

# 서울시 각 구별 미세먼지 분석 및 예측

김동률

# 목차

1. 주제 선정 동기
2. 데이터 출처
3. 각 구의 미세먼지 수치 시간대 비교
4. 2017년 1-9월 미세먼지 수치 예측하기
5. 결론

# 1. 주제 선정 동기

- ◆ 최근 미세먼지의 심각성에 대한 언론들의 보도와 대중들의 높은 관심때문에 미세먼지 분석을 주제로 선정



["문대통령님, 미세먼지 특단 대책 세워주세요!"](#) 데일리메디 | 4시간 전 | [🔗](#)

중국이 강한 의지로 **미세먼지** 문제 해결에 나서고 있다는 소식은 시사하는 바가 크다. 우리나라도 수도권 대기관리 특별법 등 적극적인 의지로 해결에 나선 적이 있지만 2014년 이후 **미세먼지** 감소 추세는 정체기다. 흥...



[\[기상특보\]기상청 오늘날씨 및 주간날씨 예보..태풍 마리아 중국서 소멸..미세...](#)

일간리더스경제 | 6시간 전 | [🔗](#)

오늘 **미세먼지** 농도는 대부분 '보통' 수준을 보이겠지만 울산은 대기정체로 대기오염물질이 축적되어 '나쁨... < 태풍 현황 > 제8호 태풍 '마리아(MARIA)'는 오늘 03시 현재 중국 푸저우 서북서쪽 약 470km 부근 육상에서...



["연 1만2천명 미세먼지로 조기 사망"...대통령 직속기구 촉구](#) JTBC PICK | 17시간 전

| 네이버뉴스 | [🔗](#)

환경재단은 중국이 시진핑 주석 주도로 최근 4년 간 주요 도시의 초**미세먼지**를 31% 줄인 점을 주목하고 있습니다. 우리도 대통령 직속 기구를 만들어야 한다며 오는 23일까지 보건 의료 전문가 300인 서명을 받아...

▶ [전문가들, 미세먼지 특단 대책 촉구...](#) 뉴시스 | 1일 전 | 네이버뉴스

▶ ["태아 사망까지 부르는 미세먼지, 대...](#) 연합뉴스 | 1일 전 | 네이버뉴스

## 2. 데이터 출처: 에어 코리아

([http://www.airkorea.or.kr/last\\_amb\\_hour\\_data](http://www.airkorea.or.kr/last_amb_hour_data))

2014 – 2017 분기별 csv파일 (17년은 3분기까지)  
분기 당 데이터 개수는 70만개 내외

**통계정보**

측정소별 확정자료  
측정망·항목별 확정자료  
대기환경 연월보  
**최종확정자료 다운로드**  
국외대기오염현황

**최종확정자료 다운로드**  
홈 > 통계정보 > 최종확정자료 다운로드

● 최종확정자료 다운로드

측정망 도시대기 지역 서울 측정소 전체  
조회기간 2017-09-28 ~ 2017-09-30 엑셀

- 국립환경과학원의 최종확정자료입니다.  
(확정이전 측정자료는 "실시간 자료조회" 메뉴에서 조회가능합니다.)
- PM<sub>2.5</sub> 측정자료는 2015.1.1부터 공개되었기 때문에 국립환경과학원 확정이 완료된 이후 공개를 예정입니다.
- 측정소 코드는 대기환경연보 [부록14. 대기오염 측정망 제원]을 참고하시기 바랍니다.

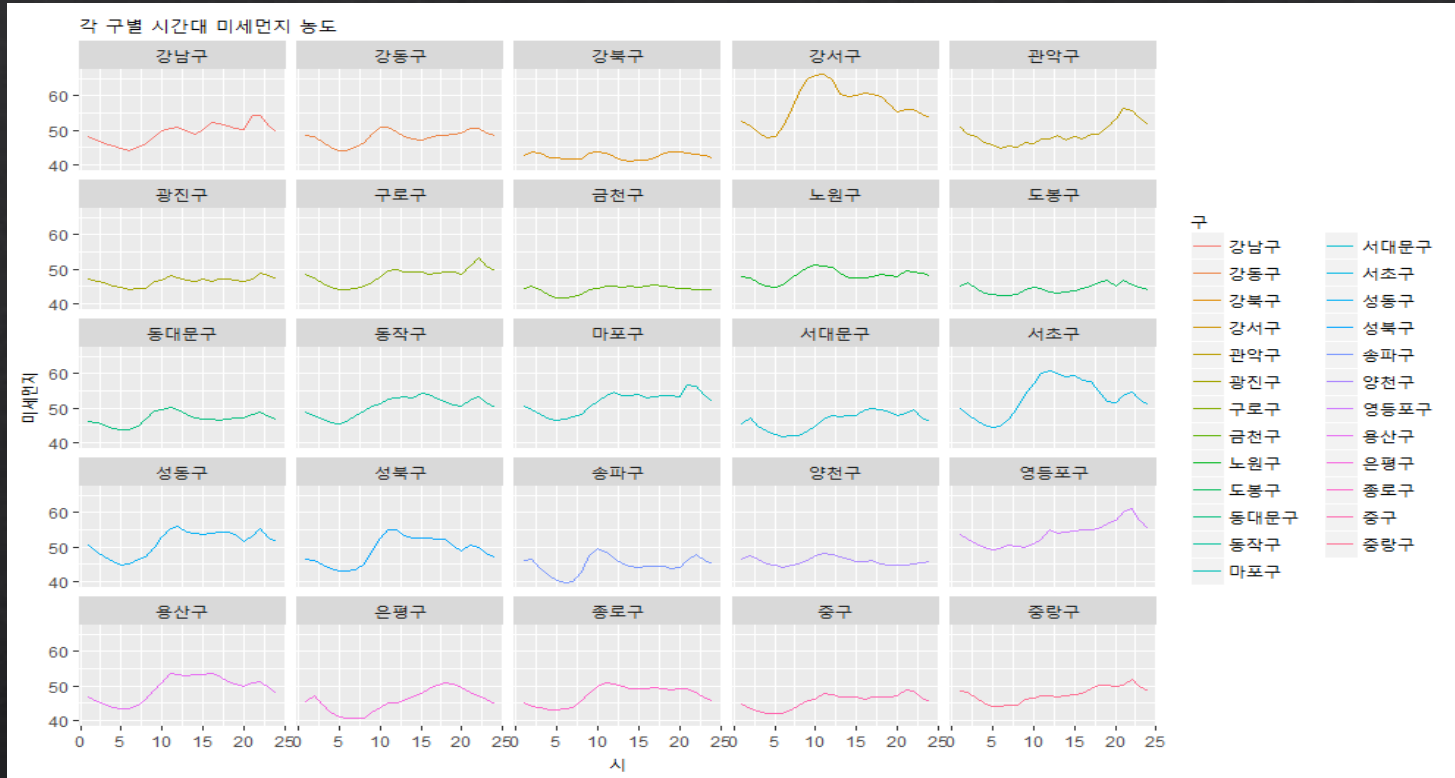
● 최종확정자료 연도별 다운로드

- 본 자료는 최종확정 자료이며, 장비점검, 통신장애 등 이상 데이터가 발생한 경우 "-999"로 표기됩니다.
- 최종확정 자료관련 문의 : 국립환경과학원 대기환경연구과 (032-560-7261)

2017년	2016년	2015년	2014년
-------	-------	-------	-------

- 2014년 1분기
- 2014년 2분기
- 2014년 3분기
- 2014년 4분기
- 2015년 1분기
- 2015년 2분기
- 2015년 3분기
- 2015년 4분기
- 2016년 1분기
- 2016년 2분기
- 2016년 3분기
- 2016년 4분기
- 2017년 1분기
- 2017년 2분기
- 2017년 3분기

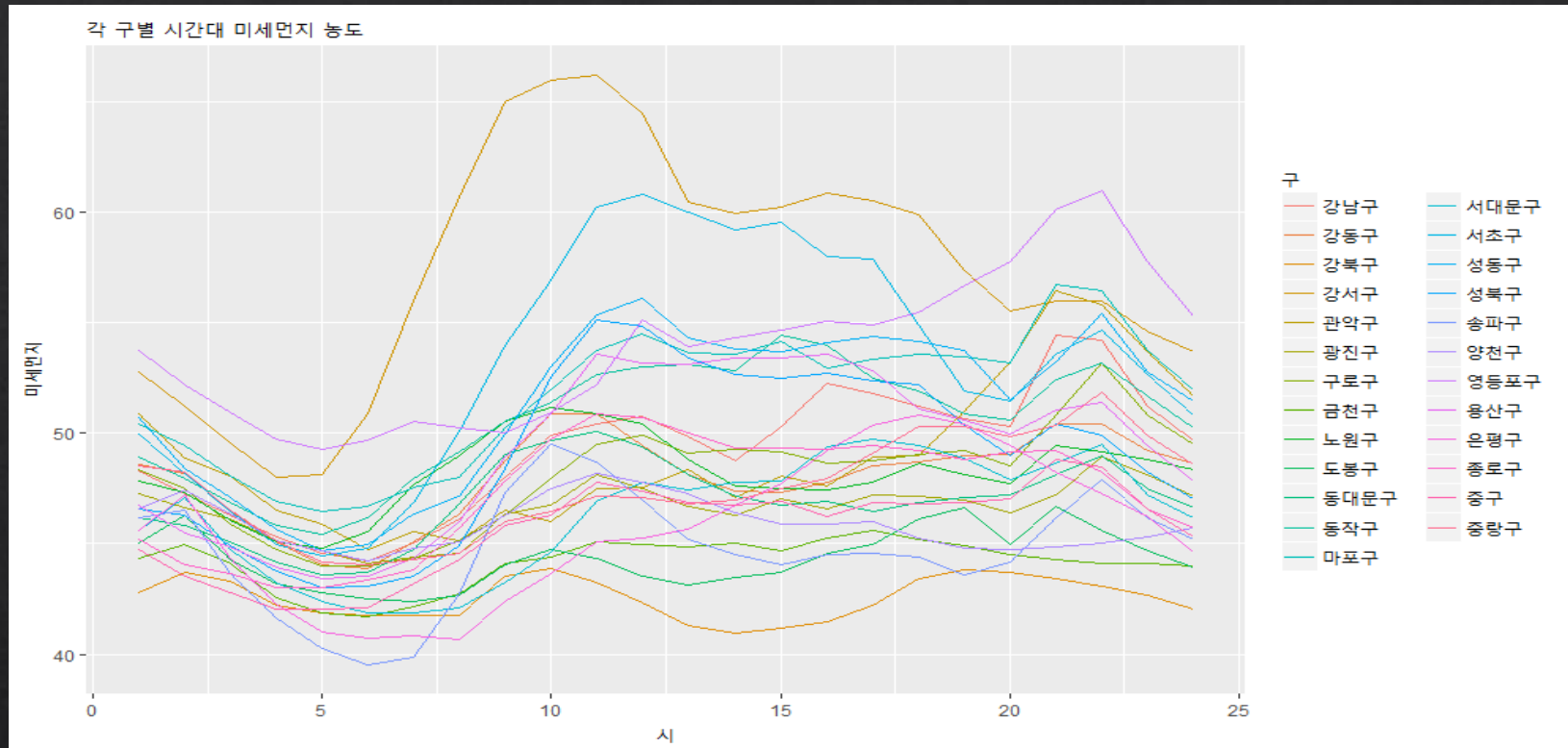
### 3. 각 구의 미세먼지 수치 시간대 비교



- ♦ 평균적인 미세먼지 수치가 가장 높은 구는 강서구(57.2). 영등포구(53.8), 서초구(53.1)가 그 뒤를 잇고 있음.
- ♦ 평균적인 미세먼지 수치가 가장 낮은 구는 강북구(42.6). 그 뒤로는 금천구(44.1), 도봉구(44.3), 송파구(44.7)가 비슷하게 수치를 이룸



### 3. 각 구의 미세먼지 수치 시간대 비교

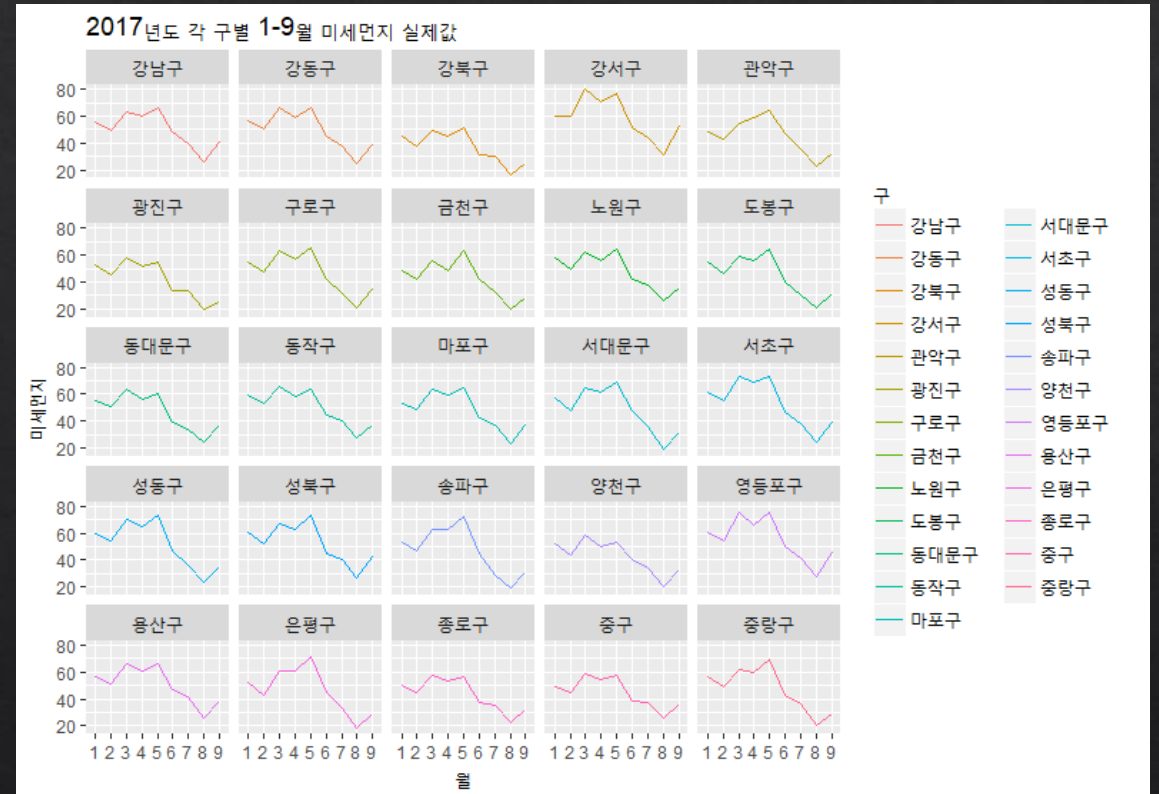
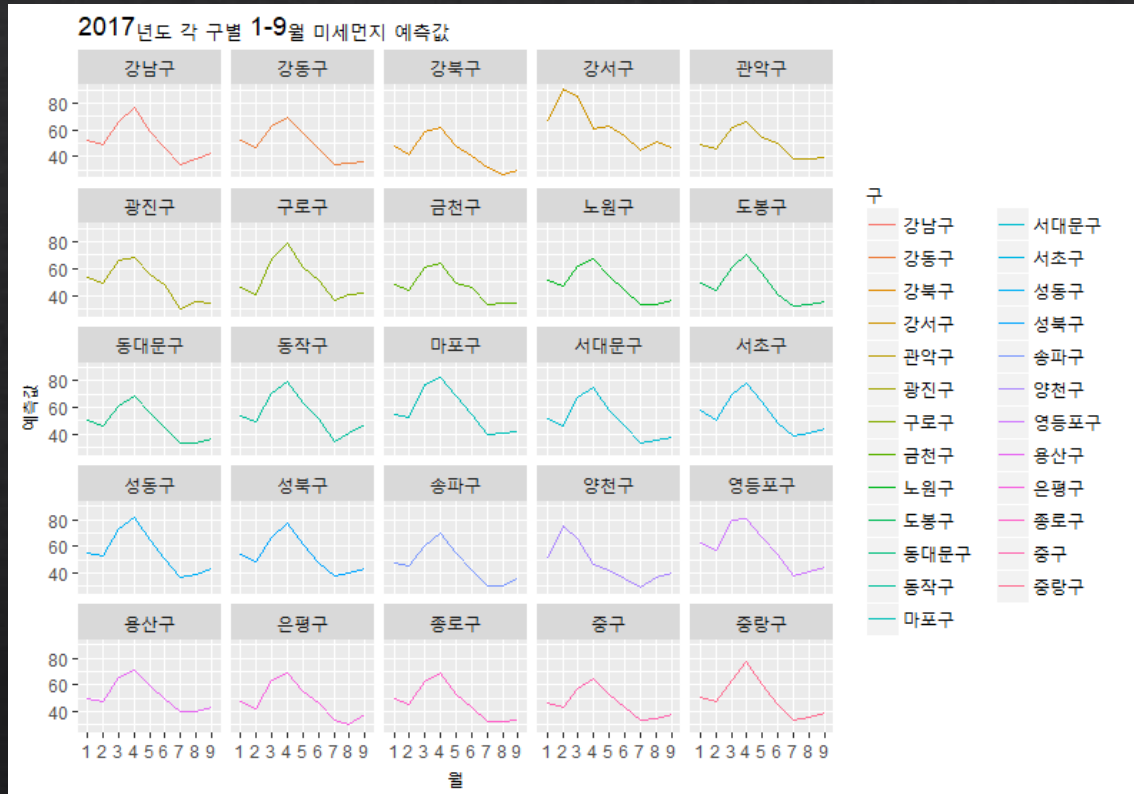


- ◆ 평균적인 미세먼지 수치가 가장 높은 시간은 11 – 12시와 21 – 22 시로 나타났고 평균 수치는 50을 기록.
- ◆ 평균적인 미세먼지 수치가 가장 낮은 시간은 새벽 4 – 7 시로 나타났고 평균 수치는 44를 기록.

## 4. 2017년 1-9월 미세먼지 수치 예측하기

- ◆ 먼저 Train data set과 Test data set을 설정
- ◆ Train data set : 2014 – 2016년 데이터
- ◆ Test data set : 2017년 데이터(1-3분기)
- ◆ 시계열 분석 모델 중 Arima model을 이용
- ◆ 시계열 평가 방법 중 하나인 auto.arima 함수를 사용해서 최적의 스펙을 결정
- ◆ 각 구의 시계열 데이터에 auto.arima 함수를 적용해서 각 구의 예측치를 구함

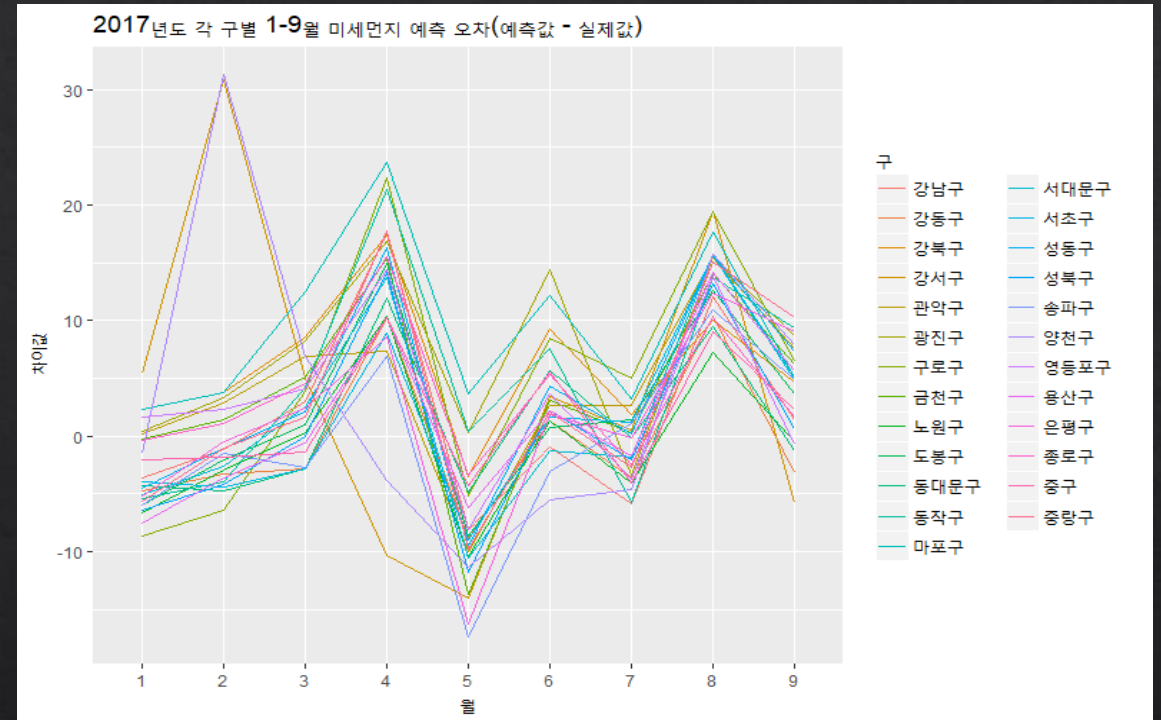
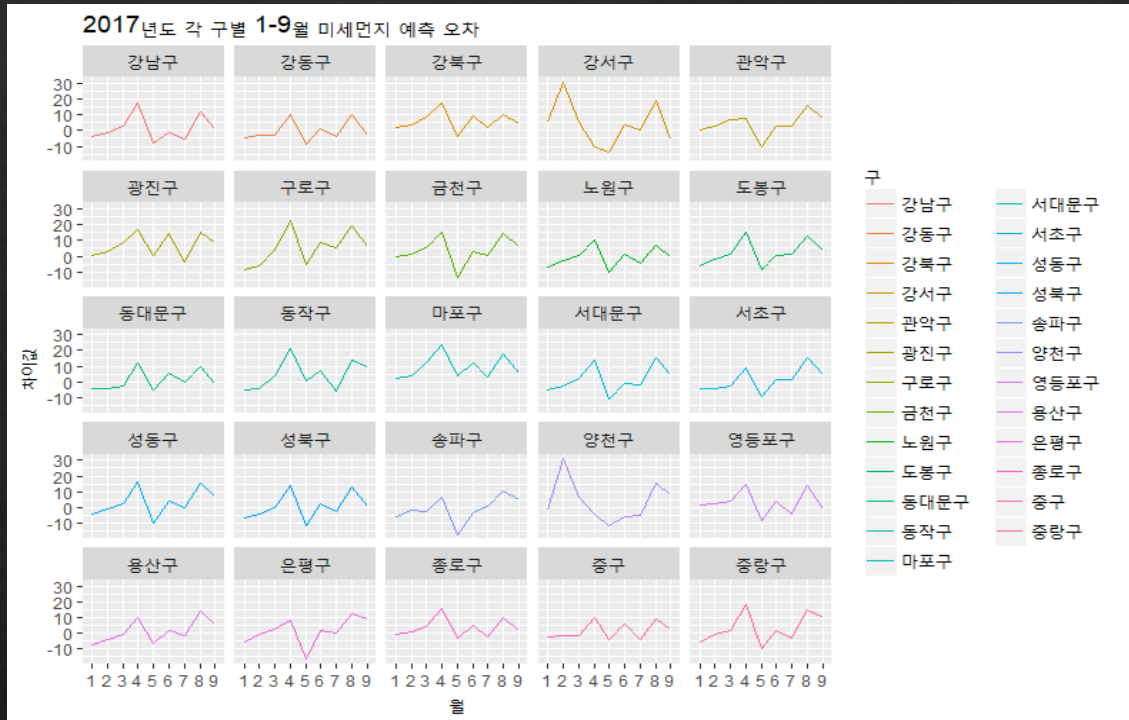
# 4. 2017년 1-9월 미세먼지 수치 예측하기



- ◇ 각 구의 2017년도 1 – 9월 미세먼지 예측값과 실제값 그래프
- ◇ 예측값 그래프의 추이가 실제값 그래프의 추이와 비슷한 걸 확인



# 4. 2017년 1-9월 미세먼지 수치 예측하기

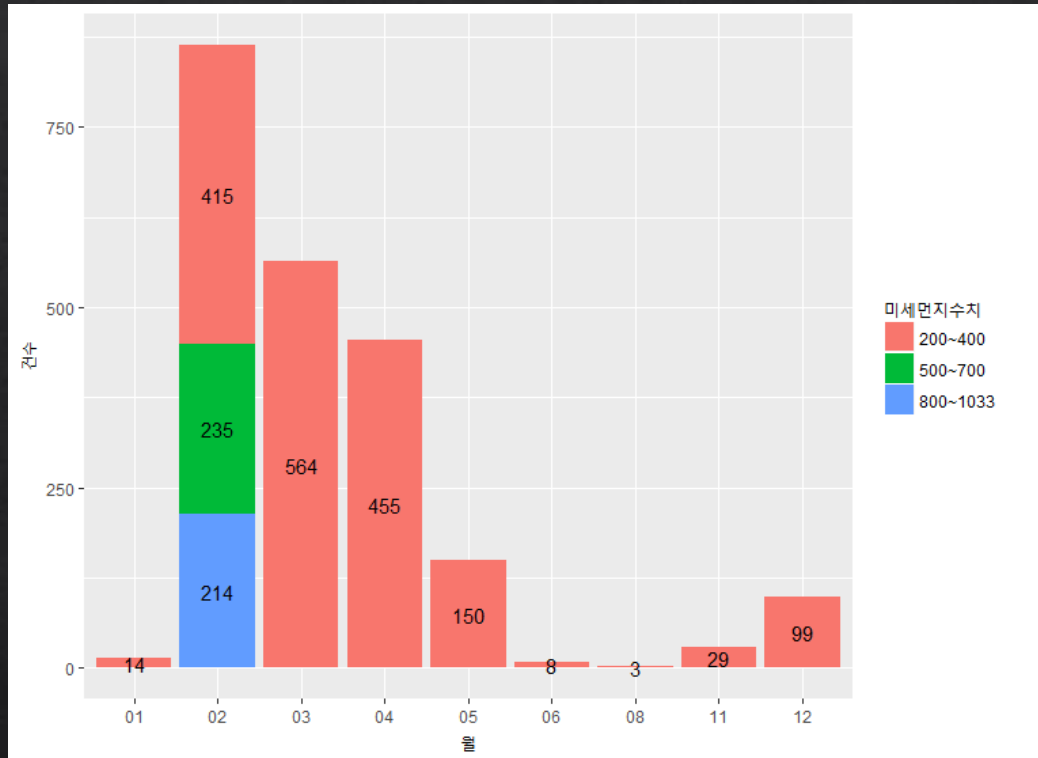


- ◇ 각 구별 예측 오차 그래프 (예측 오차 = 예측값 - 실제값)
- ◇ 0에 가까울수록 정확한 예측
- ◇ 오차가 가장 큰 달은 양천구 2월(31.27)이고 오차가 가장 적은 달은 성북구 3월(-0.08)로 확인
- ◇ 전체적으로 볼 때, 4,5,8월의 예측 오차가 높은 걸 확인

## 4. 2017년 1-9월 미세먼지 수치 예측하기

- ◆ 앞에 예측 오차를 살펴보면 전반적으로 4,5,8월의 오차가 큼
- ◆ 오차가 발생하는 원인은 raw data에 있을 것이라 생각
- ◆ 그래서 “4,5,8월 raw데이터 중에 미세먼지 수치가 비정상적으로 높은 이상값이 많이 있을 것이다.”란 가정을 하고 월별로 미세먼지 수치가 높은 건수를 분석함 (다음 페이지)

## 4. 2017년 1-9월 미세먼지 수치 예측하기



- ◇ 2014 – 2016 raw data를 분석한 결과 미세먼지 수치가 높은 데이터의 빈도가 높은 달은 02월이고 그 다음이 03월, 04로 나타남
- ◇ 각 구별로 분석한 결과도 전체를 분석한 결과와 거의 동일
- ◇ 따라서 raw data의 이상값이 Arima 모델에 영향을 미쳤을 거란 가정은 틀림

## 5. 결론

- ◆ 평균적으로 볼 때 미세먼지가 가장 높은 구는 강서구, 영등포구, 서초구
- ◆ 평균적으로 볼 때 미세먼지가 가장 낮은 구는 강북구, 금천구, 도봉구, 송파구
- ◆ 서울에서 미세먼지가 가장 높은 시간대는 오전 11 – 12, 오후 9 – 10 시
- ◆ 서울에서 미세먼지가 가장 낮은 시간대는 새벽 4 – 7시
- ◆ Arima 모델을 이용한 2017년 1- 9월 예측값이 실제값과 전반적으로 비슷
- ◆ 특정 달(4,5,8)의 오차가 심한 원인은 raw 데이터의 이상값이 아니라 Arima 모델의 정확성 때문