R을 이용한 광역시의 대기오염물질에 대한 데이터 시각화

이나영, 문봉교
동국대학교 컴퓨터공학과
e-mail : bt0002@dongguk.edu, bkmoon@dongguk.edu

R-based Data Visualization on Atmosphere Pollution Matters in Metropolitan Cities

NaYoung Lee, Bongkyo Moon
Dept. of Computer Engineering and Science, Dongguk Univ.-Seoul

요 약

1995년부터 2014년까지 약 20년간의 기상청의 대기오염 자료를 바탕으로 이를 3가지의 다른 방법으로 시각화 하였다. 첫 번째는 2014년 주요 광역시의 연중 월별로 대기 오염 물질들을 시각화 하였고, 두 번째는 1995년부터 2014년까지 주요 광역시의 연도별로 대기 오염 물질들을 시각화하였다. 마지막으로 대한민국 지도 위에 2014년 주요 광역시의 대기 오염 물질들을 점 모양으로 시각화하였다. 자료 시각화에는 R의 기본 내장 시각화 함수 plot()과 시각화 패키지 ggplot2, ggmap를 사용하였다.

1. 서론

AT&T에서 개발된 통계 언어인 S에서 영향을 받은 R은 통계 데이터 수정과 분석을 위한 스크립트언어이다. 빅데이터가 시대의 화두가 되면서 그 동안 어렵게만 여겨졌던 데이터 분석이 특정 사업영역에서 벗어나게 되면서, 단순한 데이터 처리와 활용에 강점을 가진 통계분석 언어인 R의 위상은 점점 커지고 있다.

또한, R에는 이미 통계관련 대부분의 공식이 모두 포함되어 있고, 그래프도 쉽고 빠르게 그릴 수 있다. 인터프리터 언어인 R은 오픈 소스라서 그 확장성이 매우 뛰어나며, 아직 데이터에 대한 결과를 알지 못하는 연구자들이 사용하기에 아주 적합하다.

한편, 최근 정부에서는 국민들에게 공공데이터를 개방하고 있다. 하지만, 이러한 방대한 데이터를 한 눈에 보고 이해할 수 있도록 표현하지 못한다면, 방대한 데이터는 그저 숫자들의 열거에 지나지 않을 것이다. 최근 국내 대도시들의 대기환경이 악화됨에 따라 정부에서는 다양한 그 해결책을 제시하고 있는데, 문제의 원인이 무엇인지 정확히 파악함으로써 근본적인 대책이 마련될 수 있을 것이다.

국내 대도시의 대기오염 악화의 원인을 규명하고 그 대책을 마련하는 것이 중요한 이슈로 대두되고 있다. 따라서 본 논문에서는 지난 20년간 기상청에서 관측한 대기오염 자료와 통계분석언어인 R을 이용하여 국내 대도시의 대기오염 변화에 대한 자료를 시각화 하였다. 이러한 시각화 및 자료분석을 통해 대도시 대기오염 악화의 원인과 문제해결의 실마리를 찾아보고자 한다.

2. 국내 대기질 시각화 및 분석

본 논문에서는 기상청 에어코리아 홈페이지에서 다운받은 전처리 자료를 사용하여 시각화를 수행하였다. 이 자료는 대한민국 대기질에 관한 자료를 포함하고 있는데, 특히, 대표적인 대기오염 물질인 이산화황(SO_2),이산화질소(NO_2),오존(O_3),미세먼지(PM_{10}), 일산화탄소(CO)의 자료가 포함되어 있다. 일부 유효측정 처리비율이 75% 미만인 관측치는 결측치로 처리하였다.

꺾은선 그래프에서 서울은 검은색(black), 부산은 파란색(blue), 대구는 빨강(red), 인천은 초록(green), 광주는 하늘색(cyan), 대전은 주황색(orange), 울산은 보라색(purple) 선으로 나타내었다.

2.1. 2014년 전국 7대 광역시의 연중 월별 대기오염물질 시각화 - 계절적 특징을 중심으로

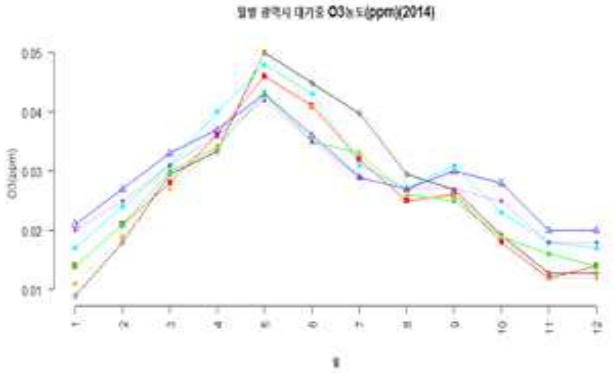
대기중의 오염물질들은 기상적인 요인(바람, 강우량, 기온 등)에 영향을 받는 것으로 보인다. 오존(O_3)를 제외한 대부분의 오염물질이 여름에 감소하고 겨울에 증가하는 특징을 보였다. 이는 주로 비나 눈에 의한 침착 때문으로 보인다.

그림1에서 전국적으로 5월을 정점으로 고온다습한 북태평양 기단의 영향을 받아 오존(O_3)농도가 감소하고 있다. 오존(O_3)는 NO_2가 태양광선에 의해서 반응하여 생성되는 2차 오염물질이다. 보통은 태양광선이 강하고 기온이 높으며 건조한 날에 높은 농도를 보이는 것이 특징이다. 따라서 상대적으로 기온이 높고 건조한 시기인 5월(7개 광역시 평균 0.044714, 단위는 ppm)에 가장 농도가 높은 것으로 보인다.

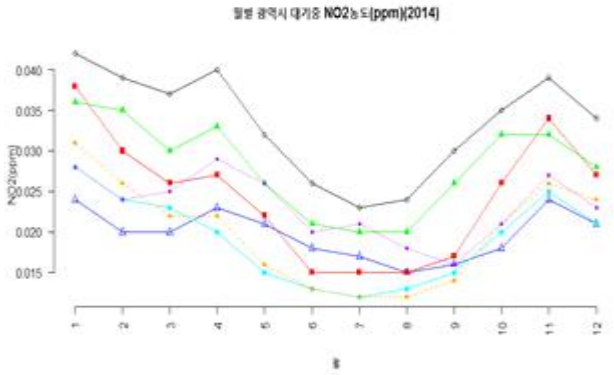
그림 2에서 이산화질소(NO_2)는 오존(O_3)와는 반대로 V자 형태의 그래프를 가진다. 최저점은 8월(7개 광역시 평균

0.016714, 단위는 ppm)이다. 이것은 이산화질소(NO_2)가 오존(O_3)의 전구물질이기 때문이다.

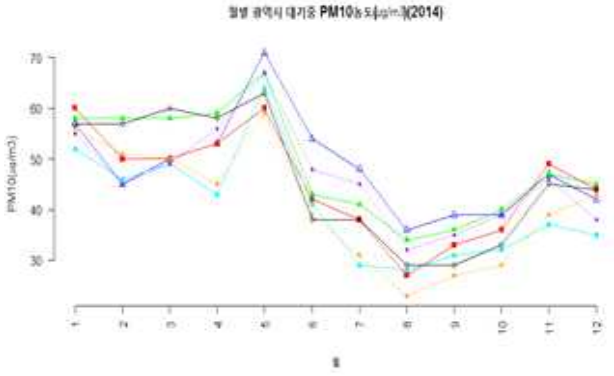
그림 3에서 미세먼지(PM_{10})은 대체로 겨울(1월~2월)과 봄(3월~5월)에 가장 높은 농도를 유지한다. 이것은 봄에 중국 고비사막에서 우리나라로 미세먼지를 실은 바람이 불기 때문이라고 생각된다. 여름(6월~8월)에는 비나 습기 때문에 감소하는 모양새를 보인다. 겨울(12월~2월)에는 대기가 안정되어 있는 날이 많고 중국발 스모그 때문에 높은 농도를 유지하는 것으로 추정된다.



(그림 1) 월별 광역시 대기중 오존(O_3)농도(ppm) (2014)



(그림 2) 월별 광역시 대기중 이산화질소(NO_2)농도(ppm) (2014)



(그림 3) 월별 광역시 대기중 미세먼지(PM_{10})농도($\mu g/m^3$)(2014)

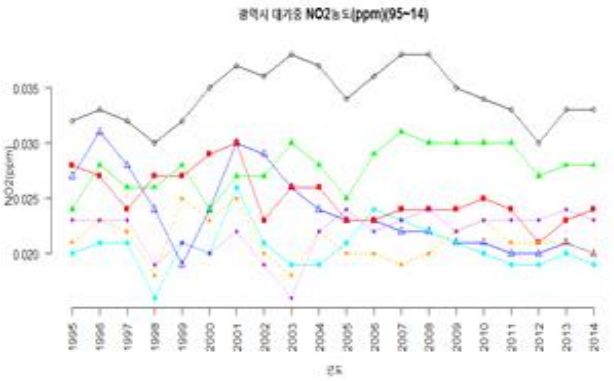
2.2. 1995년~2014년 광역시의 연별 대기오염물질 시각화

5개 대기오염물질 중 대부분이 1995년~2014년 동안 감소하는 모양이었다. 이는 경제발전에 따른 국민들의 환경에 대한 관심이 높아짐과 배출오염물질의 처리기술 발달 그리고 정부의 지속적인 환경규제정책 때문이라고 생각된다.

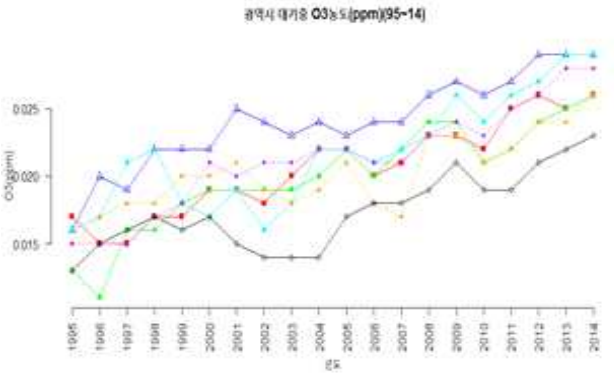
그림 4에서 이산화질소(NO_2)는 전국 7개 광역시(서울, 부산, 대구, 인천, 광주, 대전, 울산)에서 뚜렷한 증가도 감소도 보이지 않는다. 이는 경제발전 에 따른 전국 자동차 등록수의 증가와 관련 있어 보인다.

그림 5에서 다른 대기 오염물질들과 다르게 오존(O_3)은 꾸준하게 증가하고 있다. 이것은 오존 생성에 영향을 끼치는 휘발성 유기화합물 (VOCs)의 배출량이 계속해서 증가하는 것과 연관이 있다고 추측할 수 있다.

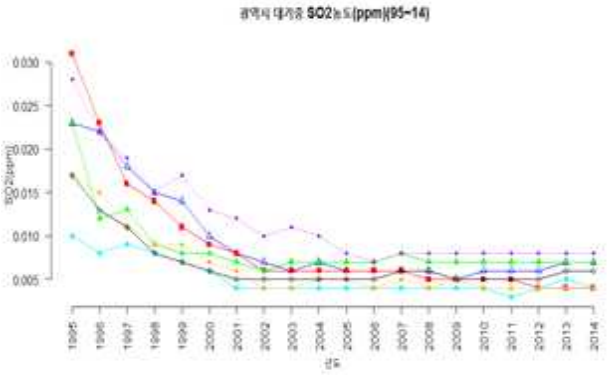
그림 6에서 이산화황(SO_2)의 경우 감소하다가 유지하는 모양을 보이고 있다. 감소요인은 지속적인 정부의 연료규제정책으로 추측된다. 연료규제정책이 경유와 중유의 황함유량을 규제함으로써 근본적인 이산화황(SO_2)발생을 억제하고 있다고 볼 수 있다.



(그림 4) 광역시 대기중 이산화질소(NO_2)농도(ppm)(95~2014)



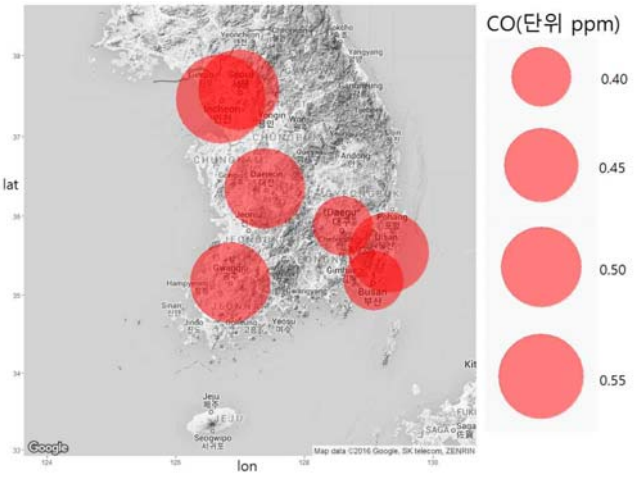
(그림 5) 광역시 대기중 오존(O_3)농도(ppm)(95~2014)



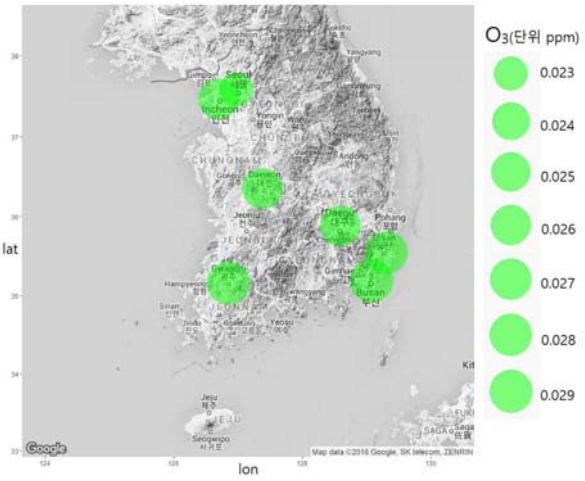
(그림 6) 광역시 대기중 SO_2농도(ppm)(95~2014)

2.3. 지도에서 2014년 전국 7대 광역시의 대기오염물질 시각화 - 지역적 특징을 중심으로

그림 7에서는 인천이 일산화탄소(CO) 농도가 가장 높게(인천 0.6, 단위는 ppm) 측정되었으나 이동오염원(자동차 등)에 의한 배출량이 많은 일산화탄소(CO)의 특성상 대도시 7개 모두 높은 농도(서울 0.5, 부산 0.4, 대구 0.4, 광주 0.5, 대전 0.5, 울산0.5, 평균0.485714, 표준편차 0.069007, 단위는 ppm)가 측정되었다.



(그림 7) 광역시 대기중 일산화탄소(CO)농도(ppm)(2014)



(그림 8) 광역시 대기중 오존(O_3)농도(ppm)(2014)

그림 8에서 오존(O_3)는 부산(0.029, 단위는 ppm)과 광주(광주 0.029, 단위는 ppm)에서 가장 큰 수치를 보였으나 7개 광역시 모두 큰 편차 없는 수치(서울 0.023, 인천 0.026, 대구 0.026, 대전 0.026, 울산0.028, 평균 0.027, 표준편차 0.002138, 단위는 ppm)를 보였다. 오존(O_3)는 이차오염물질의 특성상 배출원의 영향보다는 기상의 영향을 많이 받아 편차가 작다.

3. 결론

대용량 데이터가 쏟아져 나오는 오늘날, 자칫하면 의미없는 숫자가 될 수 있는 자료들을 통계분석과 시각화에 뛰어난 R을 이용하여 시각화 해보았다. 대한민국 7개 광역시(서울, 부산, 대구, 인천, 광주, 대전, 울산)들의 5가지 주요 대기오염물질(이산화황 SO_2,이산화질소 NO_2, 일산화탄소 CO, 미세먼지 PM_{10}, 오존 O_3)에 대해서 한눈에 알아 볼 수 있게 시각화 하였다. 그리고 시각화된 자료에 대해서 여러 가지 측면에서 해석을 해 보았다. 이러한 강력한 R의 자료 시각화 기능을 잘 사용하면 쉴 새 없이 쏟아져 나오는 데이터를 유의미하게 활용 가능할 것이라고 판단된다.

Acknowledgement

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 서울어코드활성화지원사업(IITP-2016-R0613-16-1147)의 연구결과로 수행되었음.

참고문헌

[1] 서진수, “(다양한 예제로 쉽게 배우는 R)R까기(2014)”, 느린생각

[2] 박찬성, “(통계와 R을 함께 배우는)R까기. 2, R 입문용(2016)”, 느린생각

[3] 양경숙, “R 입문 및 기초 프로그래밍(2011)”, 자유아카데미

[4] 양경숙, “(기초 자료 분석을 위한)R 입문(2011)”, 한나래

[5] 기상청 2014년 연보 자료 (http://www.kma.go.kr/weather/climate/data_sfc_ann_mon.jsp)

[6] 김현철, 이부용, 조완근, “강수시 대기오염 농도변화에 대해서(2005)”, 한국환경과학회 봄 학술발표회지 제14권(제1호), 215~217

[7] Milton, Michael, ”Head first data analysis(2013)”, 한빛미디어

[8]백영민, ”R를 이용한 사회과학데이터 분석: 기초편(2015)”, 커뮤니케이션북스