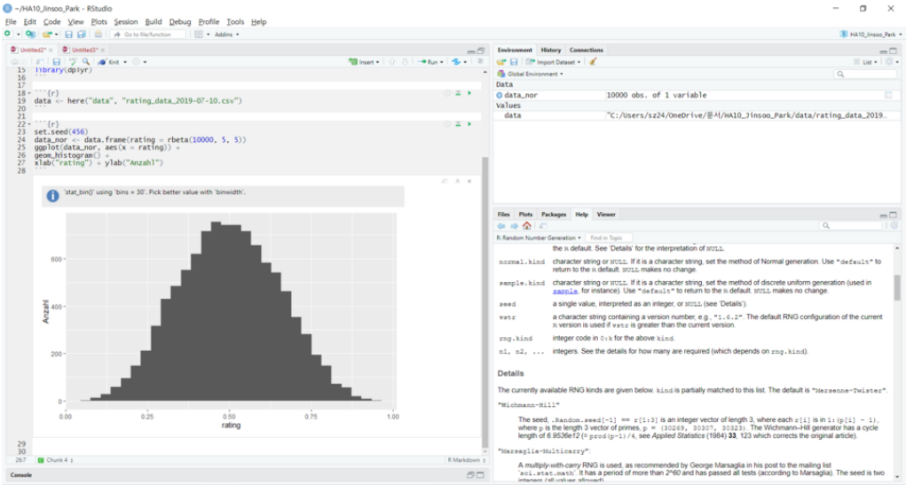
3. 데이터 처리 방안 및 활용분석기법

1.데이터 처리 방안

1. 데이터 처리 기법 소개

* R-Studio : 통계 컴퓨팅, 그래픽을 위한 프로그래밍 언어인 R을 위한

자유-오픈 소스 통합 개발환경(IDE)





R-Studio를 활용하여 강수량, 기온, 풍속, 미세먼지, 습도, 일조량 각각과

영화관 관객 수 데이터의 상관관계를 분석

* 파이썬 아나콘다 : 데이터 전처리 후 모델링에 활용



(3)결측치 처리

import pandas as pd

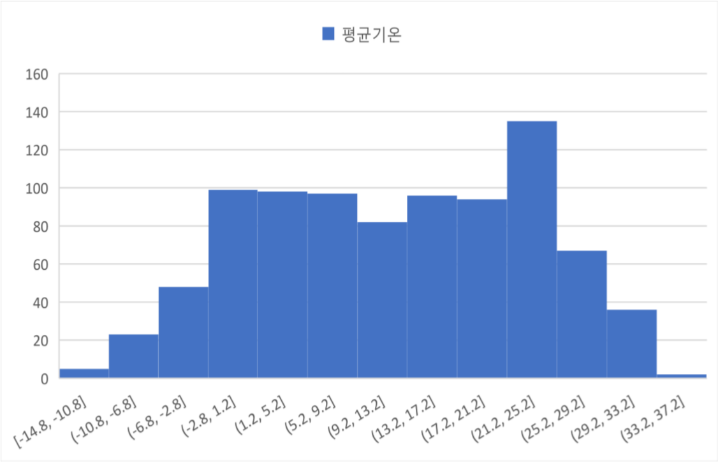
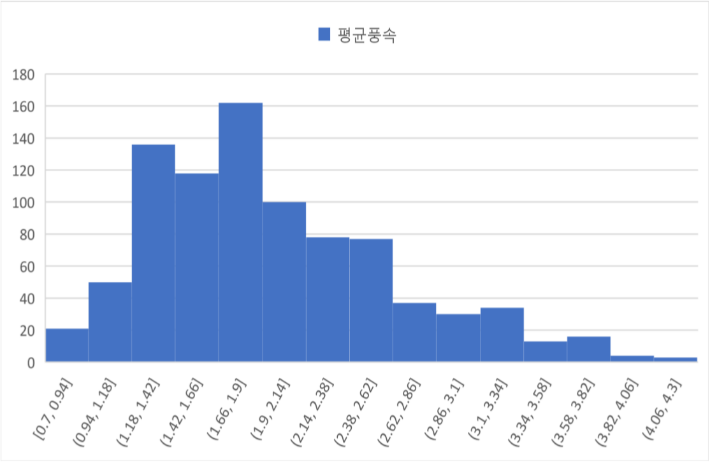
data = pd.read\_excel('total\_new.xlsx')

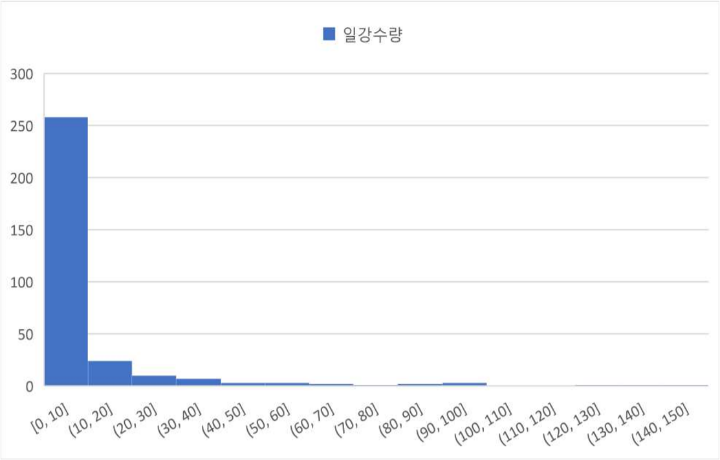
data = data.dropna(axis=0)

dropna(axis=0)을 통해 결측값 있는 행 제거

총 데이터 882개 중 795개로 결측치 처리

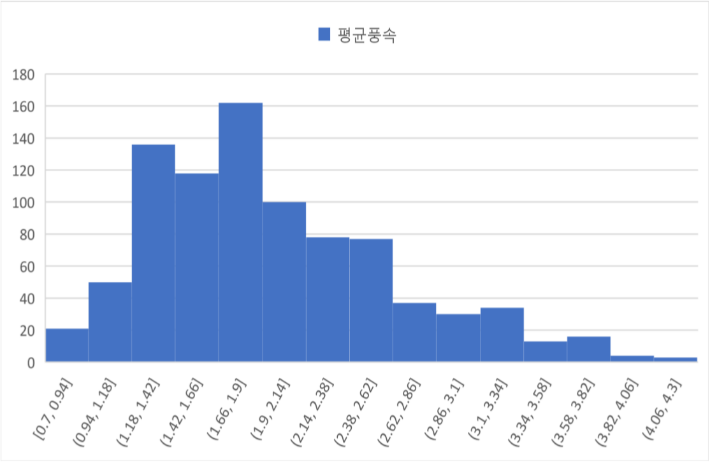
(4)데이터 그래프(히스토그램)

* 평균기온
* 영하 10도 이하
* 영하 10도 ~ 0도
* 0도 ~ 영상 10도
* 영상 10도 ~ 영상 20도
* 영상 20도 ~ 영상 30도
* 영상 30도 이상
* 일강수량

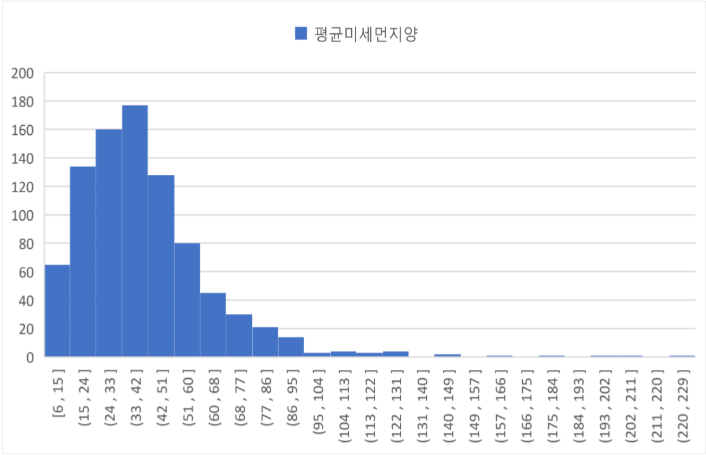


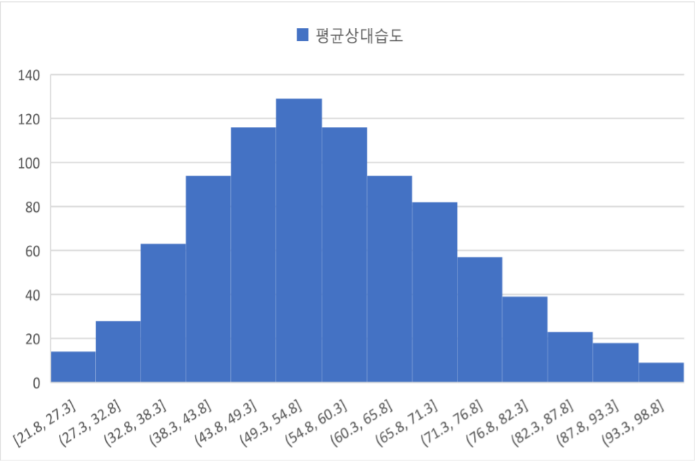
* 0mm
* 0mm – 10mm
* 10mm – 30mm
* 30mm 초과

* 평균풍속



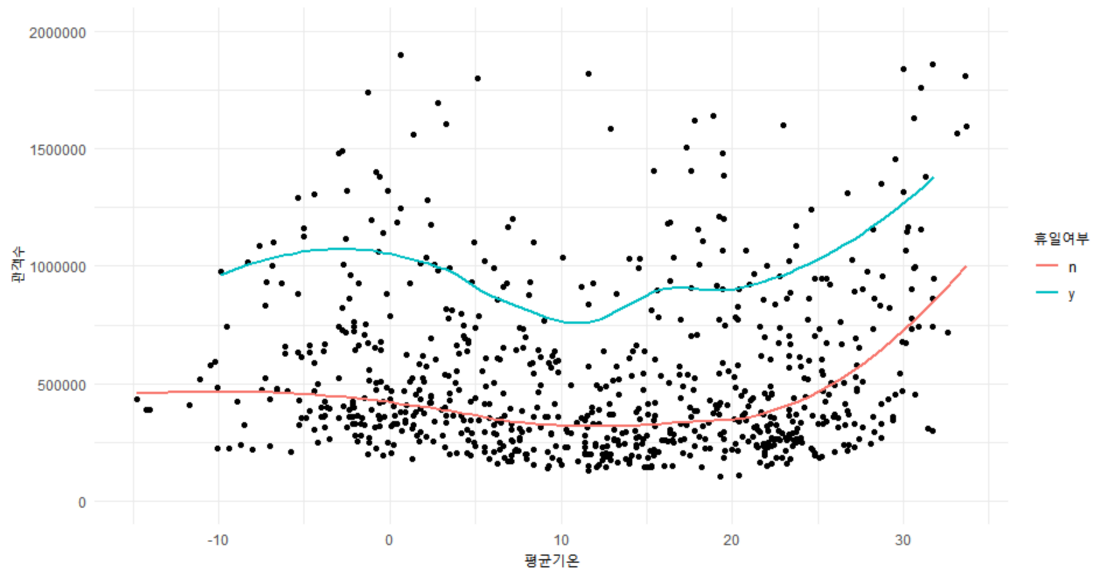
* 1 m/s 이하
* 1 m/s – 2 m/s
* 2 m/s – 3 m/s
* 3 m/s 초과
* 미세먼지 농도

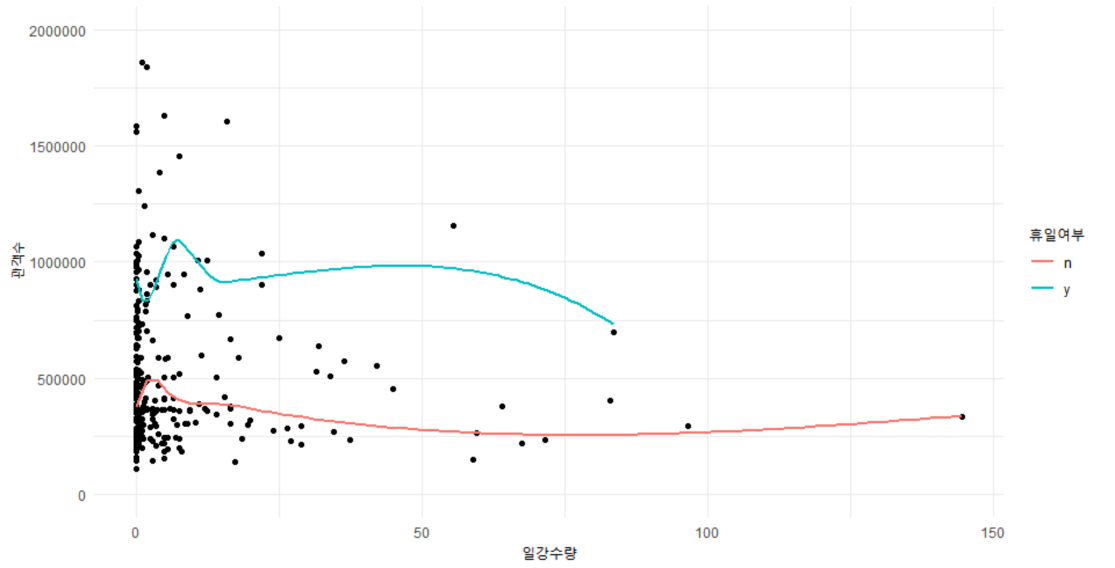


* 30㎍/㎥ 이하
* 30㎍/㎥ – 70 ㎍/㎥
* 70㎍/㎥ 초과
* 상대습도
* 50% 이하
* 50% – 60%
* 60% – 80%
* 80% 이상

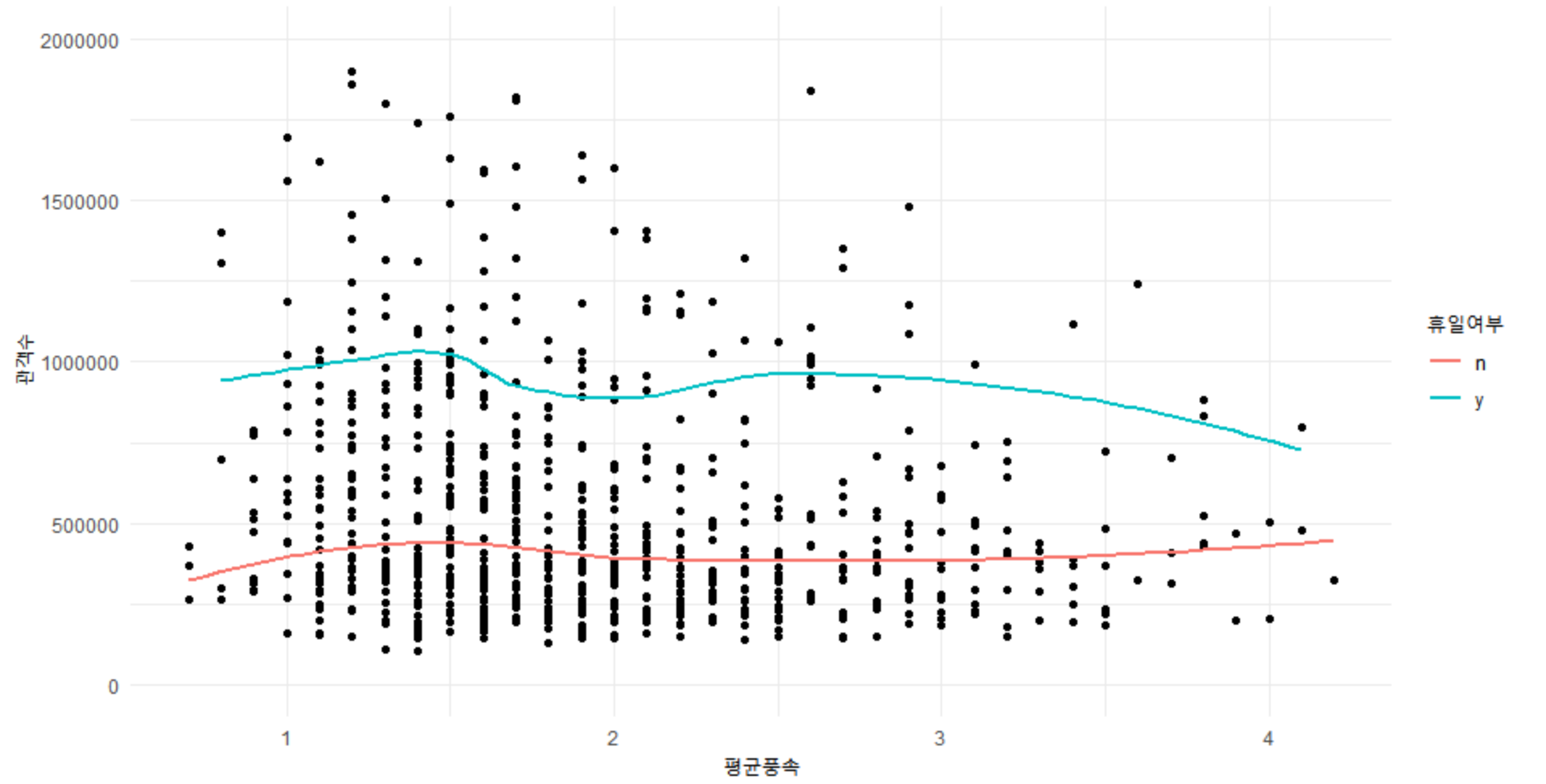
2. 활용분석기법

(1)그래프를 활용한 대략적 데이터 파악

* + 평균기온
* 날씨가 춥거나 더울수록 영화관 관객수가 늘어나는 양상
  + 날씨가 좋을수록 관객수가 늘어날 것으로 예상했으나 예상과 반대
  + 평범한 날씨에 비해 춥거나 더운 날 날씨를 피해 영화관에 가는 것으로 보임
  + 일강수량



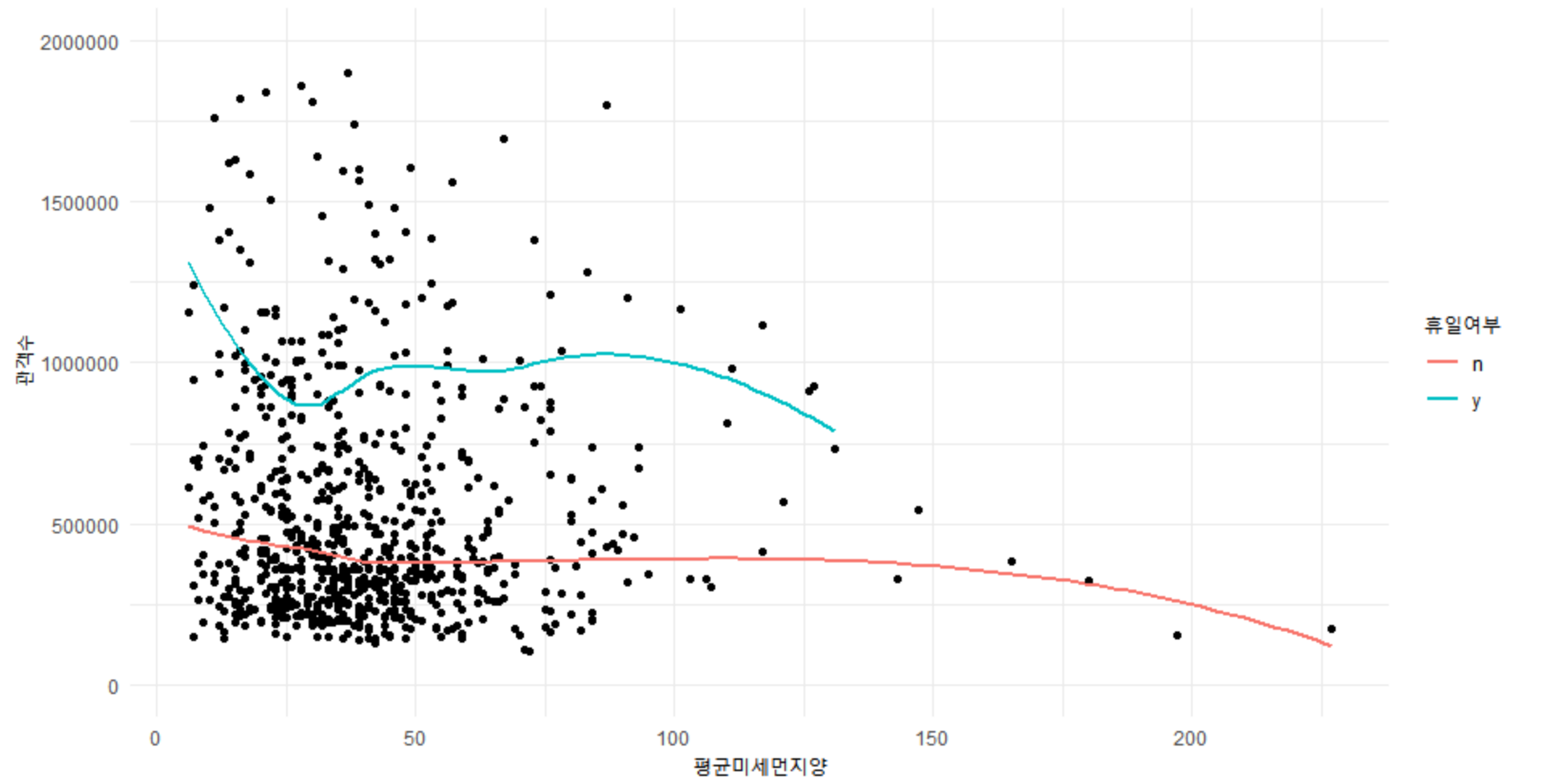
* 강수량이 0mm인 경우는 제외
* 강수량이 증가할수록 관객수가 줄어드는 양상
  + 강수가 있는 날은 외출을 꺼리기 때문으로 생각됨
  + 평균풍속



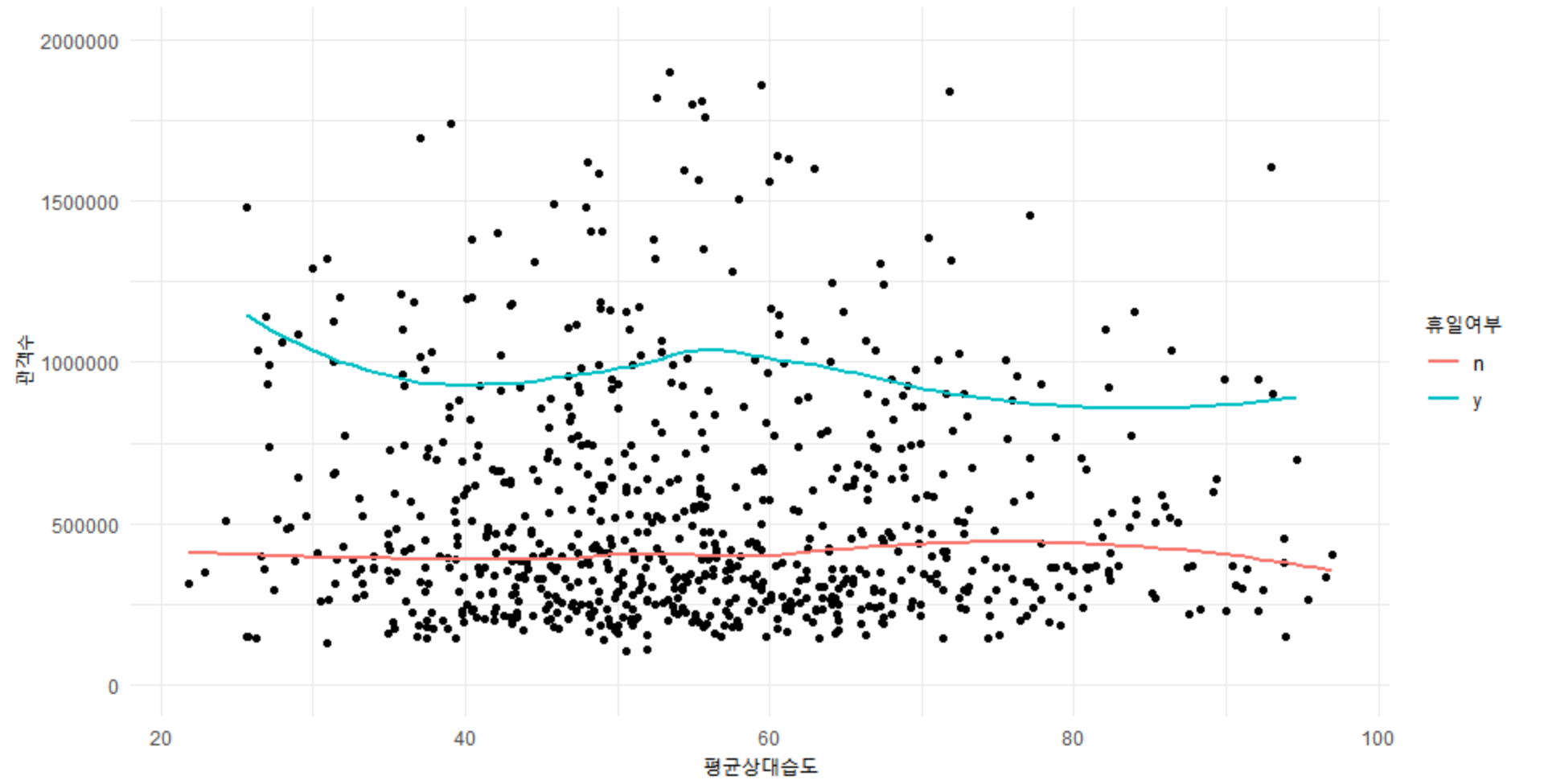
* 풍속이 증가할수록 관객수가 줄어드는 양상
  + 풍속이 증가할수록 관객 수가 늘어날 것으로 예상했으나 예상과 반대
* 휴일에는 풍속이 증가할수록 관객수가 줄어드는 반면 평일에는 증가하는 추세를 보임

4. 분석결과

* + 미세먼지 농도



* 미세먼지 농도가 증가할수록 관객수가 줄어드는 양상
* 미세먼지 농도가 증가할수록 외출을 꺼리기 때문으로 보임
  + 상대습도



* 상대습도가 낮거나 60%일 때와 높을 때 관객수가 늘어나는 양상
* 평일의 경우 습도와 관계없이 일정한 관객수를 유지하는 반면 공휴일의 경우 급격한 추세를 보임