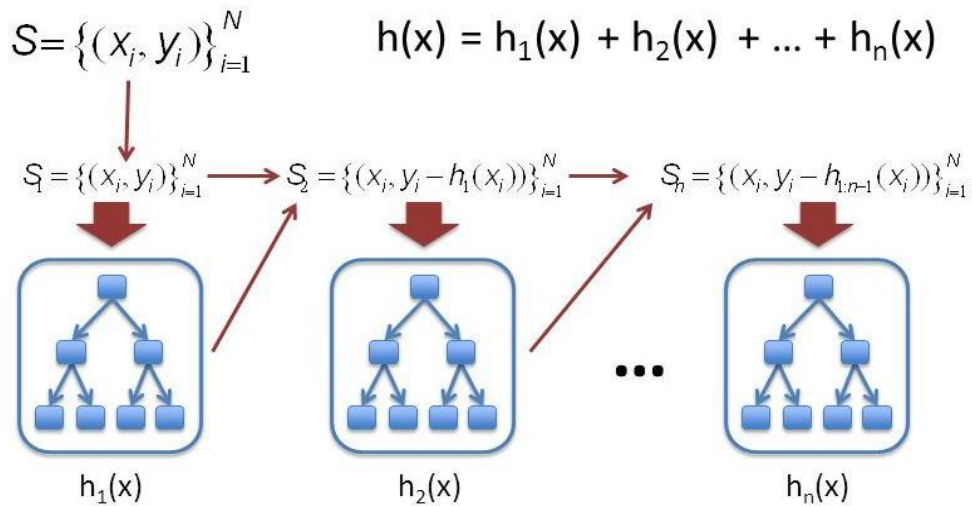
- 앙상블 모형 중 배깅 기법의 Random Forest 구현



< Random Forest 모형 >

b. 앙상블 모형 중 부스팅 기법의 xgboost 구현

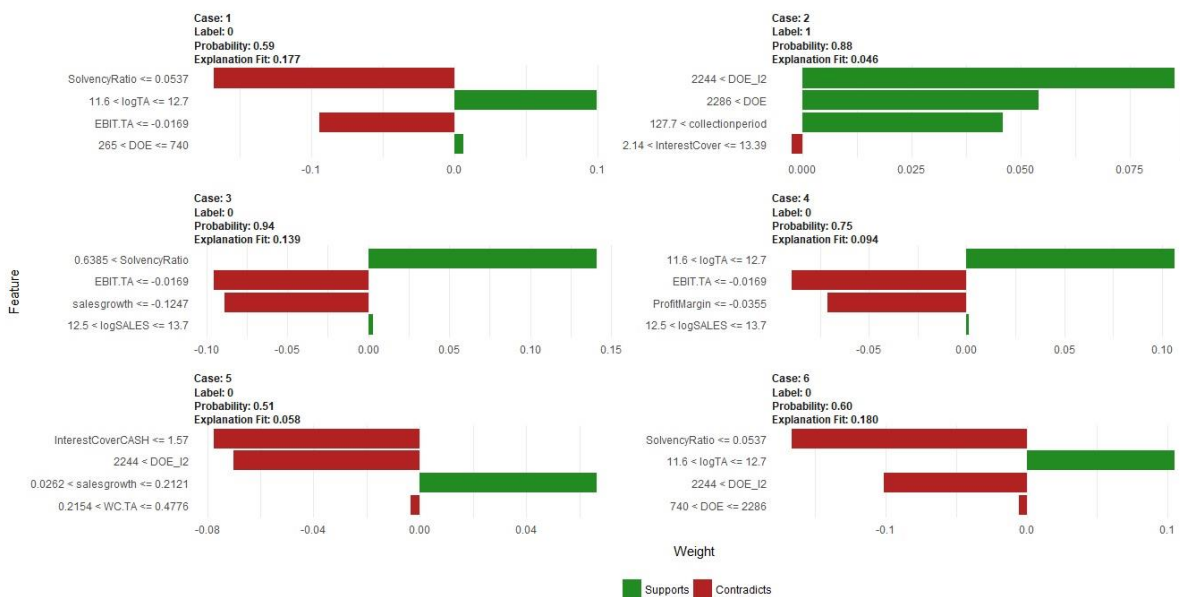
< xgboost 모형 >



c. lime 알고리즘을 통한 모델 설명력

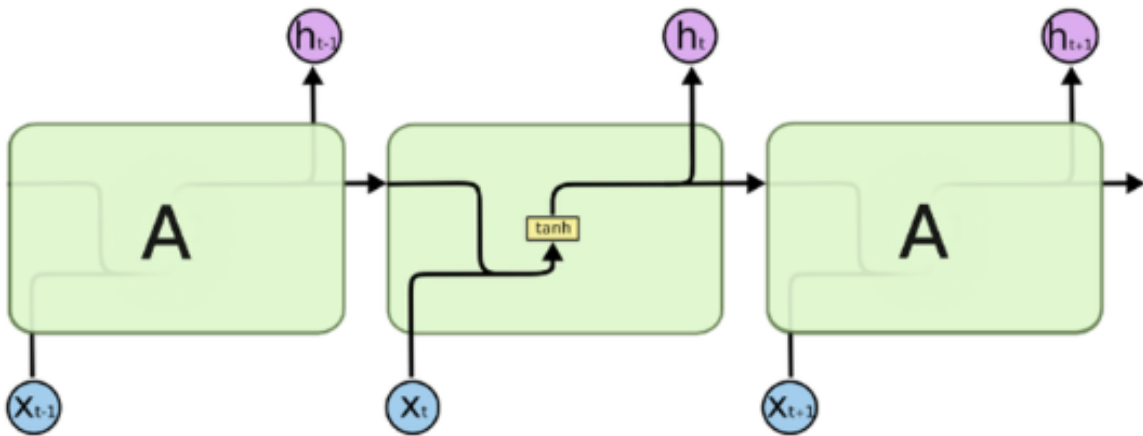
머신 러닝을 적용 할 때 직면하는 문제 중 하나는 복잡한 모델일 수록 해당 예측값에 대한 해석이 어렵다는 점이다. 가령, 은행 고객이 대출 불가 고객으로 분류되었을 때 고객에게 이유를 설명할 수 없다면 신뢰를 잃게 될 것이다. Explanation을 모델 복잡도는 낮게 유지하면서 loss를 최소화하는 문제로 정의한다.

< lime 알고리즘 예시 >



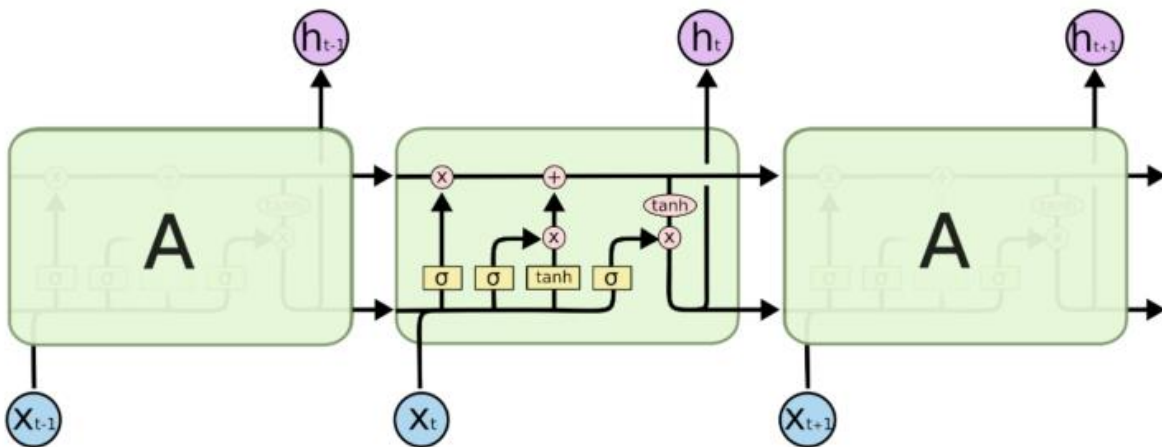
d. 시계열 자료이기 때문에 순환신경망 RNN 구현

< RNN 모형 >



e. 장기기억을 잃지 않기 위해 LSTM 구현

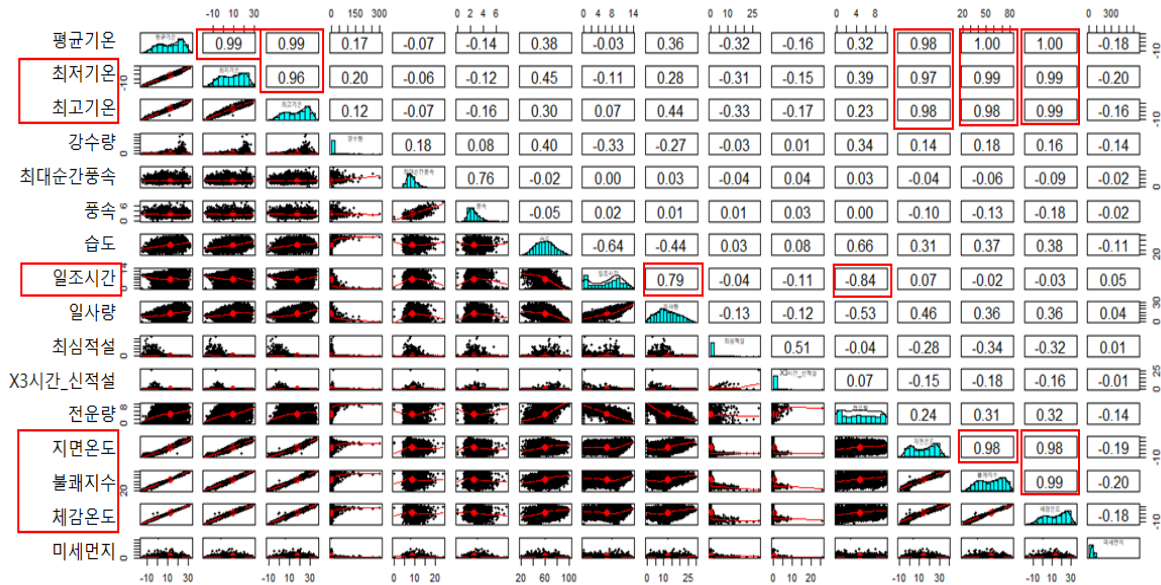
< LSTM 모형 >



## 4. 분석 결과

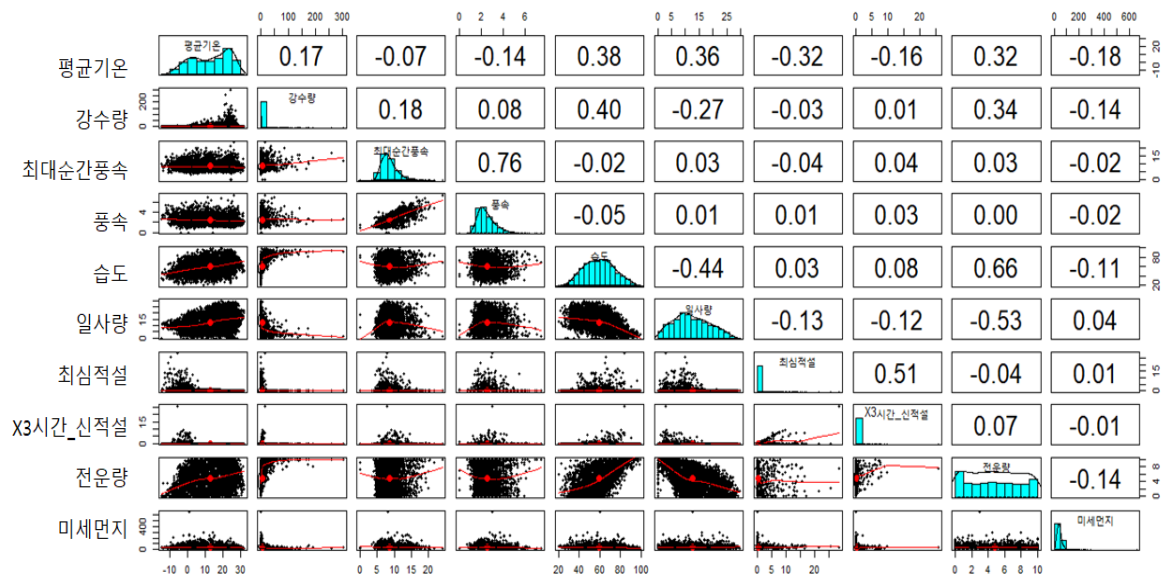
### ㄱ. 상관계수 행렬을 통한 변수 선택

#### 1. 다중공선성



- 다중 공선성을 고려하여 최저기온, 최고기온, 일조시간, 지면온도, 불쾌지수, 체감온도 변수 제거

#### 2. 비슷한 성질



- 최대순간풍속과 풍속, X3시간\_신적설과 최심신적설은 비슷한 성질을 갖기 때문에 변수 제거 .

- 최종 날씨 변수 18개

(평균기온, 강수량, 최대순간풍속, 풍속, 습도, 일사량, 최심적설, 3시간\_신적설, 전운량, 미세먼지, 요일, 호우, 강풍, 대설, 건조, 한파, 폭염, 황사)

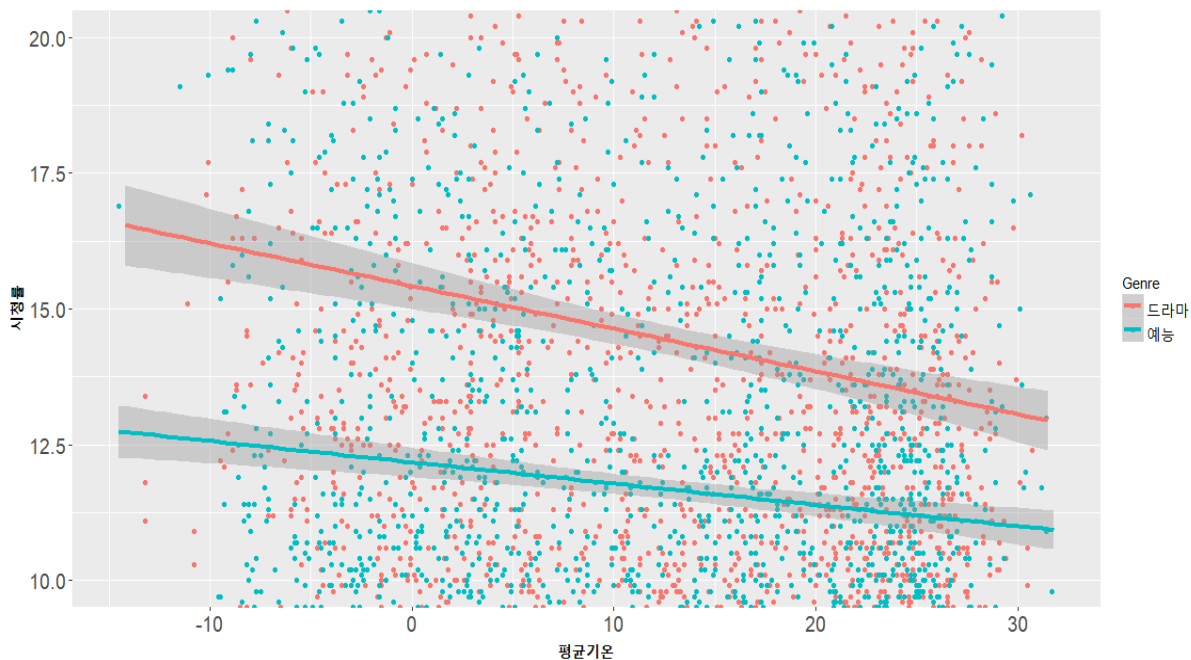
## ㄴ. 수치형 날씨변수 시각화

### 1. 예상

- a. 평균기온  
평균기온이 높거나 낮을수록 시청률이 높을 것으로 예상
- b. 강수량  
강수량이 높을수록 시청률이 높을 것으로 예상
- c. 풍속  
풍속이 셀수록 시청률이 높을 것으로 예상
- d. 습도  
습도가 높을수록 시청률이 높을 것으로 예상
- e. 일사량  
일사량이 많을수록 시청률이 높을 것으로 예상
- f. 최심적설  
최심적설이 클수록 시청률이 높을 것으로 예상
- g. 전운량  
전운량이 많을수록 시청률이 높을 것으로 예상
- h. 미세먼지  
미세먼지 농도가 높을수록 시청률이 높을 것으로 예상

### 2. 날씨별 드라마, 예능의 시청률

#### a. 평균기온 - 드라마, 예능

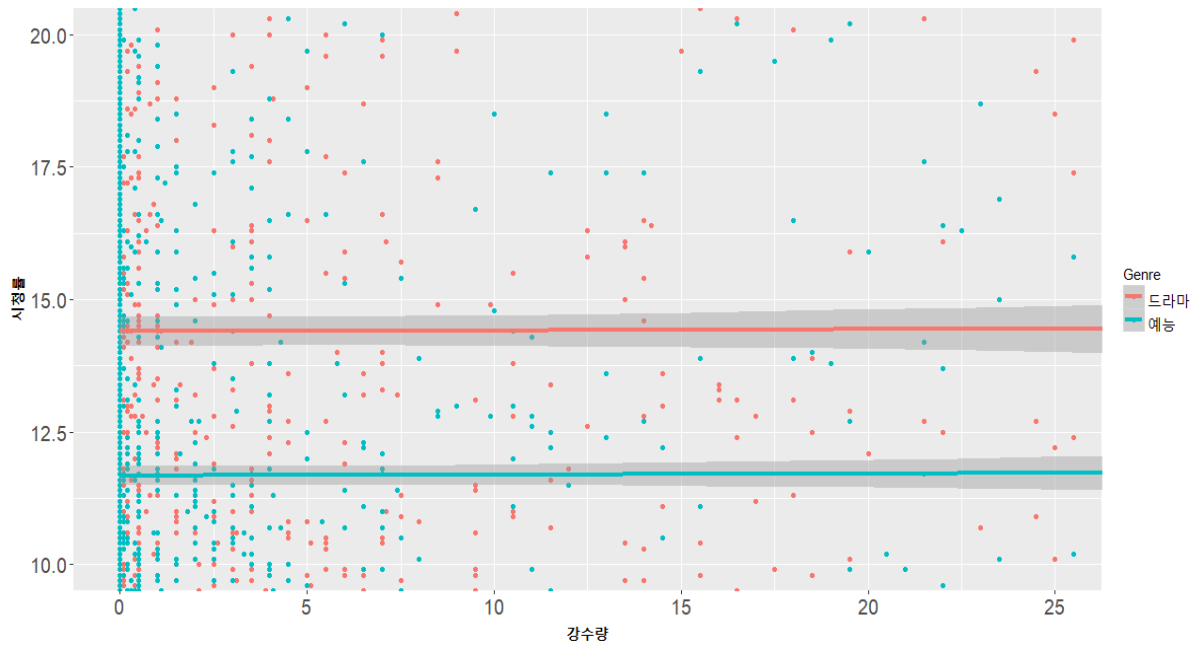


=> 평균기온이 상승할수록 드라마와 예능 모두 시청률이 떨어진다.

=> 드라마는 평균기온이 1단위 증가할 때, 시청률이 -0.2312 떨어지는 추세를 보이고,

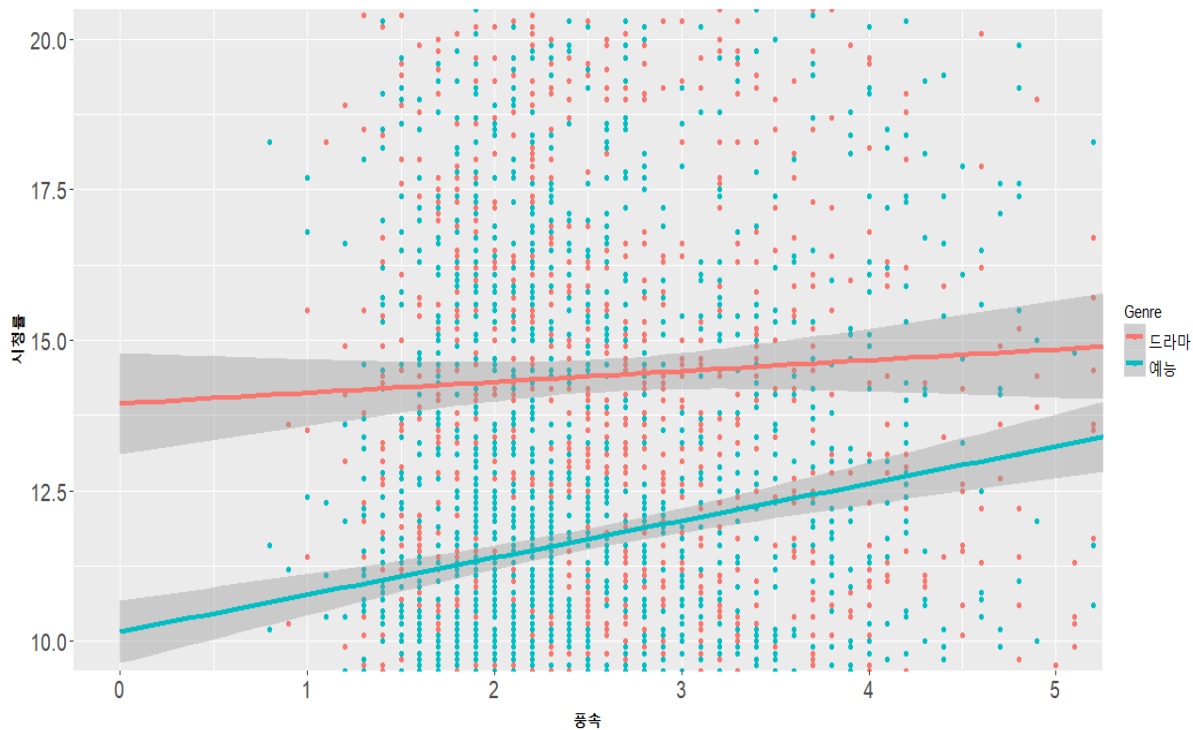
예능은 평균기온이 1단위 증가할 때, 시청률이 -0.2769 떨어지는 추세를 보인다.

b. 강수량 - 드라마, 예능



=> 드라마와 예능 모두 강수량의 변화에 따른 시청률의 차이가 미미하다.

c. 풍속 - 드라마, 예능

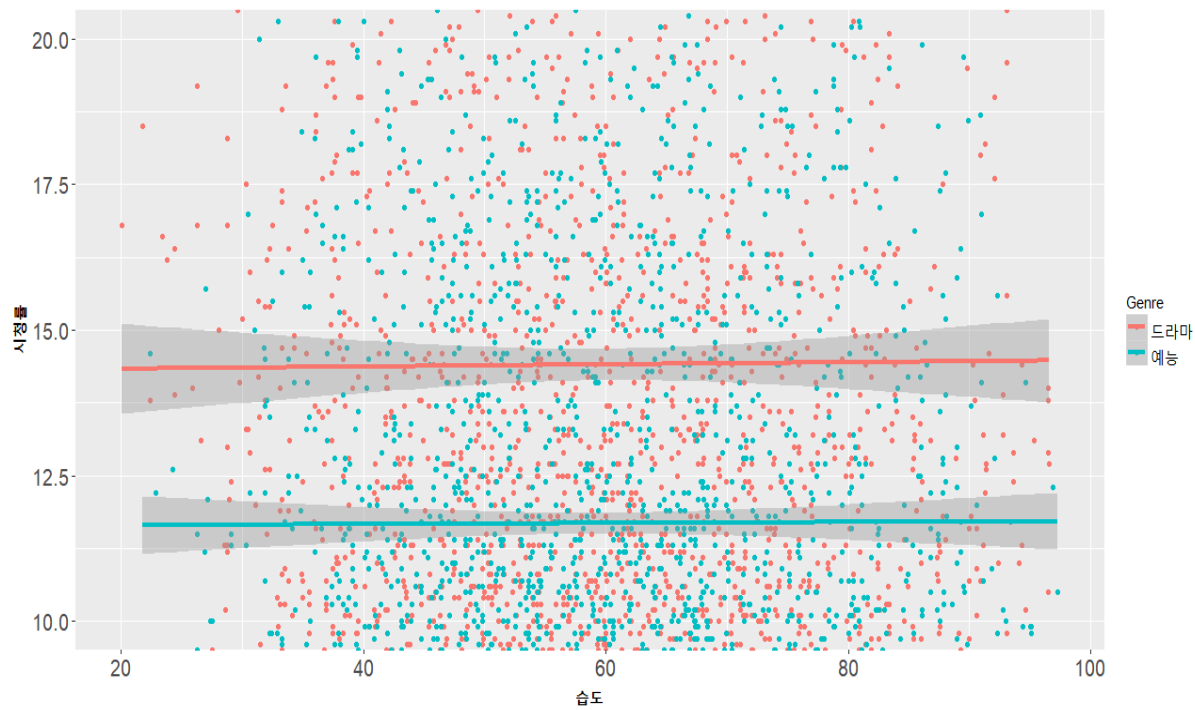


=> 예능에 비해 드라마 시청률의 변화량이 작지만, 풍속이 세질수록 드라마와 예능 모두 시청률이 올라간다.

=> 드라마는 풍속이 1단위 증가할 때, 시청률이 0.0036 올라가는 추세를 보이고,

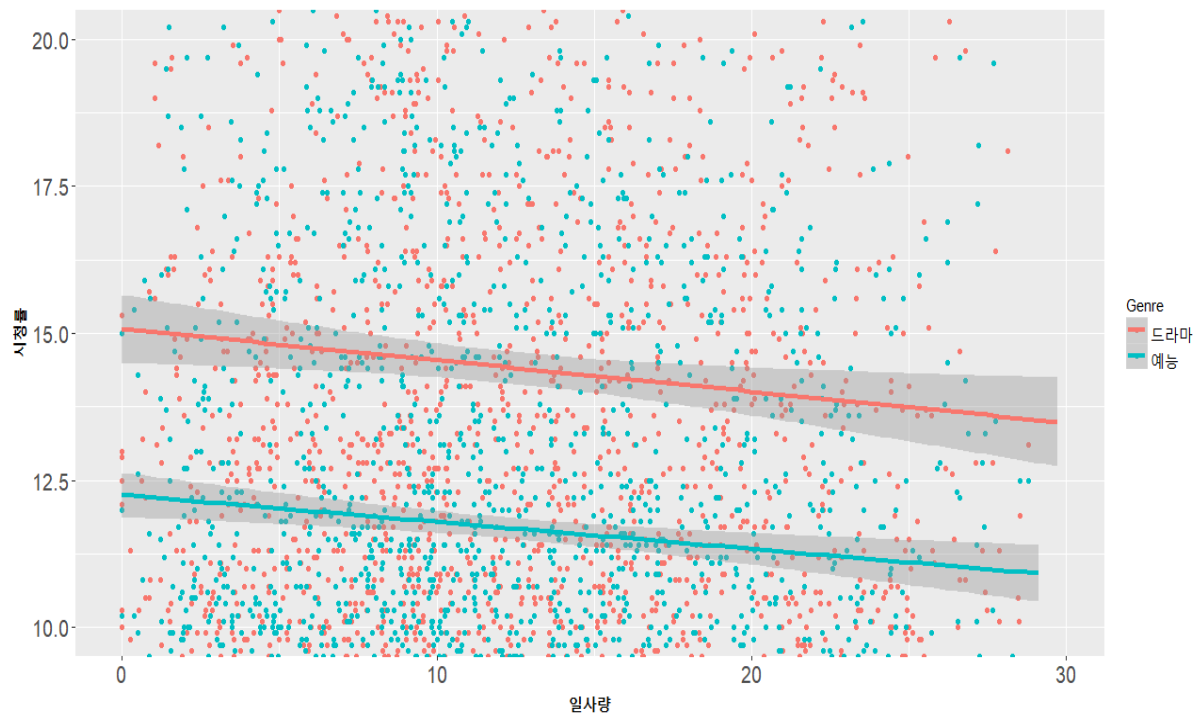
예능은 풍속이 1단위 증가할 때, 시청률이 0.0293 올라가는 추세를 보인다.

d. 습도 - 드라마, 예능



=> 습도의 변화에 따른 드라마와 예능 시청률의 차이가 미미하다.

e. 일사량 - 드라마, 예능



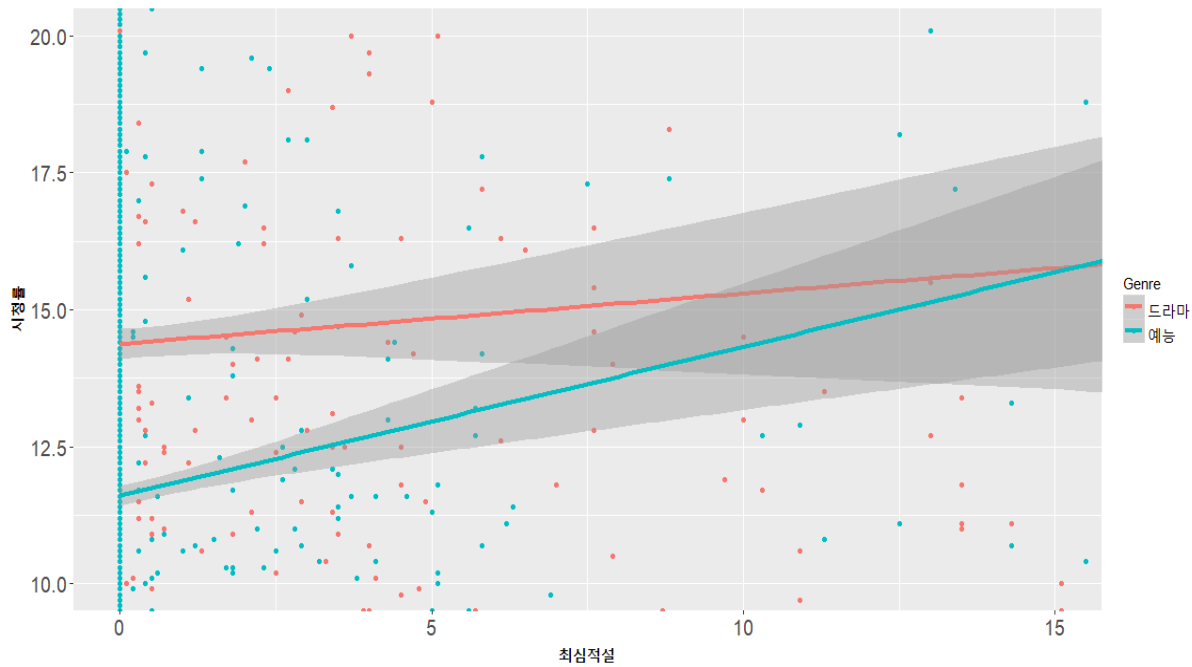
=> 일사량이 상승할수록 드라마와 예능 모두 시청률이 떨어진다.

=> 드라마는 일사량이 1단위 증가할 때, 시청률이 -0.0606 떨어지는 추세를 보이고,

예능은 일사량이 1단위 증가할 때, 시청률이 -0.1218 떨어지는 추세를 보인다.



f. 최심적설 - 드라마, 예능

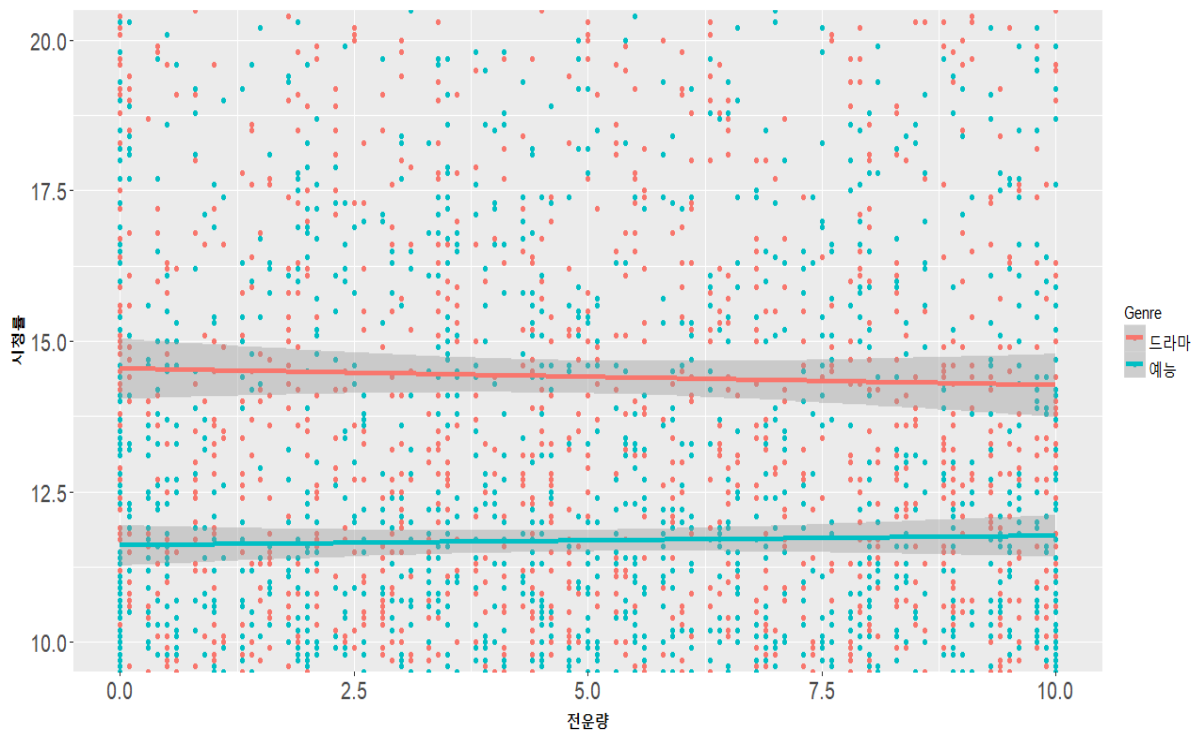


=> 예능에 비해 드라마 시청률의 변화량이 작지만, 최심적설이 상승할수록 드라마와 예능 모두 시청률이 올라간다.

=> 드라마는 최심적설이 1단위 증가할 때, 시청률이 0.0079 올라가는 추세를 보이고,

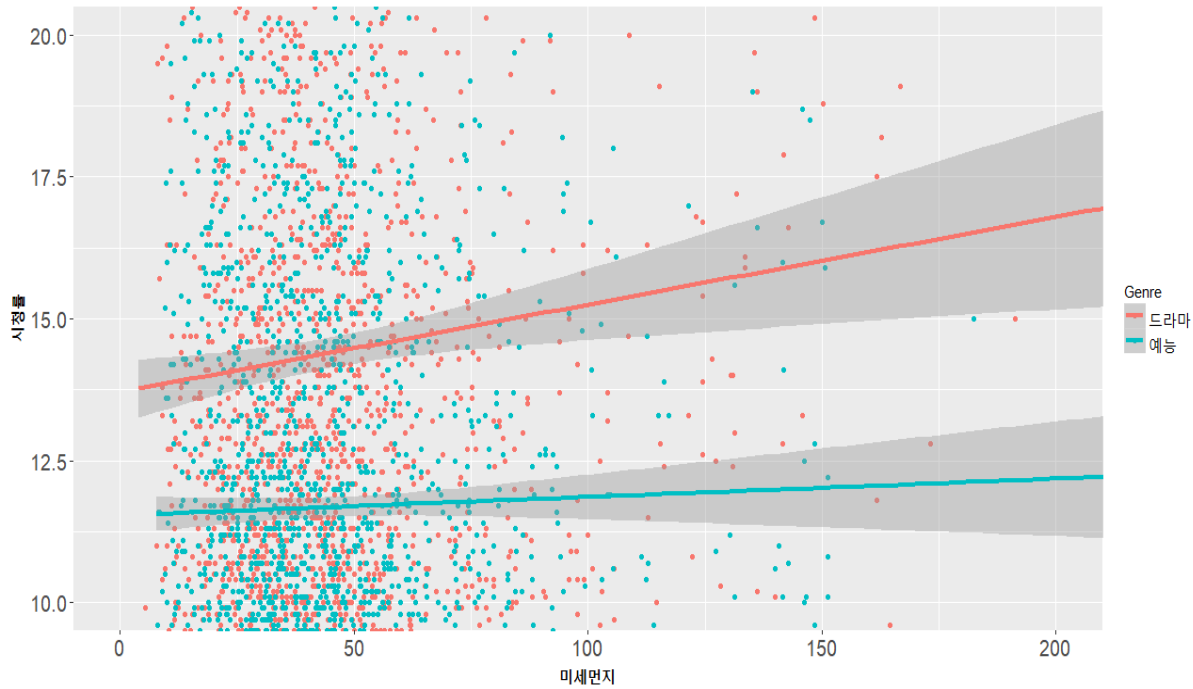
예능은 최심적설이 1단위 증가할 때, 시청률이 0.0374 올라가는 추세를 보인다.

g. 전문량 - 드라마, 예능



=> 전문량의 변화에 따른 드라마와 예능 시청률의 차이가 미미하다

#### h. 미세먼지 - 드라마, 예능

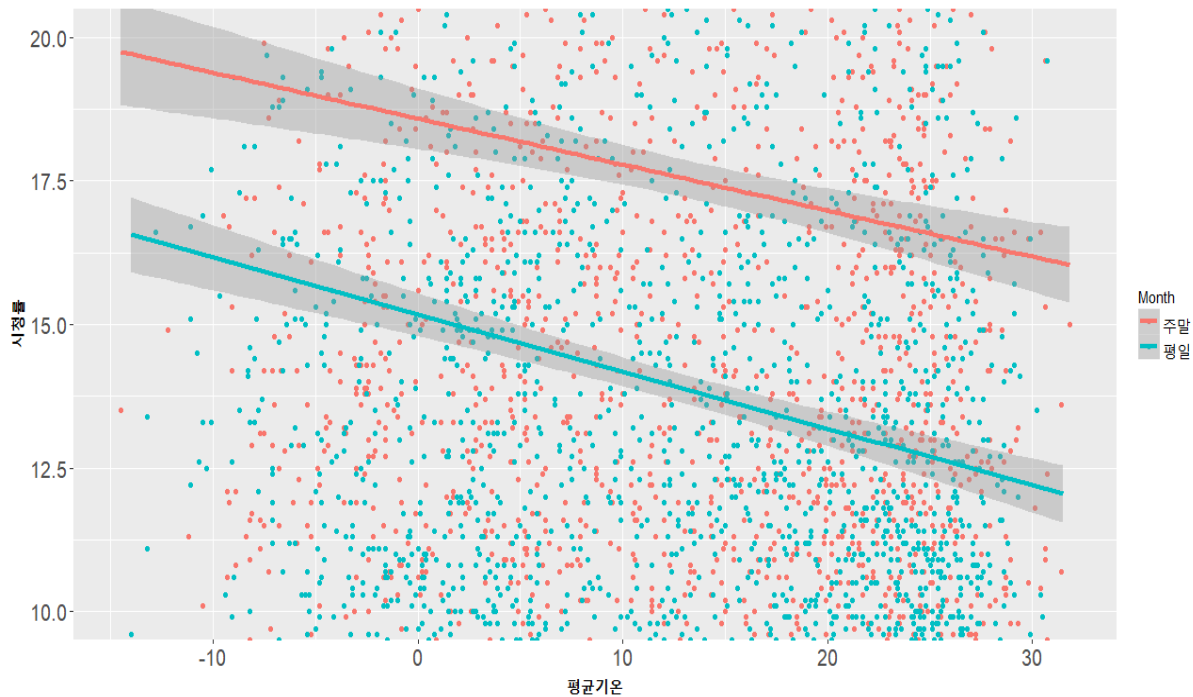


=> 드라마에 비해 예능 시청률의 변화량이 작지만, 미세먼지가 상승할수록 드라마와 예능 모두 시청률이 올라간다.

=> 드라마는 미세먼지가 1단위 증가할 때, 시청률이 0.2753 올라가는 추세를 보이고, 예능은 미세먼지가 1단위 증가할 때, 시청률이 0.1506 올라가는 추세를 보인다.

### 3. 날씨별 드라마의 평일, 주말 시청률

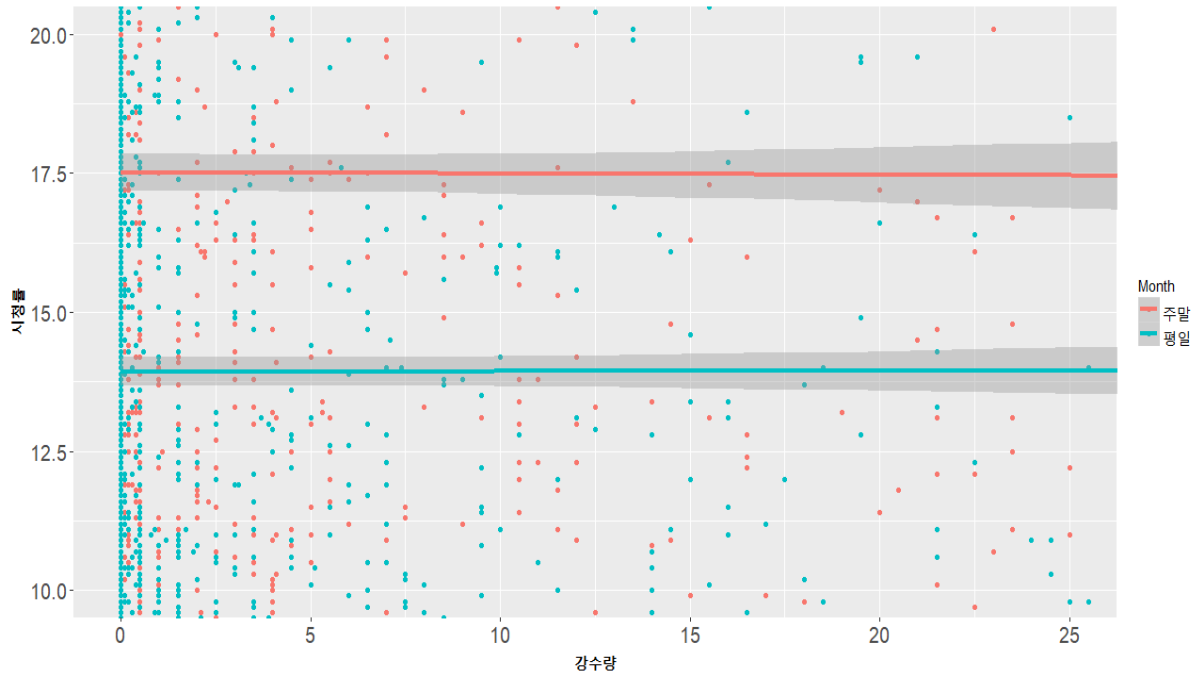
#### a. 평균기온 - 평일, 주말



=> 평균기온이 상승할수록 주말과 평일 드라마 모두 시청률이 떨어진다.

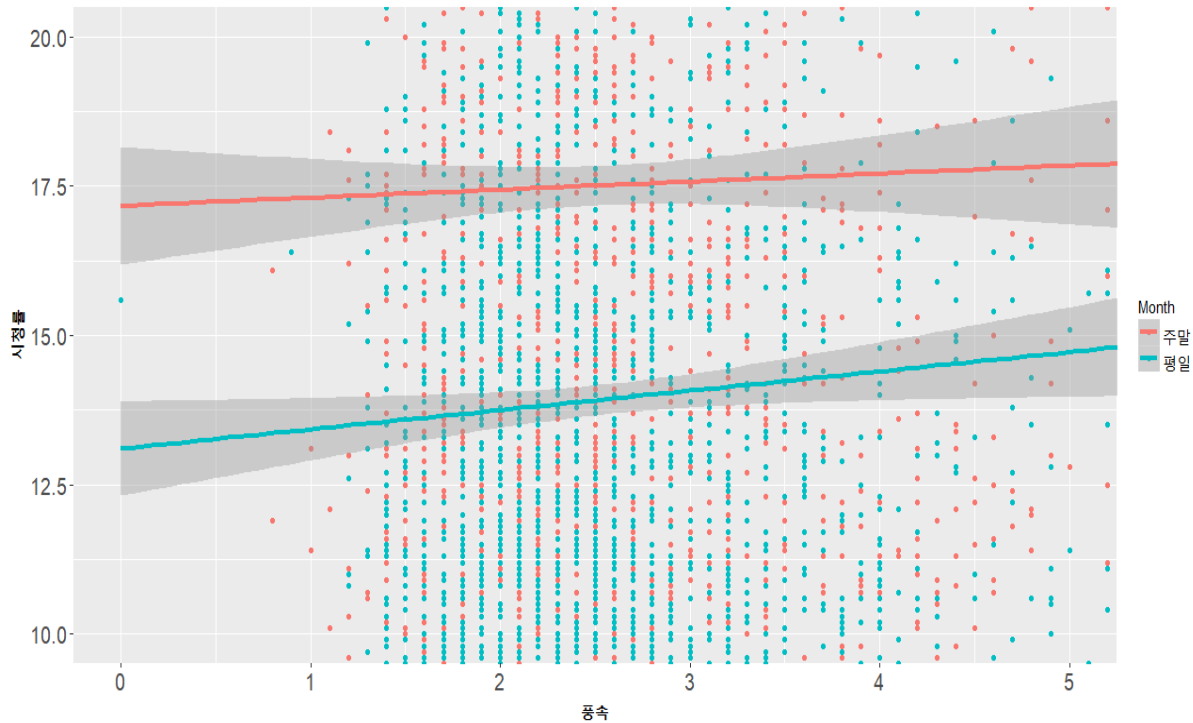
=> 주말 드라마는 평균기온이 1단위 증가할 때, 시청률이 -0.158 떨어지는 추세를 보이고, 평일 드라마는 평균기온이 1단위 증가할 때, 시청률이 -0.3568 떨어지는 추세를 보인다.

b. 강수량 - 평일, 주말



=> 주말 드라마와 평일 드라마 모두 강수량의 변화에 따른 시청률의 차이가 미미하다.

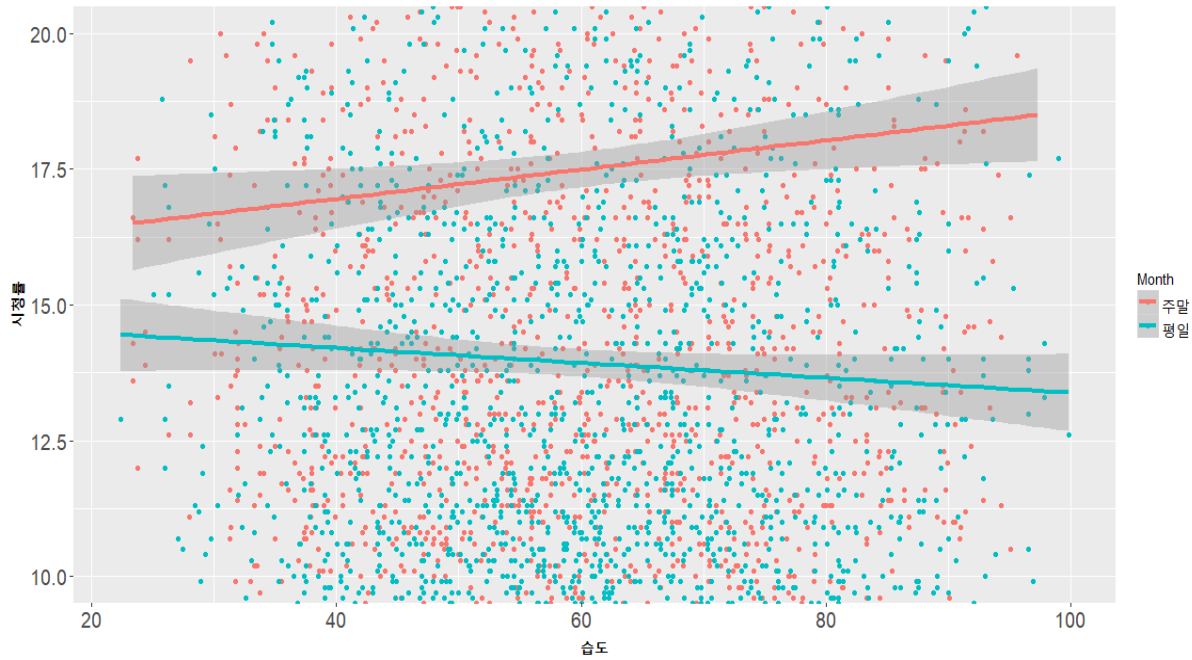
c. 풍속 - 평일, 주말



=> 평일 드라마에 비해 주말 드라마 시청률의 변화량이 작지만, 풍속이 세질수록 주말과 평일 드라마 모두 시청률이 올라간다.

=> 주말 드라마는 풍속이 1단위 증가할 때, 시청률이 0.0018 올라가는 추세를 보이고, 평일 드라마는 풍속이 1단위 증가할 때, 시청률이 0.0073 올라가는 추세를 보인다.

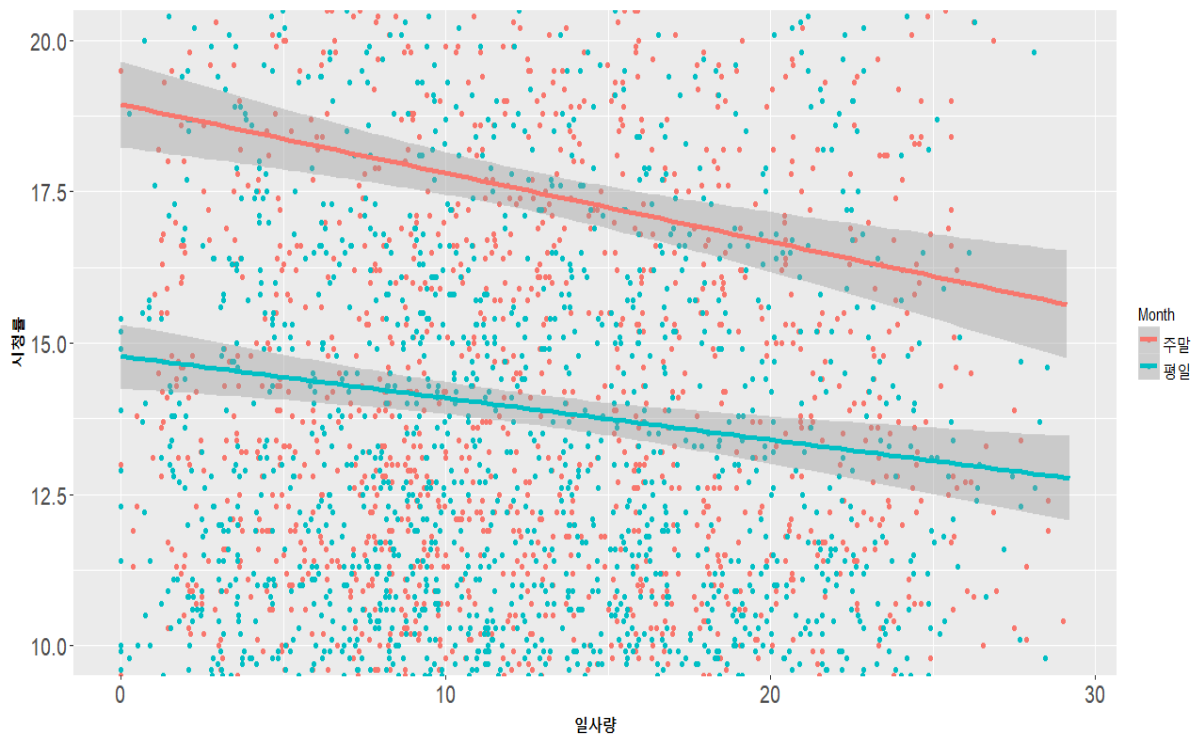
d. 습도 - 평일, 주말



=> 습도가 상승할수록 주말 드라마는 시청률이 올라가는 반면, 평일 드라마는 시청률이 떨어진다.

=> 주말 드라마는 습도가 1단위 증가할 때, 시청률이 0.1108 올라가는 추세를 보이고, 평일 드라마는 습도가 1단위 증가할 때, 시청률은 -0.0964 떨어지는 추세를 보인다.

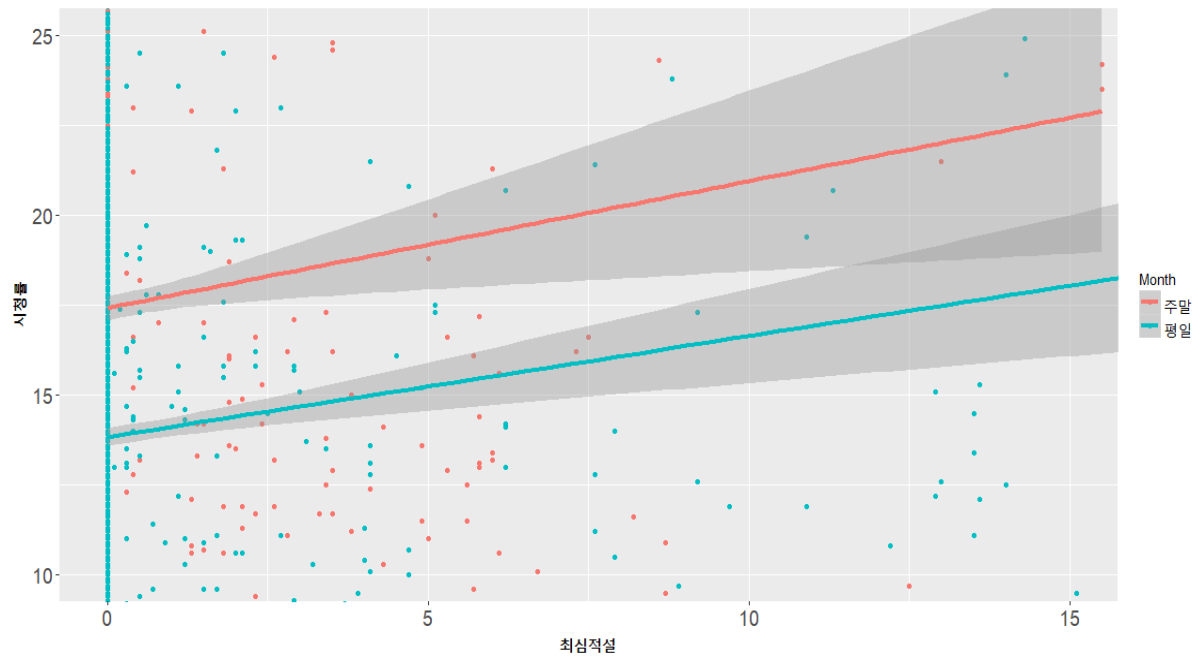
e. 일사량 - 평일, 주말



=> 일사량이 상승할수록 주말과 평일 드라마 모두 시청률이 떨어진다.

=> 주말 드라마는 일사량이 1단위 증가할 때, 시청률이 -0.0864 떨어지는 추세를 보이고, 평일 드라마는 일사량이 1단위 증가할 때, 시청률은 -0.0899 떨어지는 추세를 보인다.

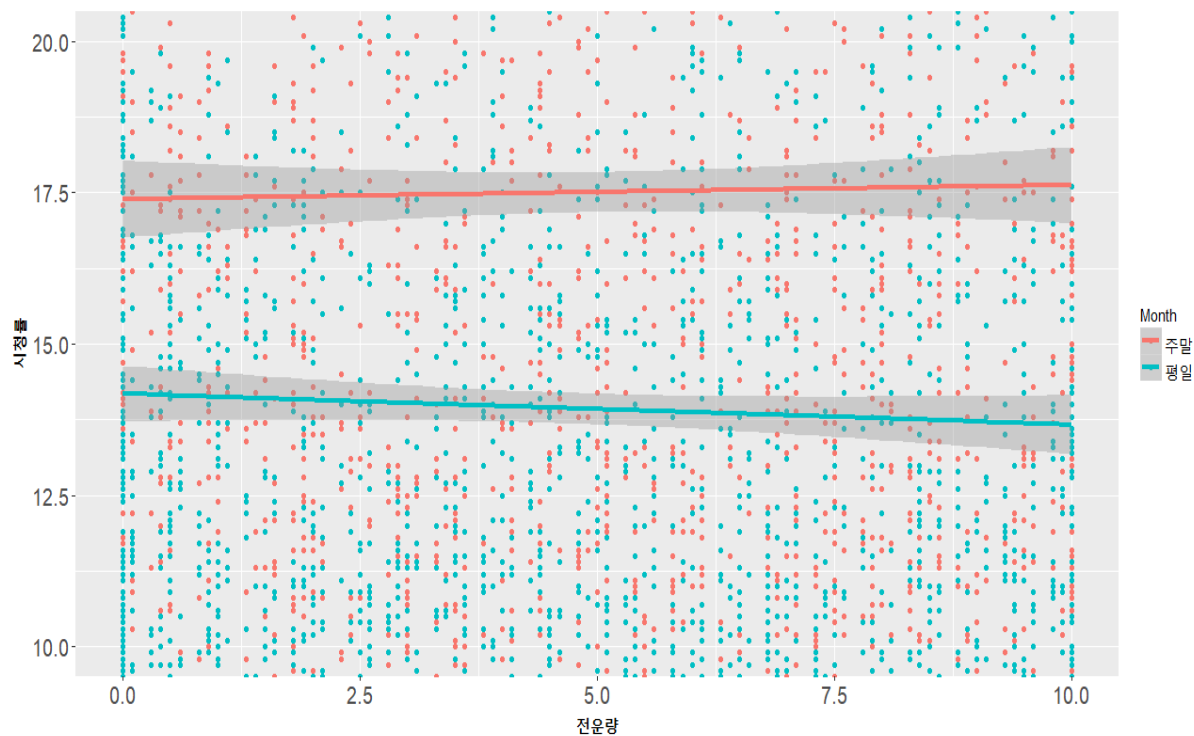
f. 최심적설 - 평일, 주말



=> 최심적설이 상승할수록 주말과 평일 드라마 모두 시청률이 올라간다.

=> 주말 드라마는 최심적설이 1단위 증가할 때, 시청률이 0.0104 올라가는 추세를 보이고,  
평일 드라마는 최심적설이 1단위 증가할 때, 시청률은 0.0303 올라가는 추세를 보인다.

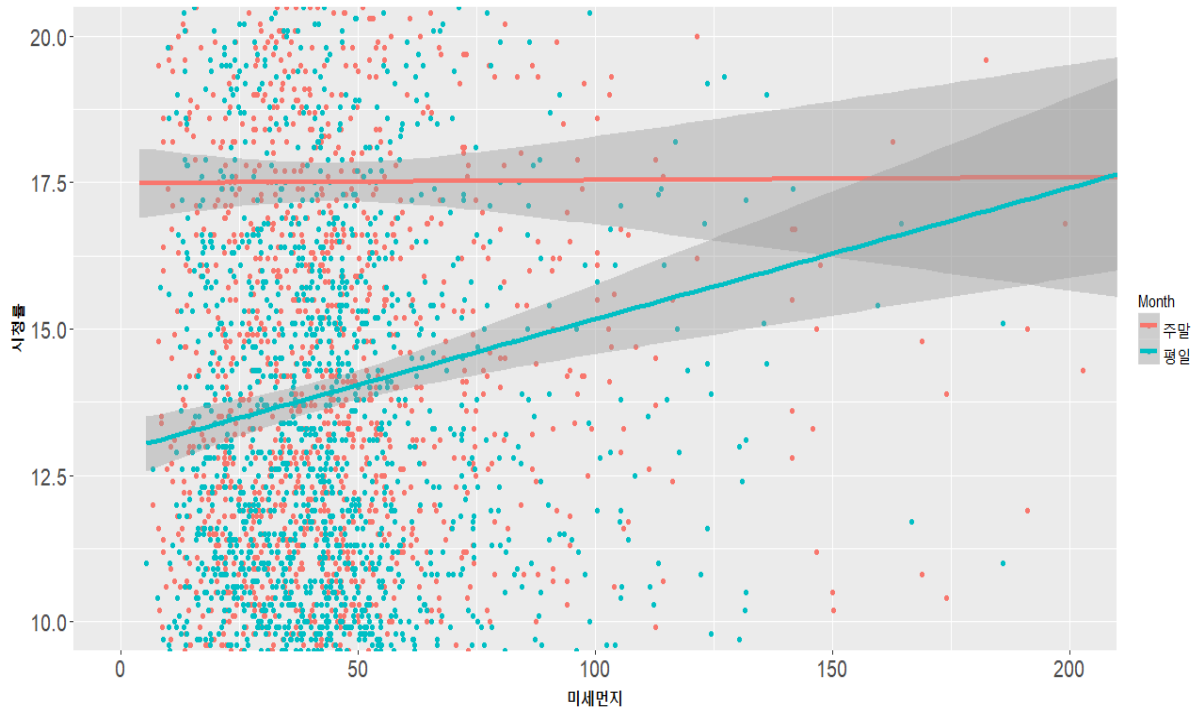
g. 전문량 - 평일, 주말



=> 전문량이 상승할수록 평일 드라마는 시청률이 떨어지는 반면, 주말 드라마의 시청률 차이는 미미하다.

=> 평일 드라마는 전문량이 1단위 증가할 때, 시청률은 -0.0154 떨어지는 추세를 보인다.

#### h. 미세먼지 - 평일, 주말

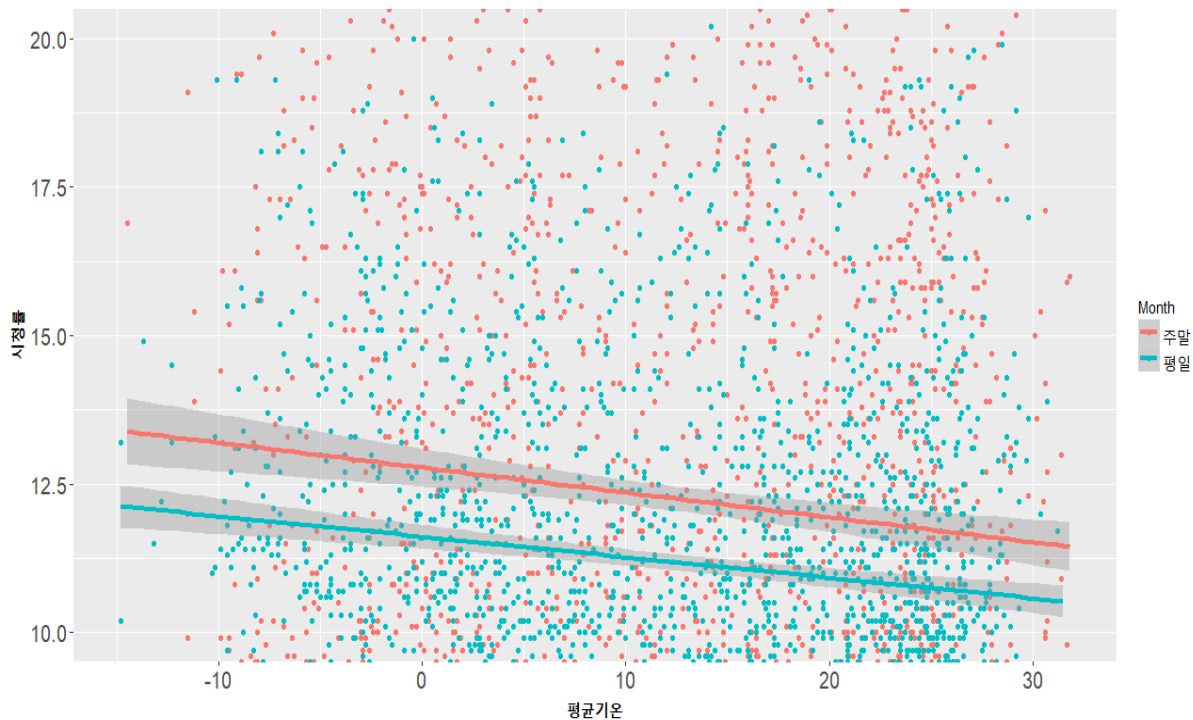


=> 미세먼지가 상승할수록 평일 드라마의 시청률은 올라가는 반면, 주말 드라마의 시청률 차이는 미미하다.

=> 평일 드라마는 미세먼지가 1단위 증가할 때, 시청률은 0.4423 올라가는 추세를 보인다.

#### 4. 날씨별 예능의 평일, 주말 시청률

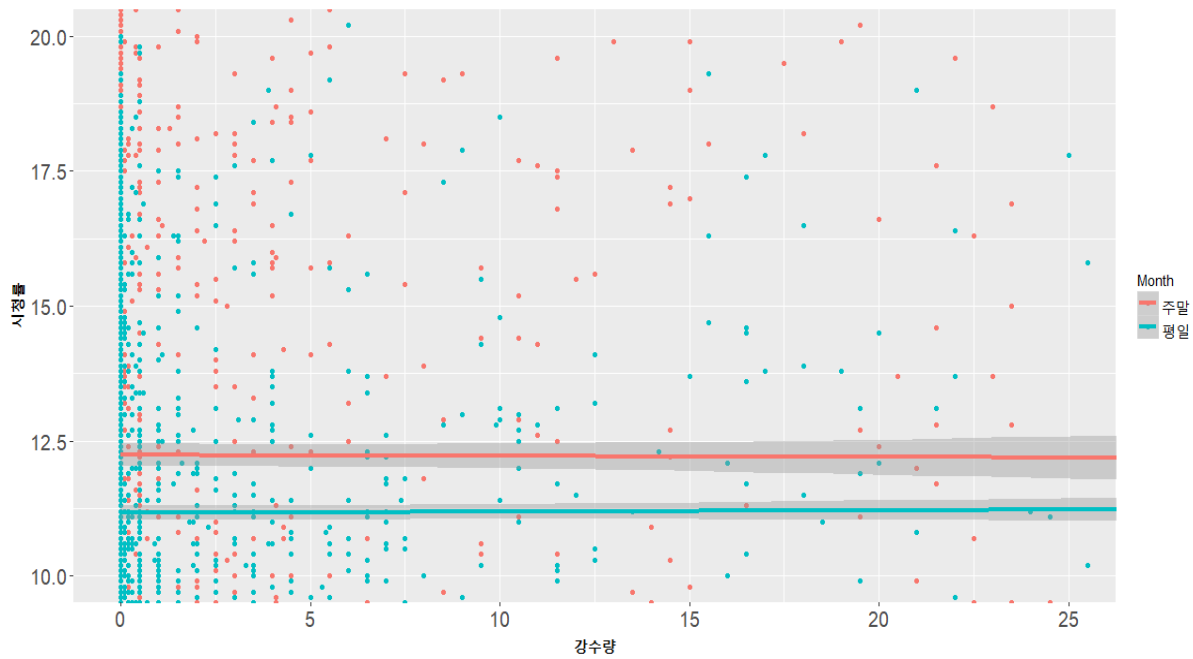
##### a. 평균기온 - 평일, 주말



=> 평균기온이 상승할수록 주말과 평일 예능 모두 시청률이 떨어진다.

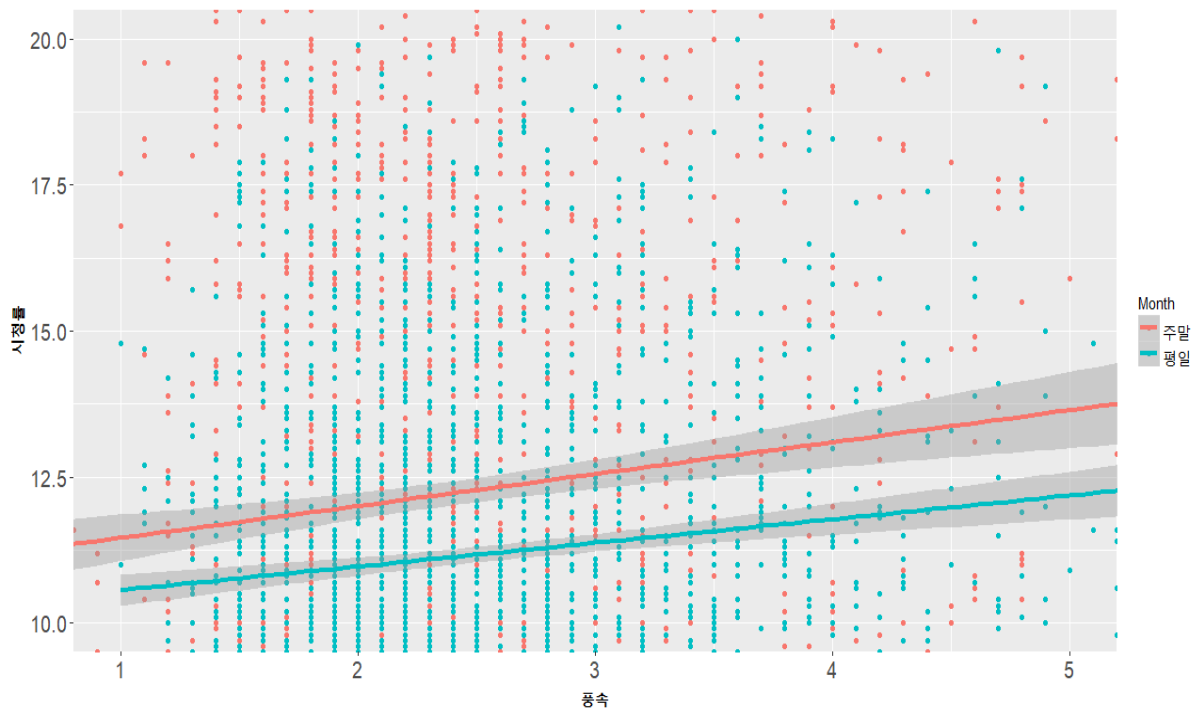
=> 주말 예능은 평균기온이 1단위 증가할 때, 시청률이 -0.2208 떨어지는 추세를 보이고, 평일 예능은 평균기온이 1단위 증가할 때, 시청률이 -0.4452 떨어지는 추세를 보인다.

b. 강수량 - 평일, 주말



=> 주말과 평일 예능 모두 강수량의 변화에 따른 시청률의 차이가 미미하다.

c. 풍속 - 평일, 주말

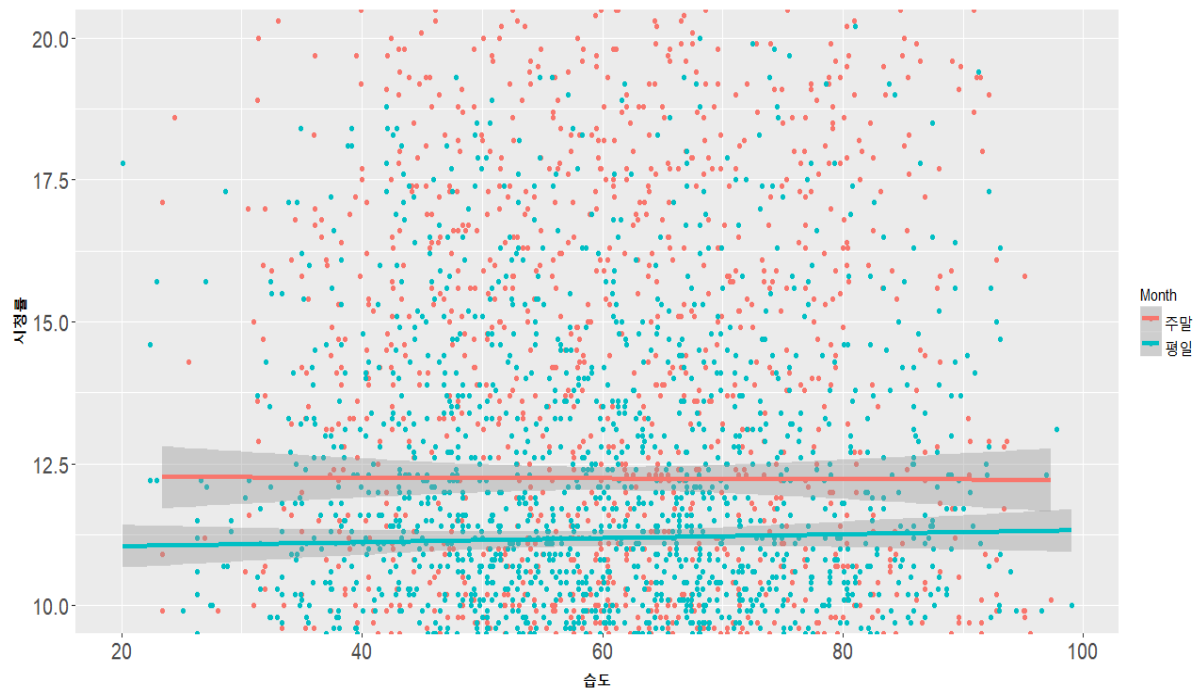


=> 풍속이 세질수록 주말과 평일 예능 모두 시청률이 올라간다.

=> 주말 예능은 풍속이 1단위 증가할 때, 시청률이 0.018 올라가는 추세를 보이고,

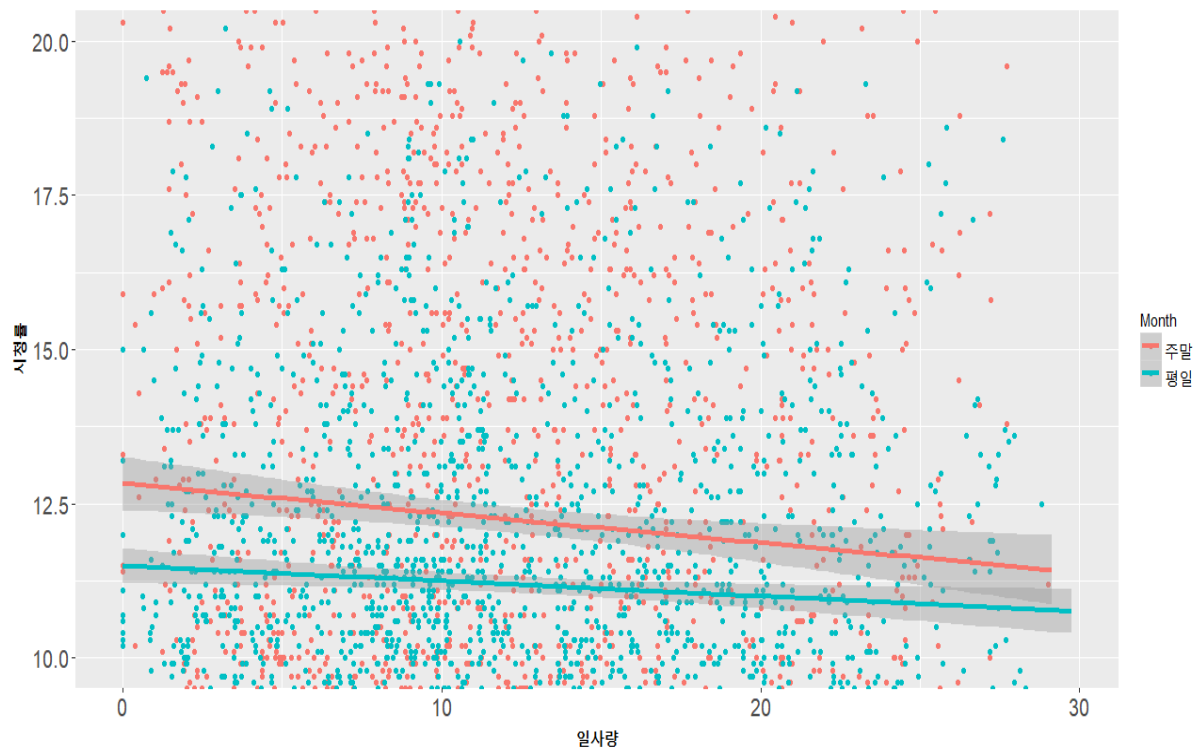
평일 예능은 풍속이 1단위 증가할 때, 시청률이 0.0314 올라가는 추세를 보인다.

d. 습도 - 평일, 주말



=> 주말과 평일 예능 모두 습도의 변화에 따른 시청률의 차이가 미미하다.

e. 일사량 - 평일, 주말

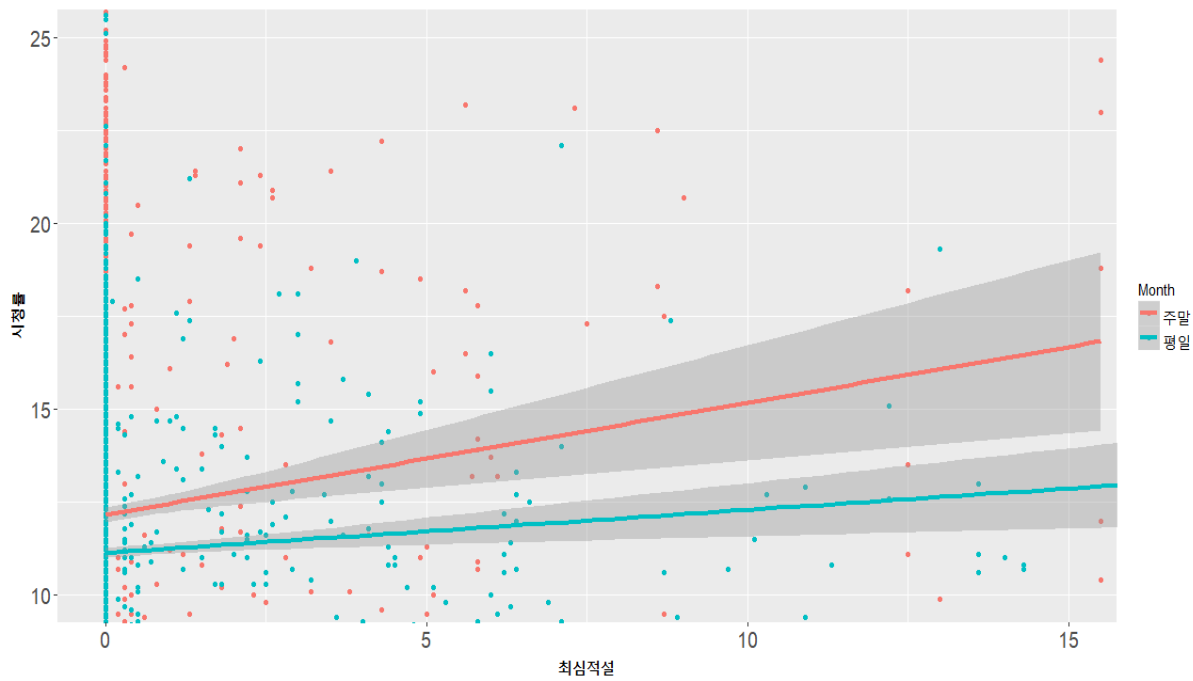


=> 일사량이 상승할수록 주말과 평일 예능 모두 시청률이 떨어진다.

=> 주말 예능은 일사량이 1단위 증가할 때, 시청률이 -0.0957 떨어지는 추세를 보이고,  
평일 예능은 일사량이 1단위 증가할 때, 시청률이 -0.1235 떨어지는 추세를 보인다.



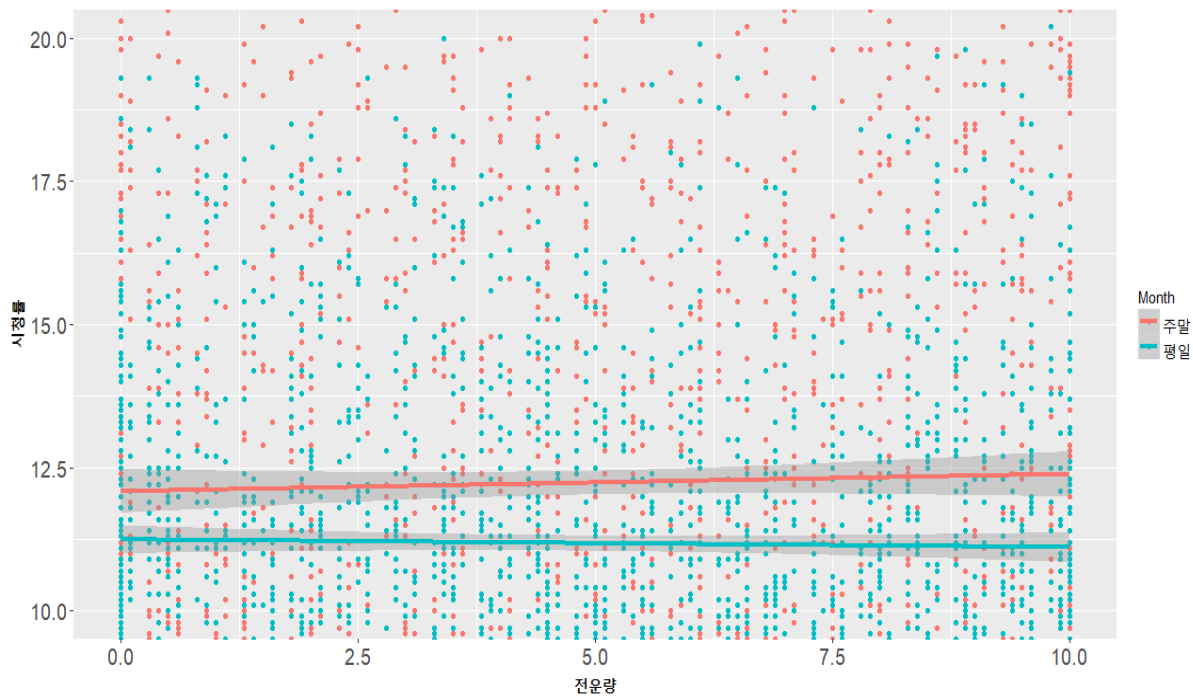
f. 최심적설 - 평일, 주말



=> 최심적설이 상승할수록 주말과 평일 예능 모두 시청률이 올라간다.

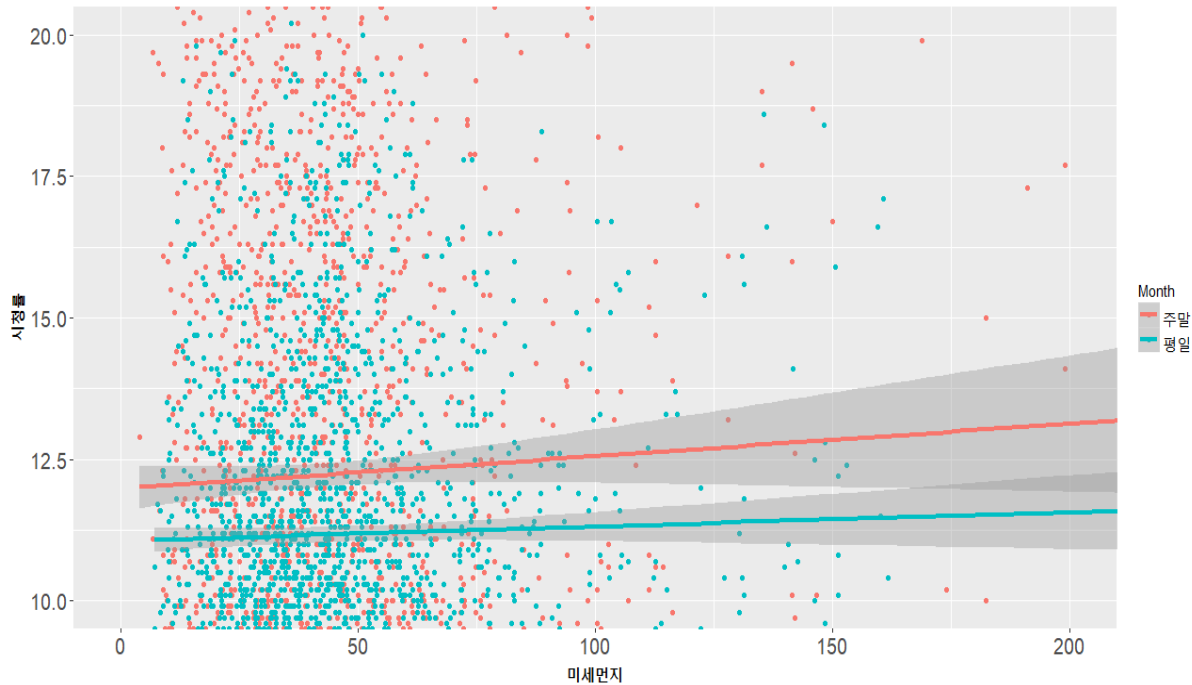
=> 주말 예능은 최심적설이 1단위 증가할 때, 시청률이 0.0233 올라가는 추세를 보이고,  
평일 예능은 최심적설이 1단위 증가할 때, 시청률이 0.0412 올라가는 추세를 보인다.

g. 전운량 - 평일, 주말



=> 주말과 평일 예능 모두 전운량의 변화에 따른 시청률의 차이가 미미하다.

#### h. 미세먼지 - 평일, 주말



=> 미세먼지가 상승할수록 주말과 평일 예능 모두 시청률이 올라간다.

=> 주말 예능은 미세먼지가 1단위 증가할 때, 시청률이 0.1911 올라가는 추세를 보이고,  
평일 예능은 미세먼지가 1단위 증가할 때, 시청률이 0.2814 올라가는 추세이다.

### 5. 예측 결과

#### a. 평균기온

예상과 달리, 평균기온이 증가할수록 시청률은 떨어졌다. 이는 기온이 높아짐에 따라 한낮 더위를 시원한 에어컨이 나오는 마트나 커피숍, 쇼핑몰 등으로 피하는 '물캉족'과, 늦은 저녁 운동을 시작하는 '올빼미족'의 영향이라고 볼 수 있다. 에어컨이 고장 났거나 멀쩡하더라도 냉방비를 아끼기 위해 냉방이 잘 된 공공장소를 찾아가기 때문이다. 이에 따라 평균기온이 증가할수록 실외로 나가지 않아 시청률이 높을 것이라는 예상과 달리, 오히려 집이 아닌 냉방이 잘 된 실내로 이동하기 때문에 시청률이 떨어지는 결과를 가져왔다.

출처: <http://news.donga.com/3/all/20180718/91115375/1>

<http://www.munhwa.com/news/view.html?no=2018072001070227109001>

#### b. 강수량

비가 오면 야외활동을 꺼려하는 경향이 있기 때문에 야외활동보다는 집에서 TV를 시청할 것이라는 예상과 달리, 강수량은 시청률에 큰 영향을 미치지 않는 것으로 나타났다.

#### c. 풍속

바람이 강하게 불수록 추위를 피하기 위해 집에서 TV시청을 할 것이라는 예상과 동일하게, 풍속이 셀수록 시청률이 올라가는 결과를 가져왔다.

d. 습도

습하면 야외활동을 꺼려하는 경향이 있기 때문에 야외활동보다는 집에서 TV를 시청할 것이라는 예상과 달리, 습도는 시청률에 큰 영향을 미치지 않은 것으로 나타났다.

e. 일사량

일사량이 많을수록 더위를 피하기 위해 집에만 있을 것이라는 예상과 달리, 일사량이 많을수록 햇빛을 피하기 위해 냉방이 잘 된 공공장소를 찾아가는 사람들이 늘어났다. 이에 따라 평균기온과 마찬가지로, 시청률이 오히려 떨어지는 결과를 가져왔다.

f. 최심적설

눈이 많이 내려 최심적설이 높아지면 야외활동에 제한이 생기기 때문에 집에서 TV를 시청할 것이라는 예상과 동일하게, 최심적설이 높을수록 시청률이 올라가는 결과를 가져왔다.

g. 전운량

구름의 양이 많으면 비가 내리거나 기상악화를 예상하게 되어 야외활동을 자제하려는 경향이 있기 때문에 시청률이 높아질 것이라 예상했지만, 이와 달리 전운량은 시청률에 큰 영향을 미치지 않는 것으로 나타났다.

h. 미세먼지

미세먼지 농도가 높아지면서 다양한 질병과 높은 조기사망률을 보이고 있다. 미세먼지는 사회적 문제에 있어 이미 큰 영향을 미치고 있으며, 혈관 질환은 물론, 온열질환자의 급상승으로 인해 야외활동을 자제하는 것이 당연히 되고 있다. 이러한 미세먼지 농도의 위험성에 따라 야외활동을 자제하게 되고 TV시청률이 올라갈 것이라 예상했으며, 이와 동일하게 미세먼지 농도가 좋음에서 매우 나쁨으로 갈수록 시청률이 올라가는 결과를 가져왔다.

출처: [http://health.chosun.com/site/data/html\\_dir/2018/06/29/2018062902026.html](http://health.chosun.com/site/data/html_dir/2018/06/29/2018062902026.html)

## ㄷ. 범주형 날씨 변수 시각화

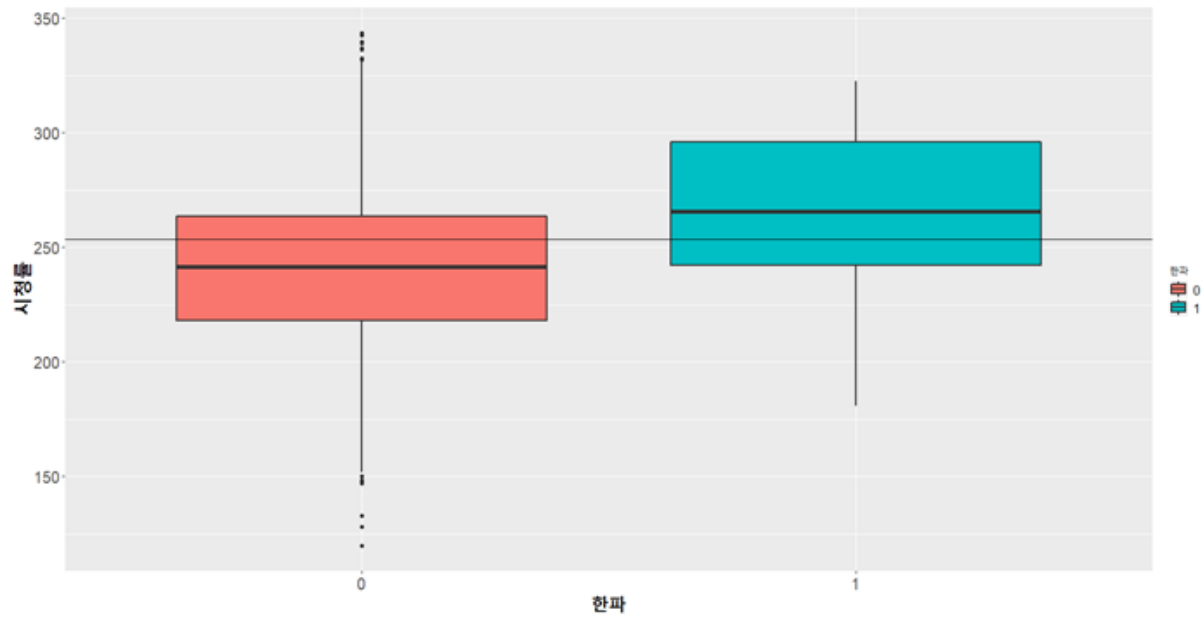
(한파, 대설, 강풍, 건조, 호우, 폭염 특보 발령인 경우 1, 아닌 경우 0 )

### 1. 예상

- 특보 발령 시 시청률이 높을 것으로 예상

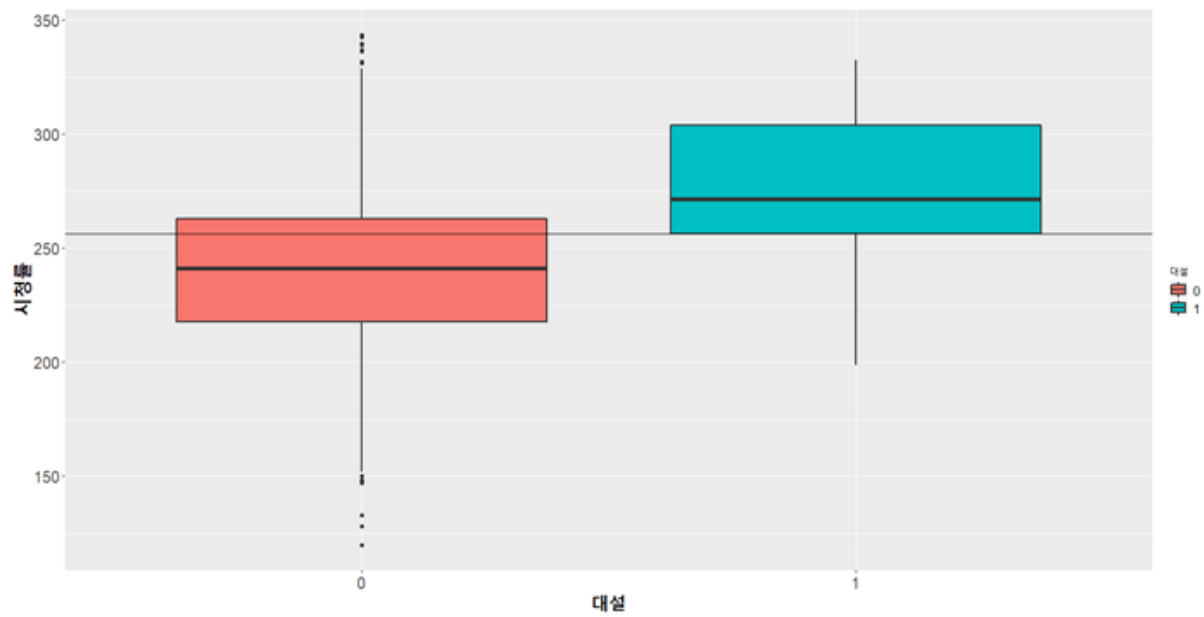
### 2. 특보 발령 여부와 시청률간의 관계

a. 한파



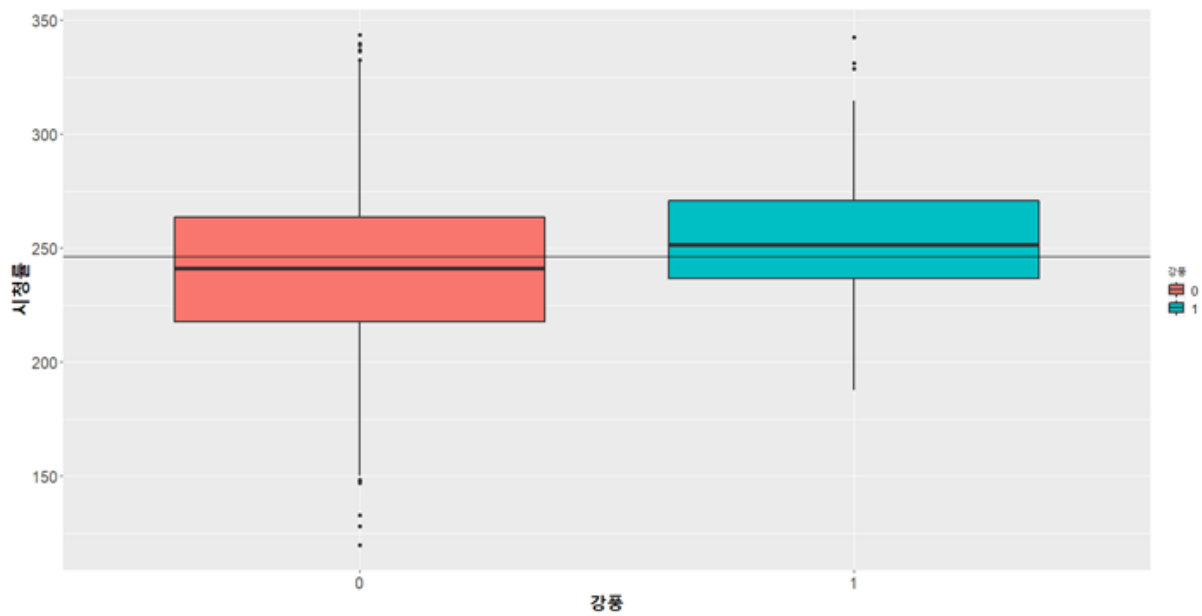
=> 한파가 아닌 날에 비해 한파인 날, 시청률이 22.2949 더 높다.

#### b. 대설



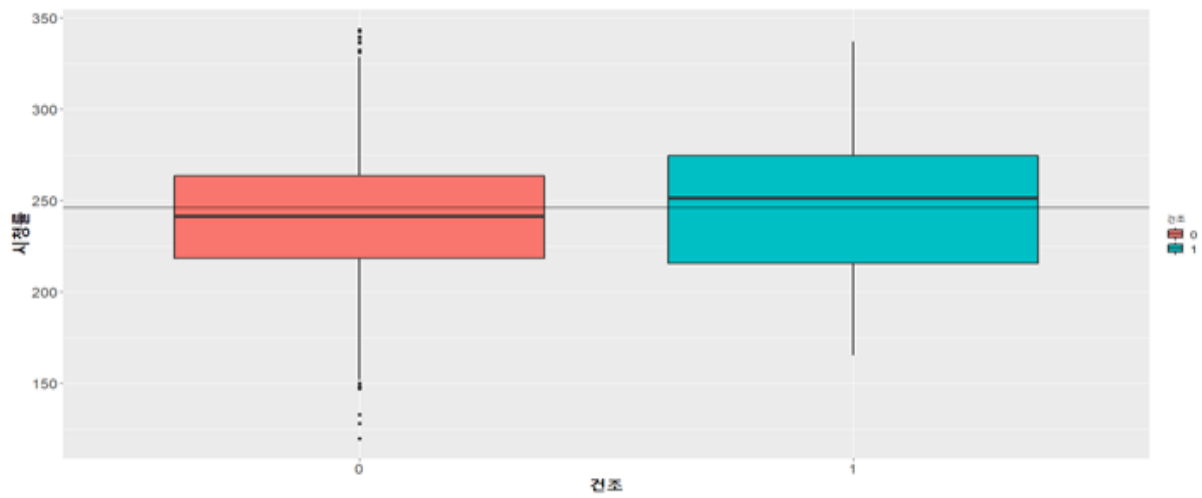
=> 대설이 아닌 날에 비해 대설인 날, 시청률이 35.0841 더 높다.

c. 강풍



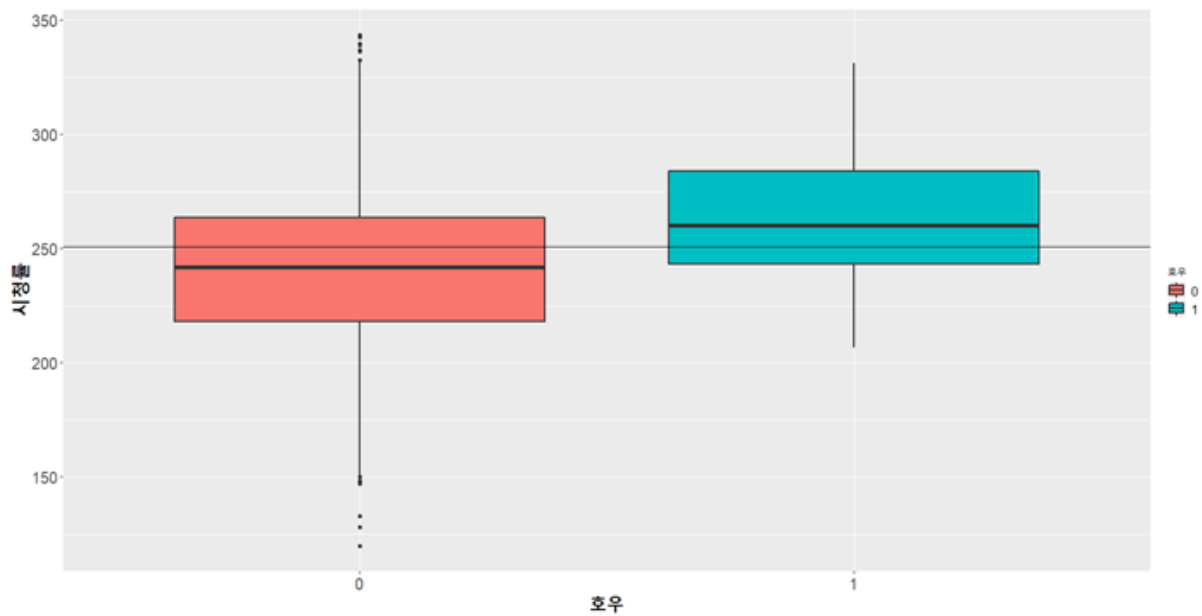
=> 강풍이 아닌 날에 비해 강풍인 날, 시정률이 13.2246 더 높다

d. 건조

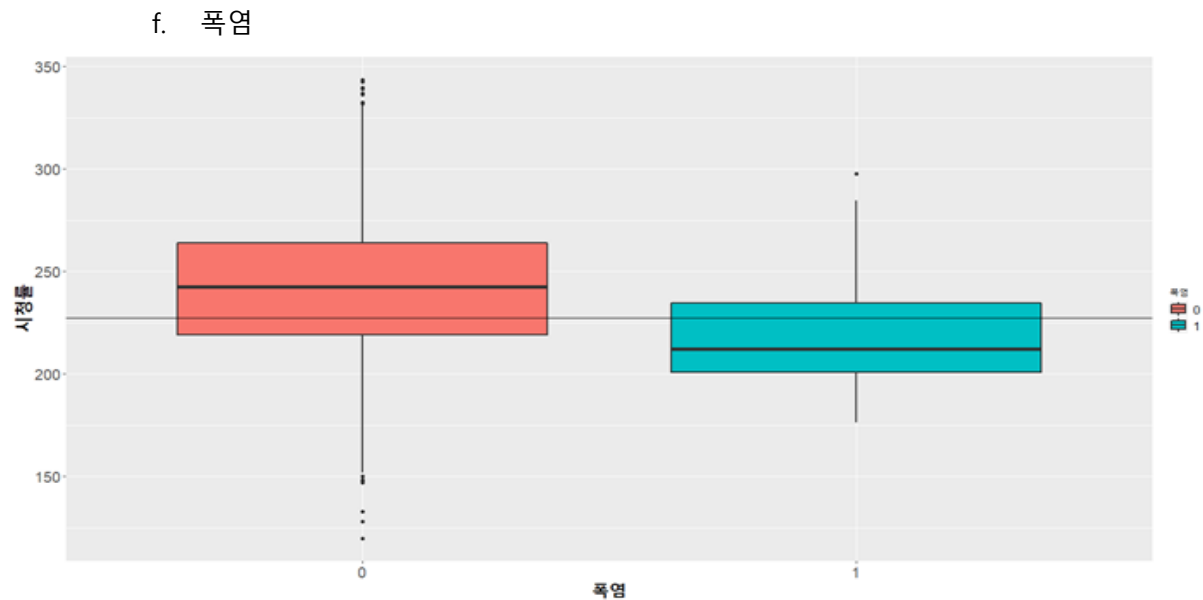


=> 건조가 아닌 날에 비해 건조인 날, 시정률이 6.136 높지만 그 차이가 미미하다.

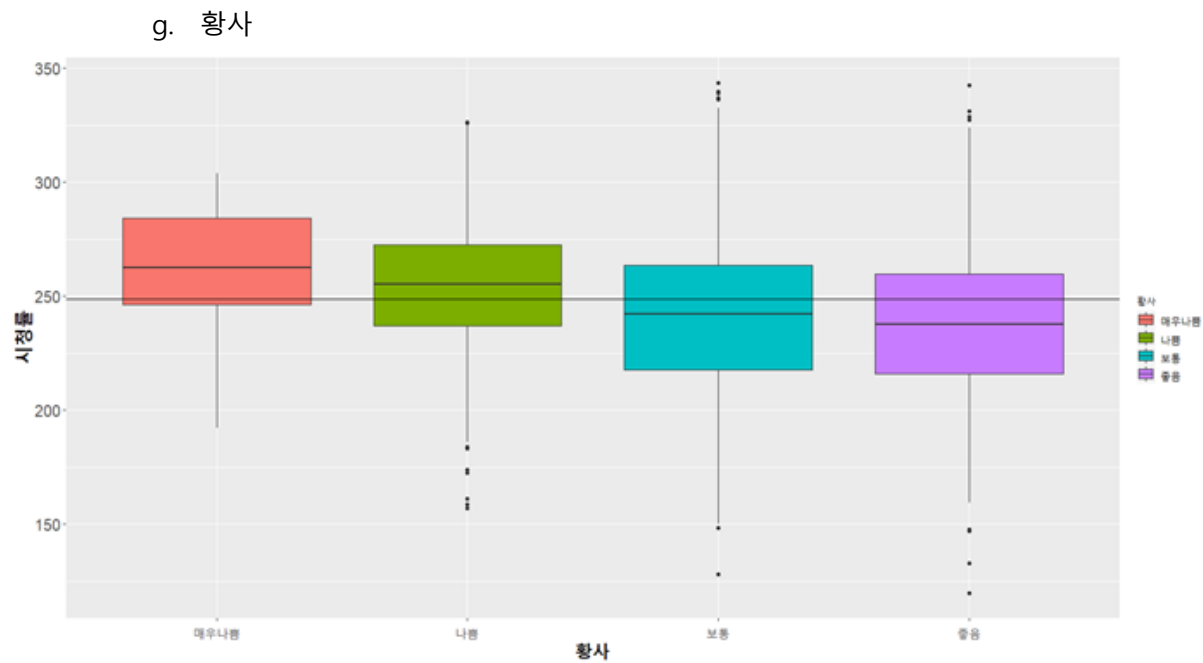
e. 호우



=> 호우가 아닌 날에 비해 호우인 날, 시청률이 22.1523 더 높다.



=> 폭염이 아닌 날에 비해 폭염인 날, 시청률이 23.2716 더 낮다.

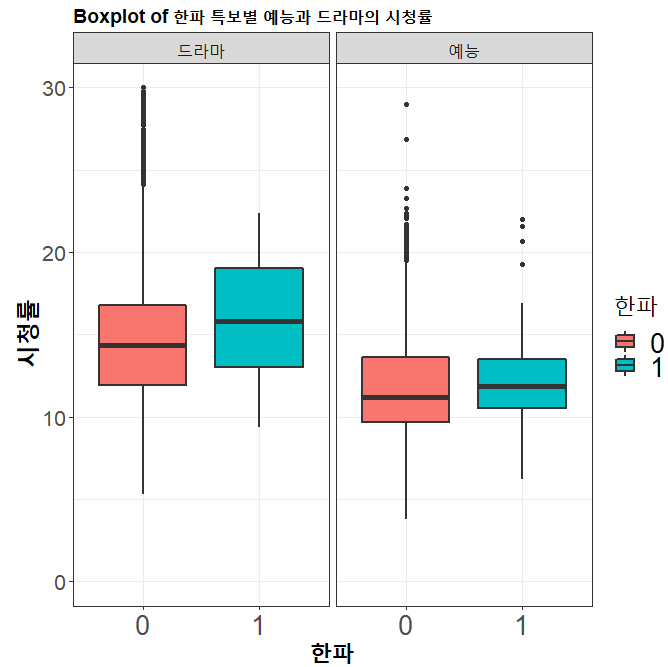


=> 미세먼지 농도가 낮은 날에 비해 높은 날, 시청률이 더 높다.

=> 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 7.617 낮고, 황사가 "보통"인 경우 시청률이 19.013 낮고, 황사가 "좋음"인 경우 시청률이 22.461 낮다.

### 3. 특보별 드라마, 예능 시청률

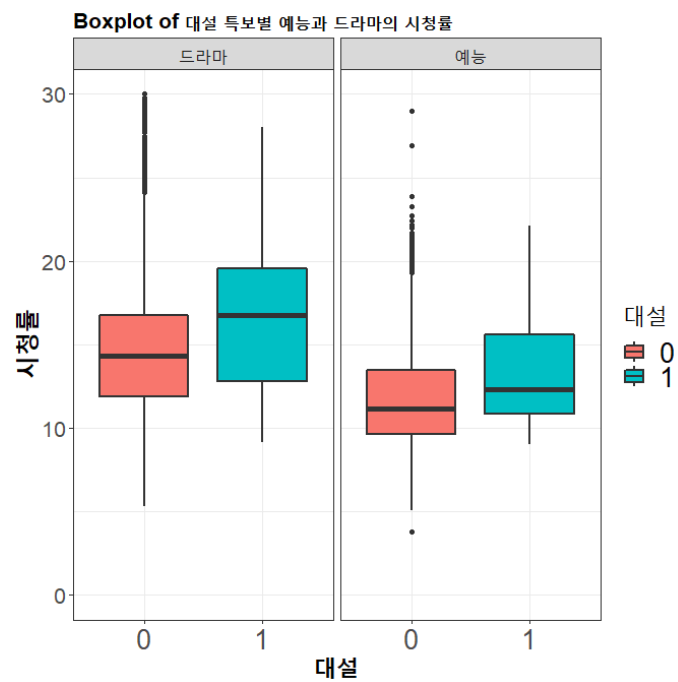
#### a. 한파 - 드라마, 예능



=> 드라마에 비해 예능 시청률의 변화폭은 작지만 한파가 아닌 날에 비해 한파인 날, 드라마와 예능 모두 시청률이 더 높다.

=> 한파가 아닌 날에 비해 한파인 날, 드라마 시청률이 1.39 높은 추세를 보이고, 예능 시청률이 0.6519 높은 추세를 보인다.

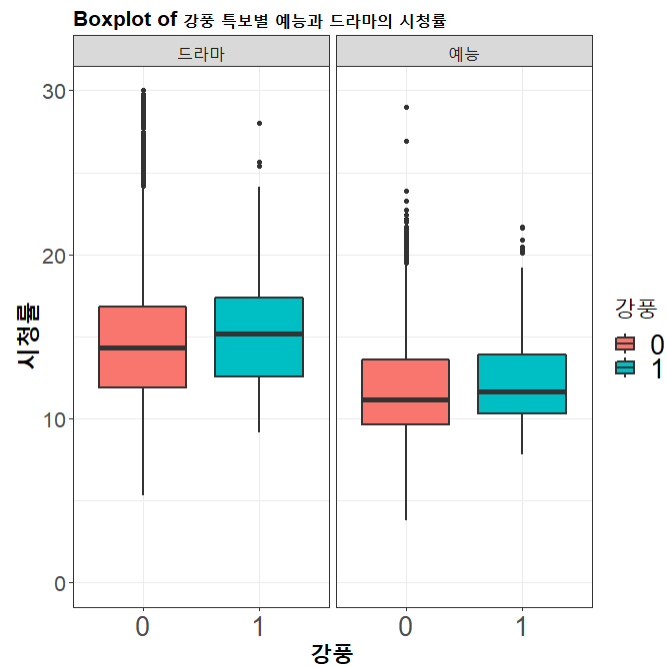
b. 대설- 드라마, 예능



=> 대설이 아닌 날에 비해 대설인 날, 드라마와 예능 모두 시청률이 더 높다.

=> 대설이 아닌 날에 비해 대설인 날, 드라마 시청률이 1.98887 높은 추세를 보이고, 예능 시청률이 1.54954 높은 추세를 보인다.

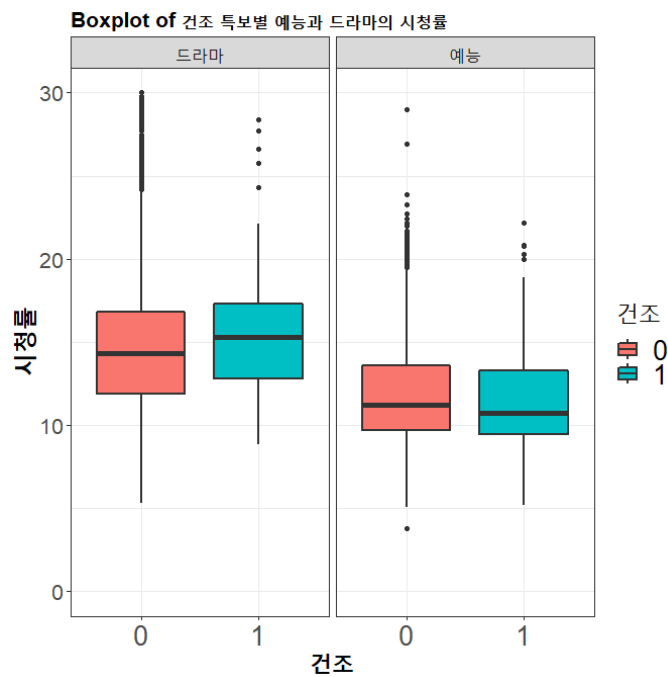
c. 강풍 - 드라마, 예능



=> 강풍이 아닌 날에 비해 강풍인 날, 드라마와 예능 모두 시청률이 더 높다.

=> 강풍이 아닌 날에 비해 강풍인 날, 드라마 시청률이 0.4359 높은 추세를 보이고, 예능 시청률이 0.83653 높은 추세를 보인다.

d. 건조 - 드라마, 예능

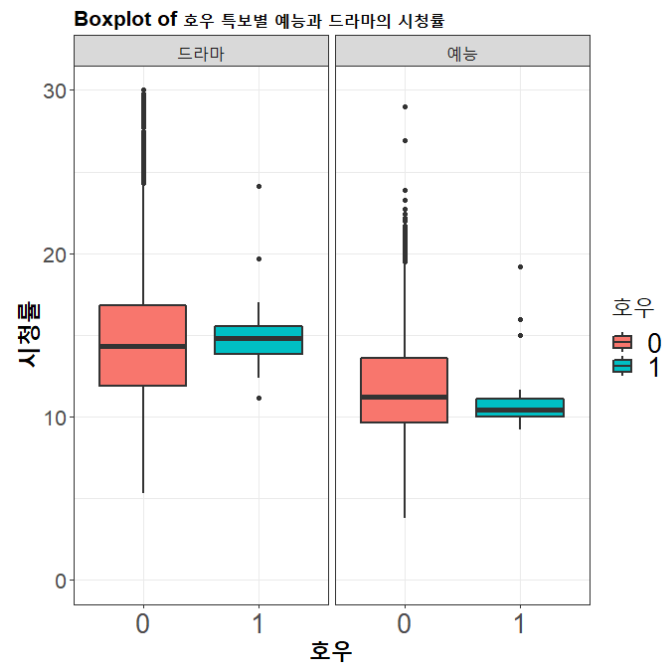


=> 드라마는 건조 특보일 때 시청률이 더 높았고 예능은 건조 특보일때가 건조 특보가 아닐 때보다 시청률이 낮다.

=> 건조가 아닌 날에 비해 건조인 날 드라마는 시청률이 0.63424 더 높은 반면, 예능은 시청률이 0.2897 더 낮다.



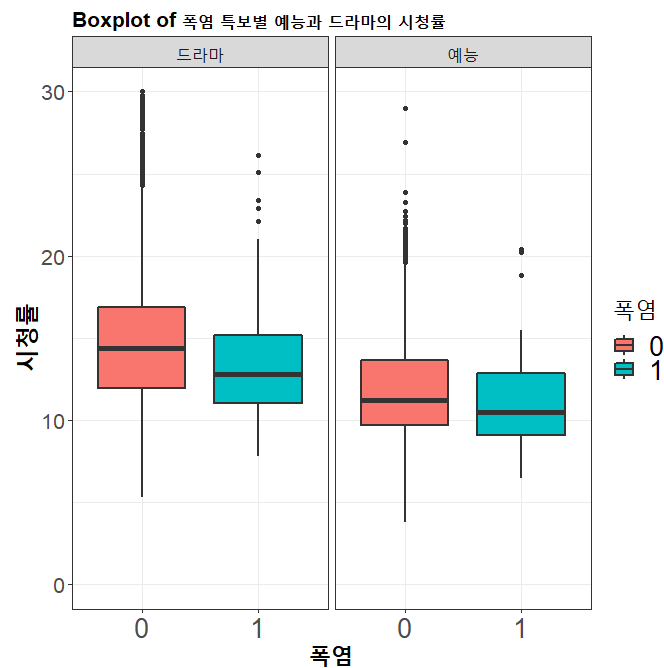
e. 호우 - 드라마, 예능



=> 호우가 아닌 날에 비해 호우인 날 드라마는 시청률이 더 높은 반면, 예능은 시청률이 더 낮다.

=> 호우가 아닌 날에 비해 호우인 날 드라마는 시청률이 0.29052 더 높은 반면, 예능은 시청률이 0.39019 더 낮다.

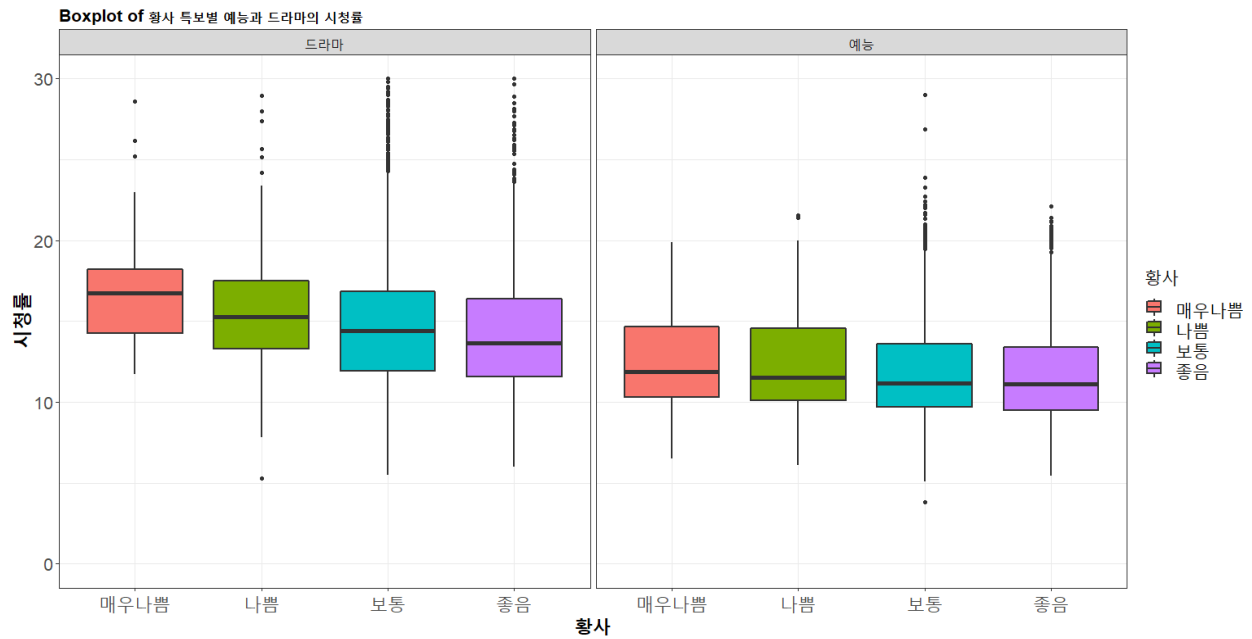
f. 폭염 - 드라마, 예능



=> 드라마에 비해 예능 시청률의 변화폭이 작지만, 폭염이 아닌 날에 비해 폭염인 날, 드라마와 예능 모두 시청률이 더 낮다.

=> 폭염이 아닌 날에 비해 폭염인 날 드라마 시청률이 1.40288 낮은 추세를 보이고, 예능 시청률이 0.63722 낮은 추세를 보인다.

### g. 황사 - 드라마, 예능



=> 드라마에 비해 예능 시청률의 변화폭이 작지만 미세먼지 농도가 높을수록 드라마와

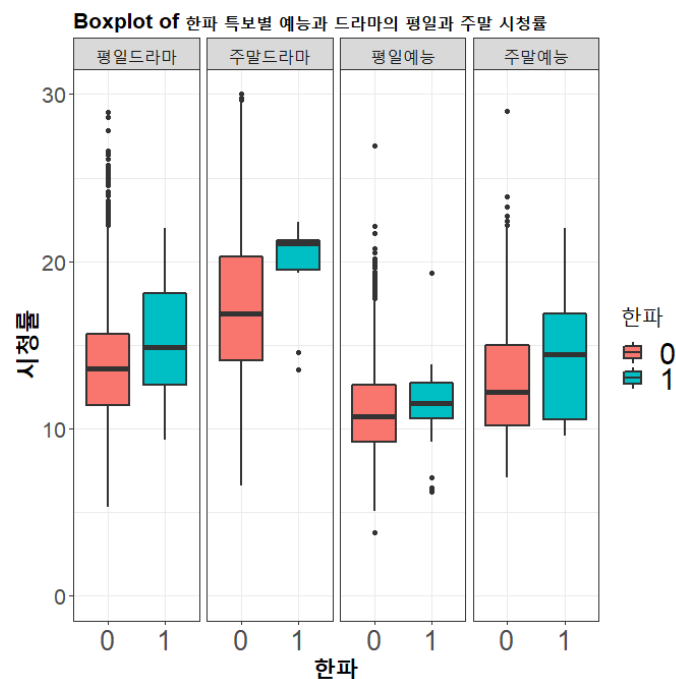
예능 모두 시청률이 더 높다.

=> 드라마 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 1.3276 낮고, 황사가 "보통"인 경우 시청률이 1.9648 낮고, 황사가 "좋음"인 경우 시청률이 2.5362 낮다.

=> 예능 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 0.03621 낮고, 황사가 "보통"인 경우 시청률이 0.49179 낮고, 황사가 "좋음"인 경우 시청률이 0.63985 낮다.

## 4. 특별별 드라마와 예능의 평일, 주말 시청률

### a. 한파 - 평일, 주말



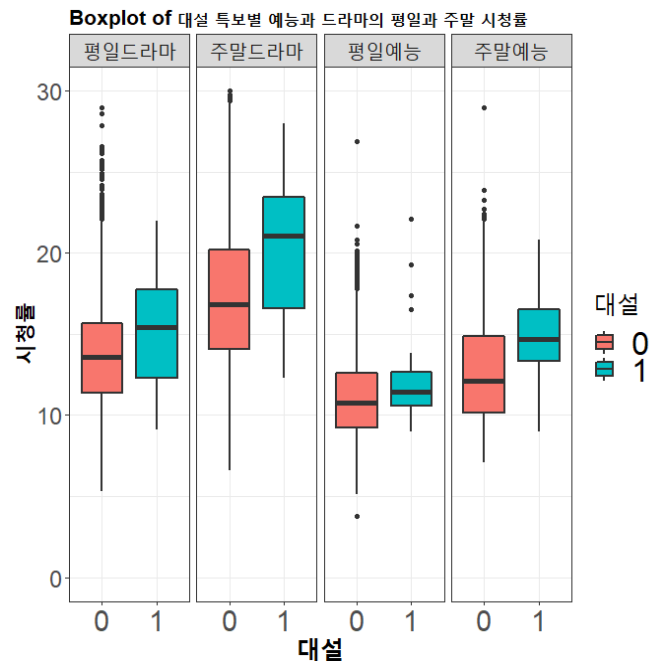
=> 한파가 아닌 날에 비해 한파인 날, 각 장르의 평일과 주말 모두 시청률이 더 높다.

=> 예능에 비해 드라마의 시청률 변화폭이 크다.

=> 한파가 아닌 날에 비해 한파인 날 드라마의 주말 시청률은 2.6678 더 높은 추세를 보이고 평일 시청률은 1.50761 더 높은 추세를 보인다.

=> 한파가 아닌 날에 비해 한파인 날 예능의 주말 시청률은 1.8004 더 높은 추세를 보이고 평일 시청률은 0.26736 더 높은 추세를 보인다.

b. 대설 - 평일, 주말



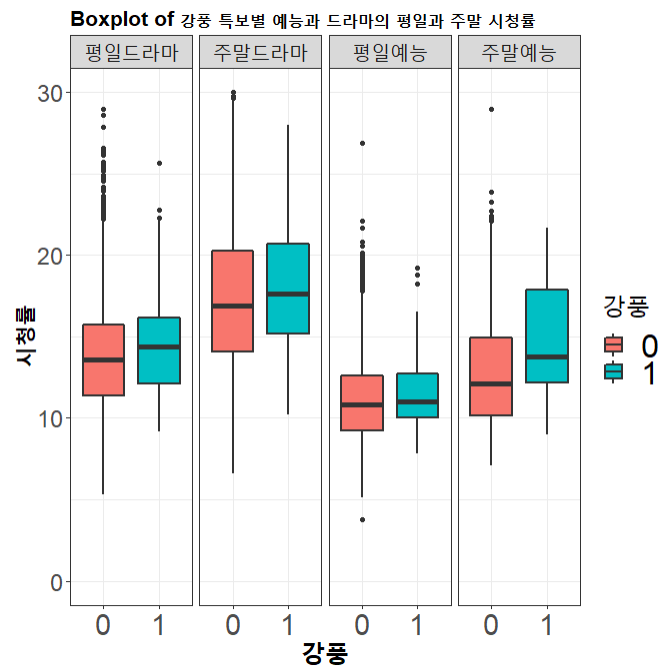
=> 대설이 아닌 날에 비해 대설인 날, 각 장르의 평일과 주말 모두 시청률이 더 높다.

=> 예능에 비해 드라마의 시청률 변화폭이 크다.

=> 대설이 아닌 날에 비해 대설인 날 드라마의 주말 시청률은 2.9725 더 높은 추세를 보이고 평일 시청률은 1.46352 더 높은 추세를 보인다.

=> 대설이 아닌 날에 비해 대설인 날 예능의 주말 시청률은 1.9895 더 높은 추세를 보이고 평일 시청률은 1.12242 더 높은 추세를 보인다.

c. 강풍 - 평일, 주말



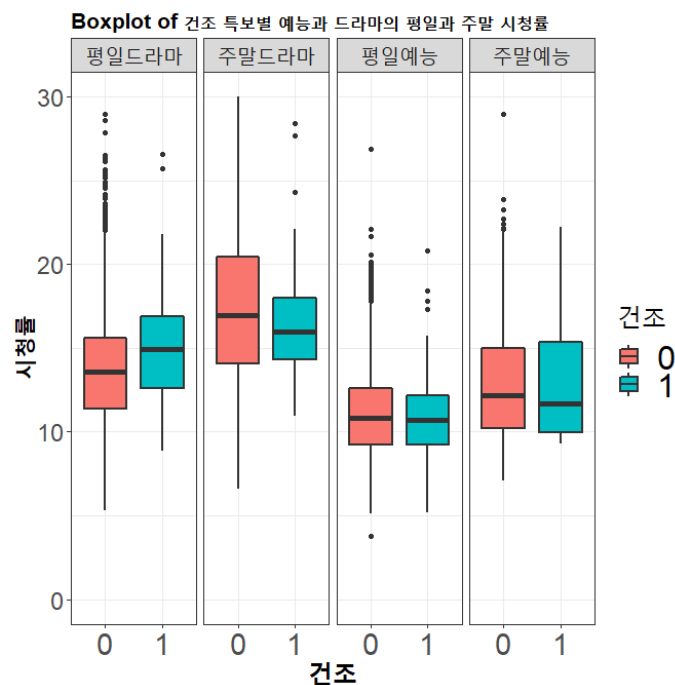
=> 강풍이 아닌 날에 비해 강풍인 날, 각 장르의 평일과 주말 모두 시청률이 더 높다.

=> 예능에 비해 드라마의 시청률 변화폭이 크다.

=> 대설이 아닌 날에 비해 대설인 날 드라마의 주말 시청률은 2.9725 더 높은 추세를 보이고 평일 시청률은 1.46352 더 높은 추세를 보인다.

=> 대설이 아닌 날에 비해 대설인 날 예능의 주말 시청률은 1.9895 더 높은 추세를 보이고 평일 시청률은 1.12242 더 높은 추세를 보인다.

d. 건조 - 평일, 주말

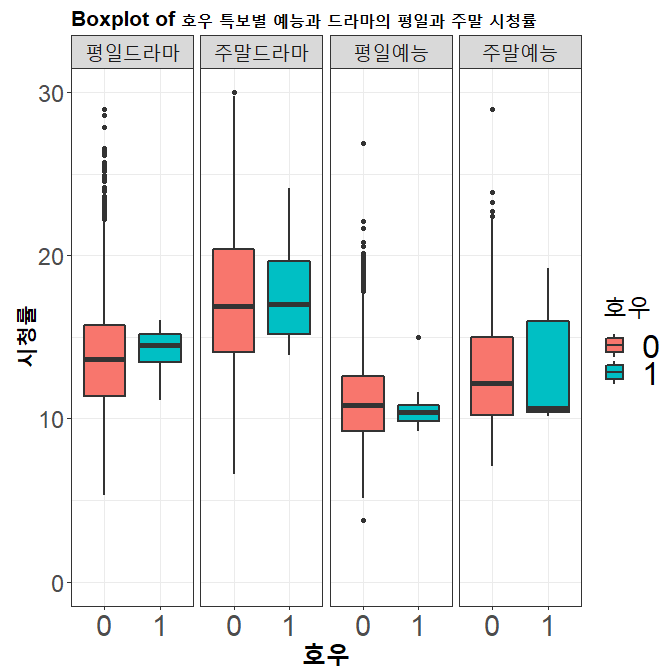


=> 건조가 아닌 날과 건조인 날, 각 장르별, 요일별의 변화 차이가 다양하다.

=> 건조가 아닌 날에 비해 건조인 날 드라마의 주말 시청률은 0.6153 더 낮은 추세를 보이고 평일 시청률은 1.24472 더 높은 추세를 보인다.

=> 건조이 아닌 날에 비해 건조인 날 예능의 주말 시청률은 0.1992 더 높은 추세를 보이고 평일 시청률은 0.40354 더 낮은 추세를 보인다.

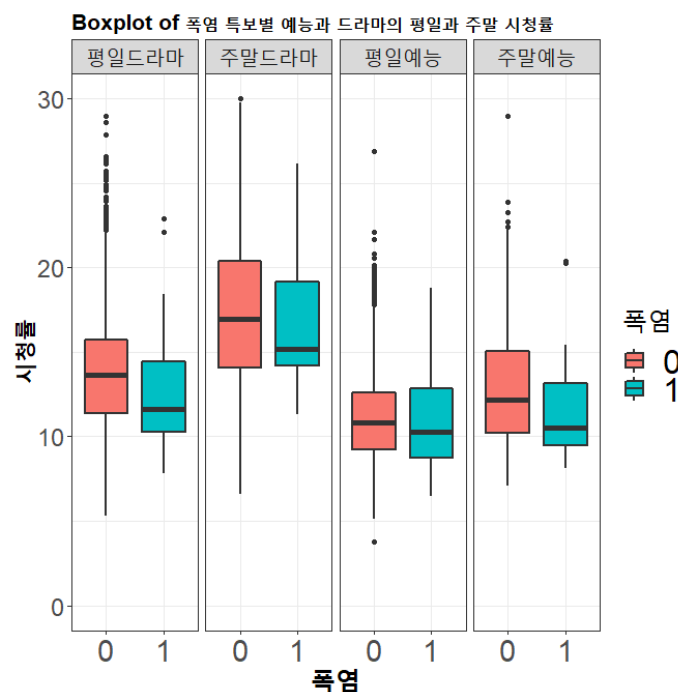
#### e. 호우 - 평일, 주말



=> 호우가 아닌 날에 비해 호우인 날, 드라마 시청률의 변화는 미미하며, 예능 시청률은 더 낮지만 미미하다.

=> 회귀식에서 유의하지 않은 변수이며 변화량의 차이가 미미하고 장르별 평일, 주말 마다 모두 다르다.

#### f. 폭염 - 평일, 주말

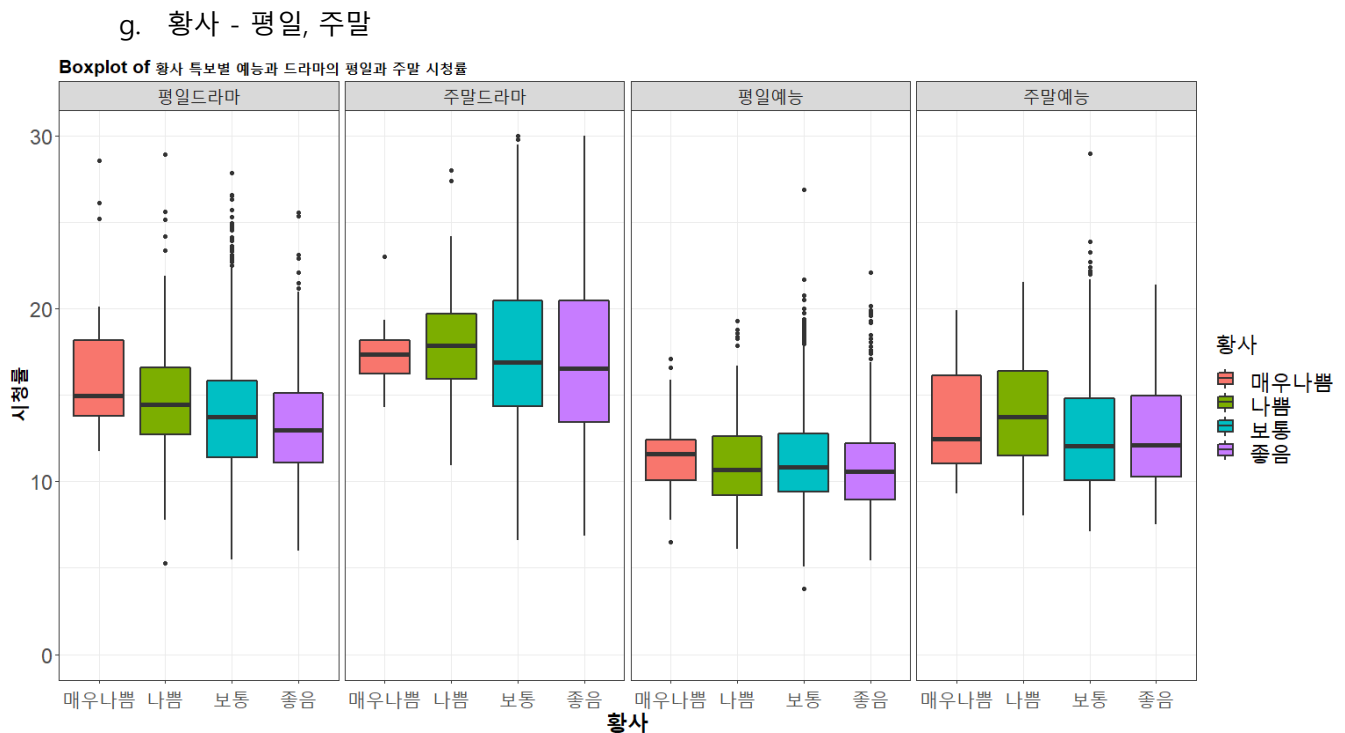


=> 폭염이 아닌 날에 비해 폭염인 날, 각 장르의 평일과 주말 모두 시청률이 낮다.

=> 예능에 비해 드라마의 시청률 변화폭이 크다.

=> 폭염이 아닌 날에 비해 폭염인 날 드라마의 주말 시청률은 1.2818 더 낮은 추세를 보이고 평일 시청률은 1.25705 더 낮은 추세를 보인다.

=> 폭염이 아닌 날에 비해 폭염인 날 예능의 주말 시청률은 0.9812 더 낮은 추세를 보이고 평일 시청률은 0.44709 더 낮은 추세를 보인다.



=> 미세먼지 농도가 낮은 날에 비해 높은 날, 각 장르의 평일과 주말 모두 시청률이 높다.

=> 예능보다 드라마 시청률의 변화폭이 크다.

=> 주말 드라마 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 0.8956 높고, 황사가 "보통"인 경우 시청률이 0.1382 낮고, 황사가 "좋음"인 경우 시청률이 0.1443 낮다.

=> 평일 드라마 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 2.0983 낮고, 황사가 "보통"인 경우 시청률이 3.0011 낮고, 황사가 "좋음"인 경우 시청률이 3.6401 낮다.

=> 주말 예능 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 0.5884 높고, 황사가 "보통"인 경우 시청률이 0.6022 낮고, 황사가 "좋음"인 경우 시청률이 0.4643 낮다.

=> 평일 예능 시청률은 황사가 "매우 나쁨"인 경우에 비해, 황사가 "나쁨"인 경우 시청률이 0.2883 낮고, 황사가 "보통"인 경우 시청률이 0.2773 낮고, 황사가 "좋음"인 경우 시청률이 0.7279 낮다.

## 5. 예측 결과

### a. 한파

한파 특보 발령시 추운 날씨로 인해서 집 밖으로 나가지 않으려는 경향이 크기 때문에 집에서 TV시청을 할 것이라는 예상과 동일하게 시청률이 올라가는 결과가 도출 되었다. 평균기온이 높아질수록 “물강족”, “울빼미족”과 같이 냉방이 잘 되어 있는 공공장소를 찾아나서기 때문에 시청률이 떨어지는 결과가 나온 것처럼 반대로 평균기온이 낮아질수록 추위 때문에 집에만 있으려는 경향이 있어 시청률이 높아진다.

### b. 대설

최심적설이 5cm 이상일 때 대설 특보가 발령 되는데 많이 내리는 눈으로 인해서 집 밖으로 나가지 않으려는 경향이 커서 TV 시청자가 많을 것이라는 예상과 동일하게 시청률이 높아지는 결과가 도출되었다. 또한 최심적설의 수치형 데이터로도 적설량이 많아질수록 시청률이 올라가는 결과와 동일하다.

### c. 강풍

강풍특보가 발령되었을 때 시청률이 높을 것이라는 예상과 동일하게 강풍특보가 발령되면 강한 바람에 밖으로 나가지 않으려는 사람들이 많아 시청률이 높은 결과가 도출되었다. 수치형 자료에서 풍속 데이터로도 풍속이 셀수록 시청률이 올라가는 결과 추세에 맞게 도출되었다.

### d. 건조

건조특보가 발령되었을 때 시청률이 높을 것이라는 예상과는 반대로 건조 특보가 발령되면 시청률의 변화의 폭도 작고 또한 장르별, 요일별로 분류하여 시각화를 통해 나온 결과 또한 시청률의 변화가 제각각인 것으로 보아 건조 특보는 시청률에 영향을 미치지 않은 것으로 나타났다.

### e. 호우

호우 특보가 발령되었을 때 비가 많이 내려 밖으로 나가려하지 않는 경향으로 시청률이 높을 것이라는 예상과는 달리 건조 특보와 마찬가지로 장르별, 요일별 시청률의 시각화 결과가 모두 제각각인 것을 보아 호우 특보는 시청률에 영향을 미치지 않은 것으로 나타났고 또한 수치형 자료의 강수량 데이터를 통해 결과를 도출해도 강수량은 시청률에 영향을 미치지 않는 것을 알 수 있다.

### f. 폭염

폭염 특보가 발령되었을 때 기온이 높아 가만히 있어도 더운 밖으로 나가지 않으려는 경향으로 집에서 TV를 시청할 것이라는 예상과는 반대로 기온이 높아 폭염 특보가 발령되면 시청률이 낮아지는 결과가 도출되었다. 이는 집보다는 시원한 에어컨이 항상 나오는 카페, 백화점, 마트 등으로 피하는 “물강족”과 선선한 저녁에 운동을 하는 “울빼미족”의 영향등으로 폭염 특보가 발령되면 빠르게 에어컨이 잘 나오는 공공장소로 사람들이 이동하여 시청률이 낮아진다.

### g. 황사

황사 특보가 발령되었을 때 밖의 미세먼지로 인해서 집에서 나가지 않으려는 경향으로 집에서 TV를 시청할 것이라는 예상과 동일하게 황사 특보가 발령되면 시청률이 높아지는 결과가 도출되었다. 미세먼지 농도는 80이상부터 “나쁨”,

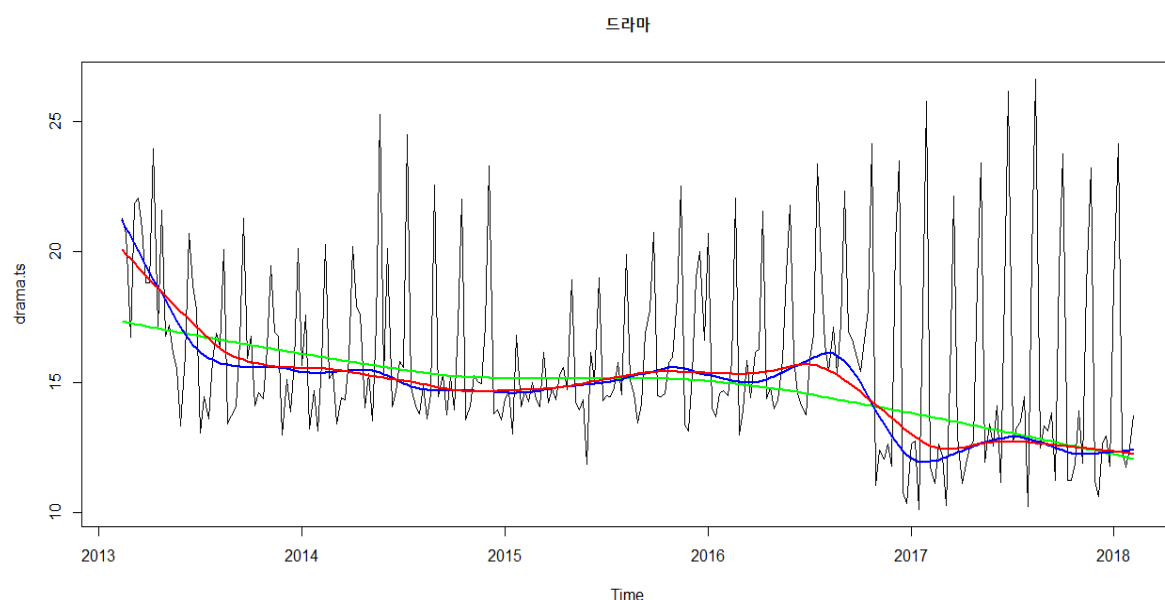
150이상이면 "매우 나쁨" 인데 최근에 미세먼지로 인한 건강 악화 기사와 미세먼지의 위험성의 인식이 점차 높아지는 등의 영향으로 미세먼지 농도가 높을수록 집 밖으로 나가지 않아 시청률이 높아지는 결과가 나타난다.

### 시간대별로 하지 않은 이유

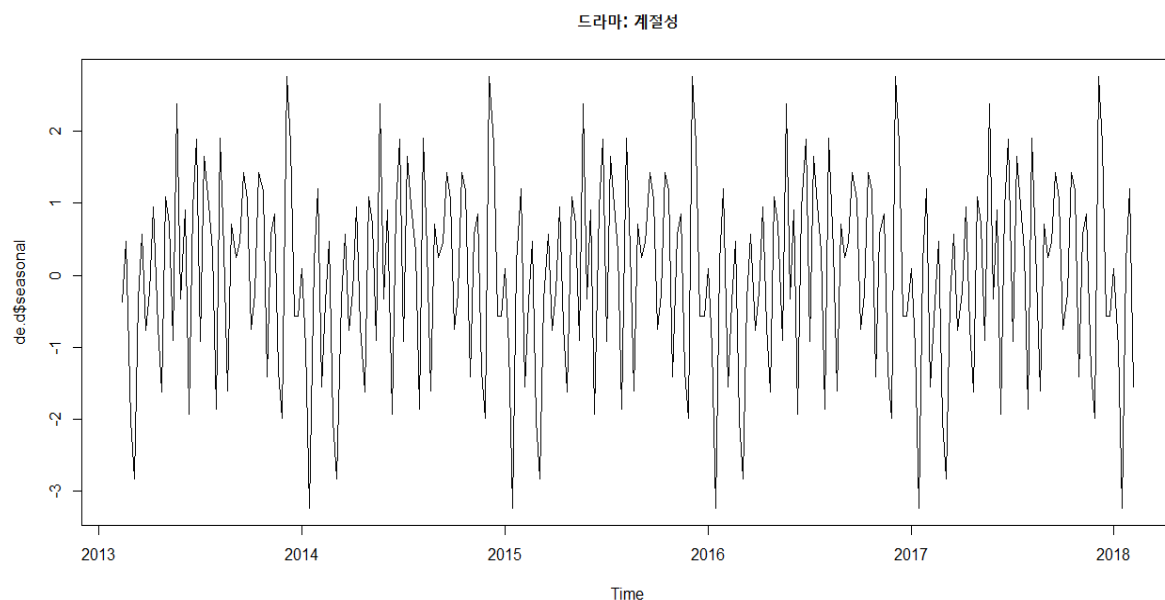
- 장르별로 시간대로 나뉘었을 때 날씨별 장르 그리고 날씨별 장르의 요일의 결과와 동일한 결과를 나타내어 시간대별로 시각화하여 명시하지 않았다.

## ㄷ. 시계열 분석

### 1. 시계열 - 드라마

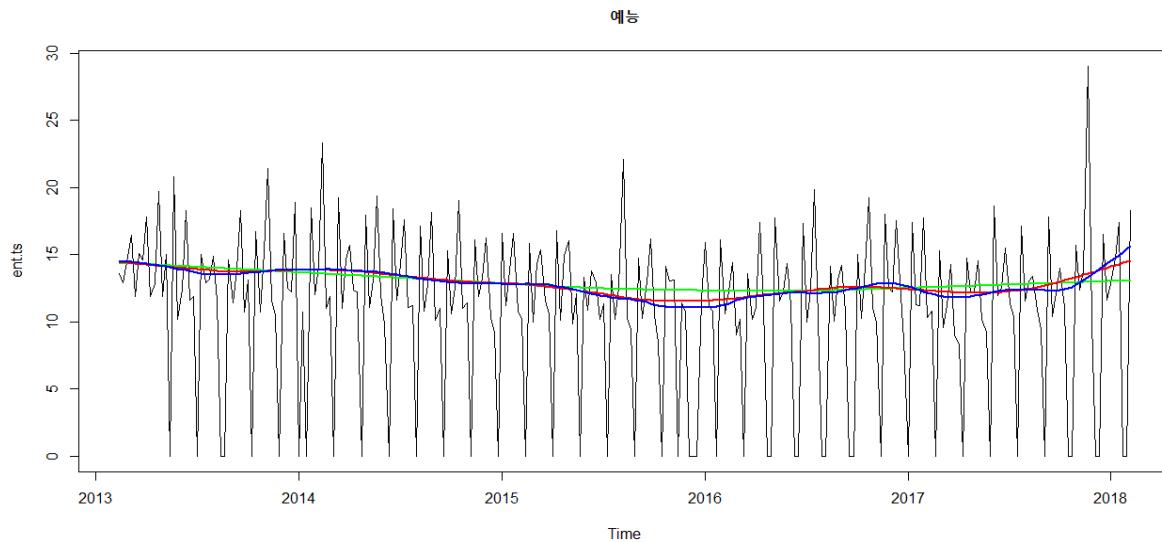


=> 드라마의 시청률은 점차 감소하는 추세를 보인다.

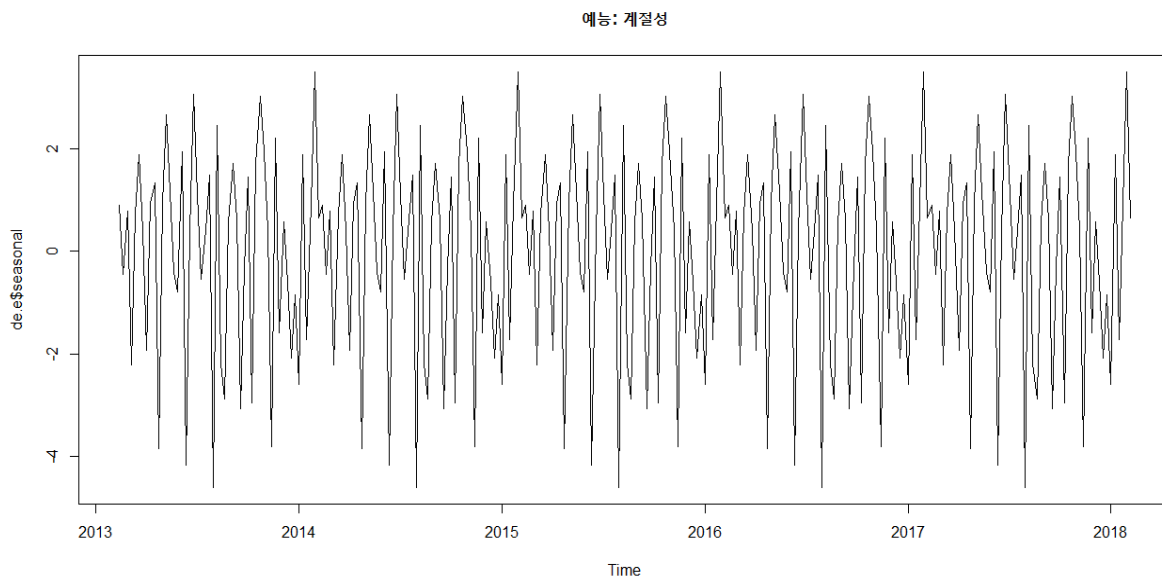




## 2. 시계열 - 예능



=> 예능의 시청률은 꾸준히 이어지는 추세이다.



## 3. 시계열 - 분석

드라마 시청률의 추세는 점차 감소하는 반면 예능 시청률의 추세는 꾸준히 이어지는 추세를 보인다.

또한 드라마와 예능 시청률 모두 계절성이 존재하는 것으로 보아, 계절적 요인이 영향을 미친다고 볼 수 있다.

따라서, 시계열 자료 분석 및 예측에 사용되는 RNN, LSTM 적용

## □. 분석 및 예측 모델 구현

## 1. 예측 데이터 셋 정의

- 시각화 결과 수치형은 평균기온, 풍속, 일사량, 최심적설, 미세먼지 변수가 날씨와 연관 있을 것으로 판단
- 특보의 경우 건조와 호우를 제외한 변수들이 날씨와 연관 있을 것으로 판단
- 예측 데이터 셋 정의
- top20 전체 시청률을 날짜별로 합산한 all\_Data
- 배달데이터와 구글트렌드 변수 추가

```
< all_Data >
```

Rate	평균기온	풍속	일사량	최심적설	미세먼지	강풍	미결	한파	폭염	Call	Trend	황사.나쁨	황사.매우나쁨	황사.보통	황사.중음
228.0	28.0	1.7	12.73	0	33.7	0	0	0	0	60829	5.608696	0	0	1	0
218.0	27.4	2.0	5.49	0	36.7	0	0	0	0	69136	6.702899	0	0	1	0
204.8	28.9	3.7	16.52	0	21.1	0	0	0	0	82485	6.702899	0	0	0	1
234.5	28.3	2.4	11.08	0	39.2	0	0	0	0	82276	6.144928	0	0	1	0
225.7	27.7	3.2	10.13	0	45.0	1	0	0	0	56185	6.144928	0	0	1	0
222.5	26.4	2.4	4.71	0	36.1	0	0	0	0	60502	5.804348	0	0	1	0
214.5	28.7	2.3	11.01	0	41.8	0	0	0	0	58637	5.804348	0	0	1	0
223.8	29.3	3.7	9.72	0	51.8	0	0	0	0	61436	5.949275	0	0	1	0
233.5	29.2	4.1	4.98	0	78.2	0	0	0	0	72126	5.949275	0	0	1	0
225.7	28.0	2.4	6.17	0	74.8	0	0	0	0	86000	5.123188	0	0	1	0

2. 머신러닝 - 앙상블 기법인 randomforest와 xgboost 구현하여 R2 비교

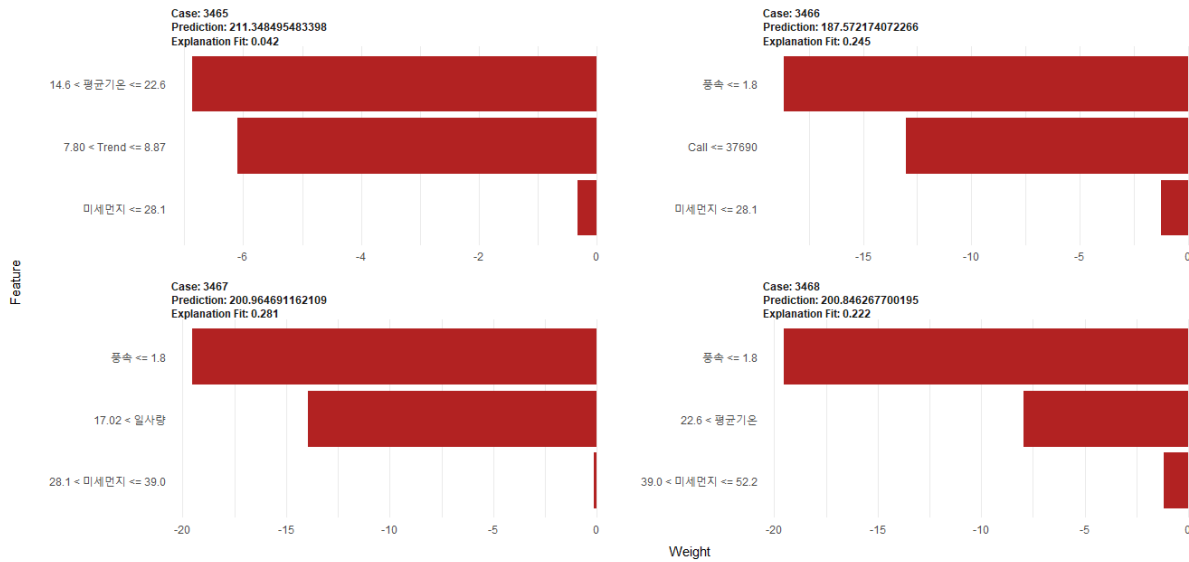
### < all\_Data 모델 비교 >

Data set	Model	RMSE	R2
All Data	Random forest	12.21	85.32 %
	Xgb boost	15.31	76.56 %

=> 두개의 모델 비교 결과 Random forest 모델이 RMSE 85.34%로 더 좋은 성능을 나타냈다.

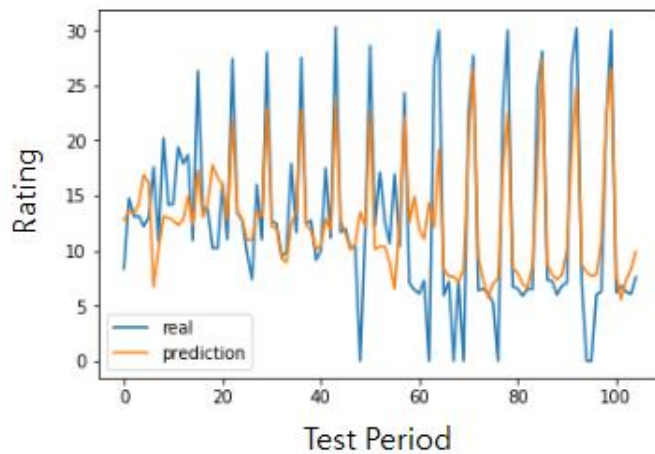
### 3. Lime 알고리즘 이용 각 모델 예측값 설명력 해석

model_prediction	feature	feature_value	feature_weight	feature_desc
206.2893	미세먼지	25.600000	-0.3172955	미세먼지 <= 28.1
206.2893	평균기온	22.300000	-6.8712798	14.6 < 평균기온 <= 22.6
206.2893	Trend	8.253623	-6.0974903	7.80 < Trend <= 8.87
192.2722	미세먼지	21.400000	-1.2747108	미세먼지 <= 28.1
192.2722	풍속	1.100000	-18.6625664	풍속 <= 1.8
192.2722	call	35033.000000	-13.0218818	call <= 37690
191.0202	미세먼지	28.200000	-0.1578313	28.1 < 미세먼지 <= 39.0
191.0202	풍속	1.500000	-19.5553278	풍속 <= 1.8
191.0202	일사량	18.790000	-13.9588933	17.02 < 일사량
195.3270	미세먼지	39.900000	-1.1666904	39.0 < 미세먼지 <= 52.2
195.3270	풍속	0.900000	-19.5422318	풍속 <= 1.8
195.3270	평균기온	24.400000	-7.9440216	22.6 < 평균기온



- model\_prediction은 모델을 통해 예측된 테스트 데이터의 예측값이다.
- feature\_weight는 예측값을 설명하는 변수의 가중치가 되고,  
feature\_desc는 해당 예측값을 설명해주는 변수의 범위가 된다.
- 예를 들어 1~3행은 하나의 테스트 데이터로 예측값이 206.28수치가 나온 이유를 미세먼지가 28.1보다 작고, 평균기온이 14.6과 22.6 사이에 있으며 구글트렌드가 7.8과 8.87 사이에 속하기 때문이라고 설명 할 수 있다.

#### 4. 딥러닝 - RNN에서 확장되어 장기기억 보존을 해주는 LSTM 구현



< all\_Data 시계열 예측 >

=> LSTM을 통해 약 100일의 테스트 값을 예측 후 실제값과 비교 해본 결과 위와 같은 예측을 나타내고 있으며 초반 부 0~10일 사이를 제외하고 시계열데이터의 예측이 잘 되었다고 여겨진다.

## 5. 서비스 활용 방안

### ㄱ. 방송사

#### 1. 예측 정확도

기상 정보를 이용해 시청률의 예측 정확도를 높여 보았다. 이 정보를 통해 가장 이익을 극대화할 수 있는 집단은 어디일까. 먼저는 '방송사'이다.

지상파든 공중파이든 너나 할 것 없이 모든 방송사에서 염두에 두고 있는 것은 '시청률을 높이는 것'이다. 창출되는 모든 이익이 높은 시청률을 달성한 시간대의 '광고료'에서 비롯되기 때문이다.

때문에 시청률을 높이기 위해 톱스타를 섭외하거나 뛰어난 연출을 보여주는 등의 기존 방식으로 방송사들 사이에선 끊임없는 시청률 쟁탈전을 벌여 오고 있다. 여기에 날씨 변수로 도출한 시청률 예측치로 효율적인 시간당 광고 할당 방식을 도입한다면 탁월한 이익 창출을 기대할 수 있다.

날씨로 시청률을 예측하는 것은 또한 방송사에서 시청자의 니즈(needs)를 충족시킬 수 있는 또 하나의 방법이다. 갑작스런 날씨 변화로 대다수의 외출 일정이 취소 될 것이 예상 된다면 해당 방송사에서는 프로그램 편성을 조정해 동시간대 타 방송사에 비해 시청자의 화면 고정 비율을 높일 수 있다.

#### 2. 설명력

물론 예측 정확도도 중요하지만 예측된 시청률이 어떤 변수에 의해 작용했는지 또한 중요하다. 단순히 시청률이 높을 것으로 결론 내린 정보를 편성에 즉각 반영하기보다 어떤 날씨 변수가 시청률에 영향을 끼치는지 면밀히 파악 후 해당 기상 정보에 맞추어 프로그램을 배치하는 것이 더욱 효과적인 시청률 증진을 보일 수 있다. 가령, 일주일 뒤의 시청률이 평소보다 5퍼센트가 증가 할 것으로 예측 되고, 기상 변수가 대설·한파 특보와 같은 기상 이변 이라면 방송사에서는 따뜻한 음식 콘텐츠를 보이거나 추운 몸을 녹여 줄 따뜻한 효과를 연출하여 시청자의 니즈(needs)를 충족시킬 수 있다.

### ㄴ. 광고주

#### 1. 예측 정확도

두 번째로 수혜를 얻을 집단은 '광고주'이다.

광고주 역시 방송사와 유사한 경로로 이익을 얻는다. 방송사를 통해 광고를 런칭(launching)하는 광고주들은 보다 높은 시청률, 높은 화면 고정률인 프로그램을 찾는 것이 목적이다.

기상 변수로 시청률을 예측하는 것은 즉, 광고주 입장에서도 동일한 비용으로 보다 높은 광고 효과를 기대할 수 있다. KBS에서 35%의 시청률을 기록한 'A'드라마의 경우 15초 간 광고를 내보내는데 1,500만원의 비용을 보였다. 이 경우, 기상 정보로 예측된 시청률을 적용한다면 더 적은 비용으로 더 높은 시청률을 예상하는 날짜와 시간대로 변경하여 최대 효과를 창출할 수 있다.

## 2. 설명력

광고주 또한 모델 정확도만큼이나 해당 예측값의 설명력이 중요하다. 예를 들어 보자. 지금으로부터 일주일 뒤의 동 시간대 시청률이 평소보다 5퍼센트가 오를 것으로 예상되는 상황이다. 그때 영향을 끼칠 변수가 미세먼지 또는 한파라고 한다면 공기청정기 회사나 겨울 외투 회사는 반드시 그 시간대 광고 유치에 힘써야 한다. 해당 프로그램 내에 광고하는 ppl 요소 또한 이와 같은 논리로 효과적인 접목이 가능할 것이다.

## 6. 서비스 기대효과

연 2조 9천억원에 이르는 방송 시장에서 방송사와 광고주 사이, 또는 광고주들 간의 광고 쟁탈전 최고의 무기는 시청률이다. 지상파에서 35~40%의 시청률에 육박하는 프로그램의 경우 15초당 광고사로부터 1,200만원 광고료가 방송사와 광고주 사이에서 오고 간다. 방송사는 광고주로부터 최대수익을 얻기 위해 시청률을 높이고 싶어 하고 광고주는 최소비용으로 최대효과를 누리고 싶어 할 것이다. 이렇게 TV 광고 시장에서 날씨정보가 영향을 미친다면 이는 절대 간과해서는 안된다. 이를 적극 활용해 효율적인 경영이 이루어져야 한다.

날씨데이터를 활용해 시청률을 예측 및 설명을 할 수 있게 된다면, 방송사에게는 프로그램편성 의사결정과 해당 프로그램의 콘텐츠를 제공해 줄 수 있고,

광고주에게는 최소비용 최대효과의 광고 런칭과, 날씨에 영향을 받는 재화 및 서비스에 날씨 마케팅을 해주는 등 여러가지 기대효과를 누릴 수 있다.