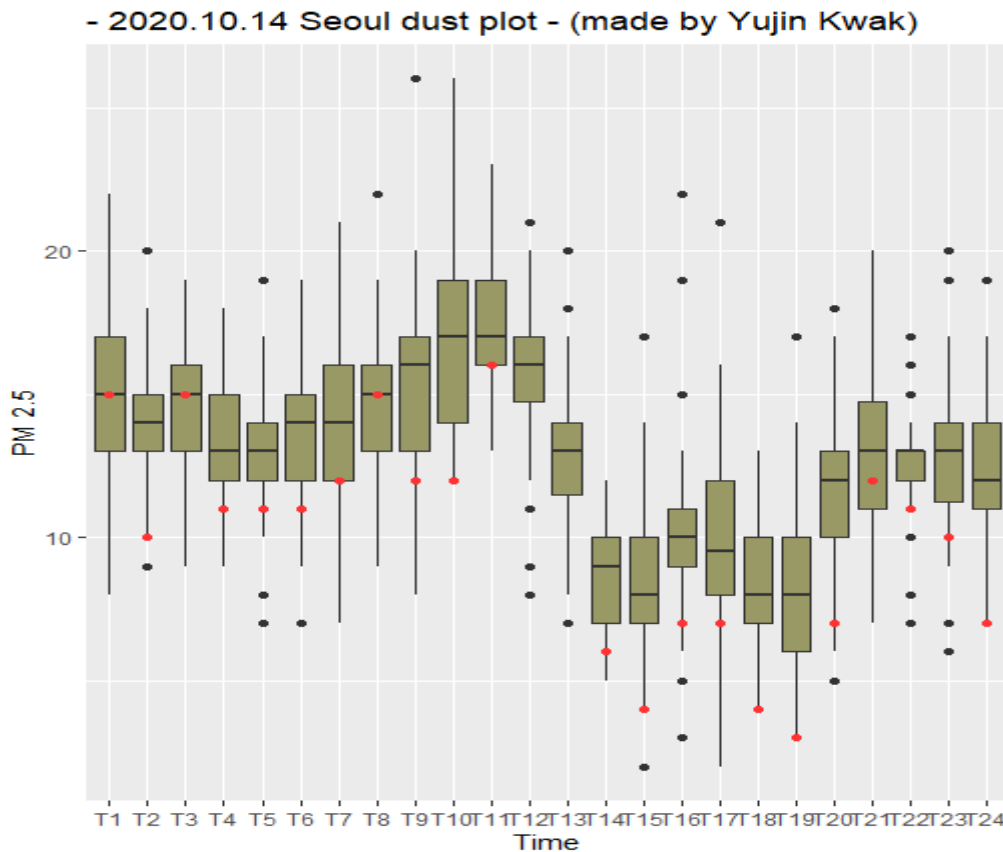


<환경생태 실습과제>

정보보안암호수학과

20172228

곽유진



[그래프 분석]

먼저 그래프를 분석해 보자면 검정색점 들이 각각 서울시에 있는 00 구들의 시간대별 초미세먼지 양을 의미하고 빨간색 점은 내가 살고 있는 서울시 성북구의 시간대별 미세먼지 농도를 표시한 것이다. boxplot으로 그린 box는 해당 열에 나열된 점(값)들을 4분위수를 이용해서 box로 표현한 것이다. 상자 밑면의 선은 1사분위수 즉 하위 25%위치의 값을 나타내고 상자안의 중간에 있는 선은 2사분위수는 하위 50%의 위치 값을(즉, 중앙값을 의미한다.) 그리고 상자 윗면의 선은 3사분위수인 하위 75%의 위치 값에 해당한다. 그렇다면 이제 굵은 선과 노란색 점이 의미하는 것은 무엇인지 알아보자. 검정색의 굵은 선은 플롯을 그릴 때 계산되는 범위에서의 값들에서 그 의미를 찾을 수 있다. 위로는 하위 75%-100%내에 해당하는 값을 의미하고, 아래로는 하위 0-25%내에 해당하는 값을 의미한다. 노란색점은 정상 범주에서 크게 벗어난 값(미세먼지 양이 정상 범주보다 많거나 적을 경우)을 의미한다.

[초미세먼지가 가장 심한 시간과 맑은 시간 비교]

다음 그래프는 서울의 10 월 14 일의 초미세먼지 그래프이다. 가장 심한 시간과 가장 맑은 시간은 일단 시간을 기준으로 비교를 할 것이기 때문에 x 축을 기준으로 y 값이 가장 높고, 낮은 경우를 찾아줄 것이다. 여기에서 우리는 보다 일관성 있는 정보를 얻기 위해 평균값을 이용해서 비교를 해볼 것이다. 일단 초미세먼지가 가장 심한 시간부터 알아보자. x 축(Time)을 기준으로 T10 과 T11 이 비슷하게 다른 그래프들의 평균값보다 가장높이 위치해 있다는 것을 알 수 있다. 실제로 값을 비교해보면 아주 근소한차이로 T11(11 시)의 평균값이 높다는 것을 R 을 통해 확인할 수 있었다. (이유는 boxplot 을 보면 박스가 T10 은 넓게 분포되어 있는데 이는 점이 퍼져 있다는 것을 의미하여 점들이 더 모여 있는 T11 의 평균이 더 높은 값을 가지는 것 같다.)

맑은 시간은 위와 같은 방법으로 평균값이 낮은 값을 찾아주면 된다. 낮은 값은 그래프만 보고 구분하기 힘들 정도로 T15, T18, T19 의 값이 매우 유사하게 보이는데 이것도 R 로 비교해보면 T19 가 가장 작은 값인 것을 비교 T19(19 시)의 값이 가장 작다는 것을 알 수 있었다.

[내가 사는 동네의 초미세먼지 농도 분석]

이제 그래프에서 붉은 점으로 표시한 내가 살고 있는 서울시 성북구의 미세먼지 농도를 알아보자. 12 시와 13 시에는 데이터 값이 찍히지 않은 것을 보니 측정이 불가능했다는 것을 알 수 있다. 또한 하루 동안 서울시 성북구의 미세먼지 농도가 가장 낮을 때가 19 시이고 높을 때가 11 시 인 것을 알 수 있다. 이 데이터를 보기 전까지는 사람들이 출퇴근하는 시간인 8-9 시과 6-7 시쯤이 초 미세먼지 값이 가장 높게 나올 것이라 예측했다. 하지만 데이터 분석 결과 예상과 반대로 나왔다는 것이 조금은 놀라웠다. 사람들이 퇴근하는 시간대인 19 시에 초미세먼지가 적었던 이유는 10 월 14 일날 사람들이 적게 활동하고 있었다는 그 날의 특성이 반영된 것이 아닐까 조심스럽게 예측해보았다.