

Re.SEOUL

참가자 : 채무진, 김현민, 편승엽

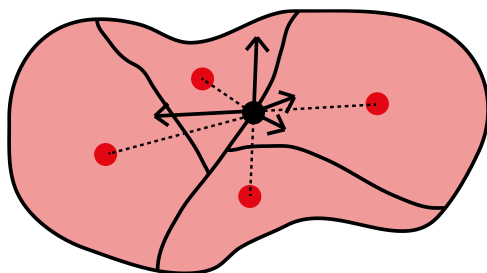
1. Motivation

- 서울시 공공데이터 공개 : 2009년 오바마 대통령 취임과 동시에 정부부처와 및 공공기관이 참여하는 'Open Government Initiative' 를 마련하고, 공공데이터를 전부 공개하고 있다. 이에 발맞춰 서울시에서도 작년부터 공공데이터를 공개하고 사람들의 이용을 독려하고 있다.

- Cartogram 기법 적용 : 데이터를 독자가 직접 분석하기는 쉽지 않다. 따라서 이러한 데이터의 시각화 필요성이 크게 대두되고 있다. 본 프로젝트에서는 Cartogram 기법을 이용하여 서울시에서 생산되고 있는 지도기반 데이터를 직관적이고 쉽게 접할 수 있도록 하려고 한다.

2. Algorithm

- 본 프로젝트에 사용된 'Rubber Sheet Distortion Method Cartogram Algorithm'[1] 은 Cartogram을 그릴 때 가장 효과적인 알고리즘 중 하나이다. 각 구역의 경계에 점들과 그 구역의 무게 중심 사이의 각을 가지고, 구역이 변해야 할 넓이에 비례하는 힘을 가지는 벡터를 만든다.



● 다각형의 무게중심 ● 임의의 경계점 i

i) $d_{ij} < p_j$ 일 때,

$$F_{ij} = (p_j - q_j) (p_j / d_{ij})$$

ii) $d_{ij} > p_j$ 일때,

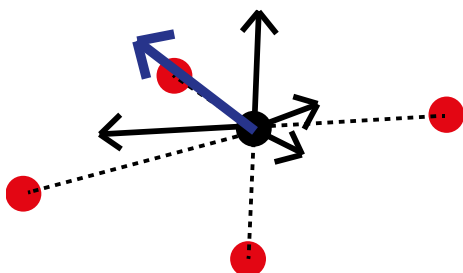
$$F_{ij} = (p_j - q_j) ((4p_j - 3d_{ij}) / p_j) (d_{2ij} / p_{2j})$$

F_{ij} = 다각형 j 와 점 i 사이의 힘

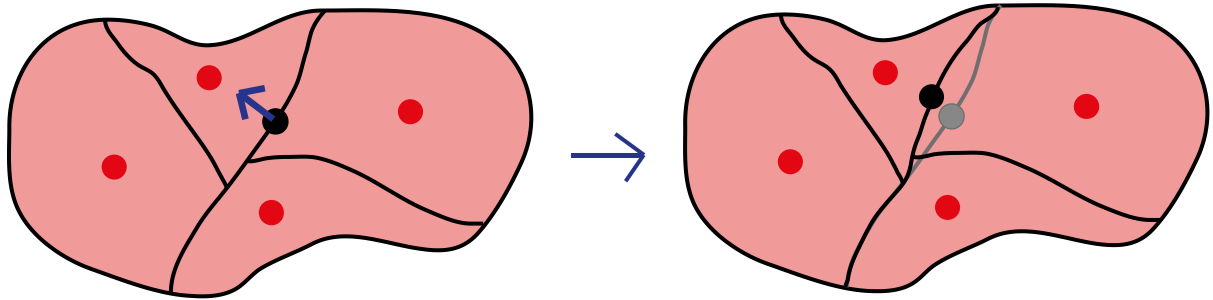
p_j = 현재 다각형과 넓이와 같은 원의 반지름

q_j = 변해야 할 다각형의 넓이와 같은 원의 반지름

d_{ij} = 점 i 에서 다각형 j 의 무게중심까지의 거리



각 경계의 점들은 구역의 갯수만큼의 힘 벡터를 가지게 되고 힘의 합성을 통하여 움직일 점을 정한다. (d_{ij} , p_j 의 대소에 따라 힘을 다르게 하는 이유는 다각형의 무게중심과 경계점이 너무 가까우면 힘의 크기가 너무 커지기 때문이다.) 위 그림2에서 파란색 화살표가 합성된 힘이다. 합성된 힘을 Normalize 하여 각 점의 힘벡터의 평형점을 찾아간다.



위와 같은 과정을 각 다각형의 현재 넓이와 변해야 할 넓이의 차가 정해진 값보다 작아 질 때 까지 반복한다.

3. Experimental Process & Result

- 개발 환경 : HTML5, Javascript, Canvas를 이용하여 웹기반으로 개발 하였다.
- 지도 데이터 수집 : 통계청에서 운영하는 통계지리정보서비스(<http://sgis.kostat.co.kr>) 에서 서울의 각 지역구별 경계에 해당하는 데이터를 받아온다. 홈페이지에 API 호출을 하면 GeoJSON(<http://geojson.org> 참고) 형태로데이터를 넘겨준다.
- Data 선정 및 수집 : 서울시에서 운영하는 서울 열린 데이터 광장(<http://data.seoul.go.kr>) 에서 아래의 세가지 데이터를 구별로 수집했다.
- 결과분석 : 각 데이터 마다의 표준편차를 구해보고 iteration을 얼마나 하면 평형상태에 도달하는지 측정했다. 각 구별 현재 넓이와 변해야 할 넓이 차의 합이 평균 5 이하로 떨어지면 평형상태에 도달 했다고 가정했다. (실제 이정도 오면 화면상 1px의 크기보다 작은 만큼 움직인다.) 각 데이터별 차이가 있지만 평균적으로 30~60 iteration 을 수행하면 평형상태에 도달 한다는 것을 알 수 있다.

구 분	기초생활 수급자	의료 인력수	교육재정 지원액
데이터 표준편차	2326.86	3743.77	2581.25
Iteration 횟수	61	60	45

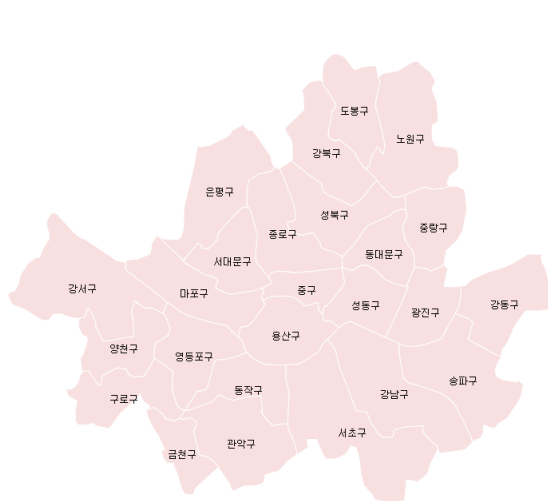


그림1. 기본 서울지도

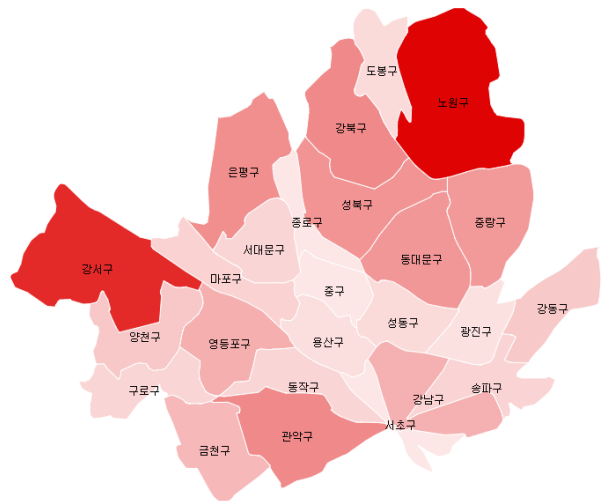


그림2. 기초생활 수급자수 적용

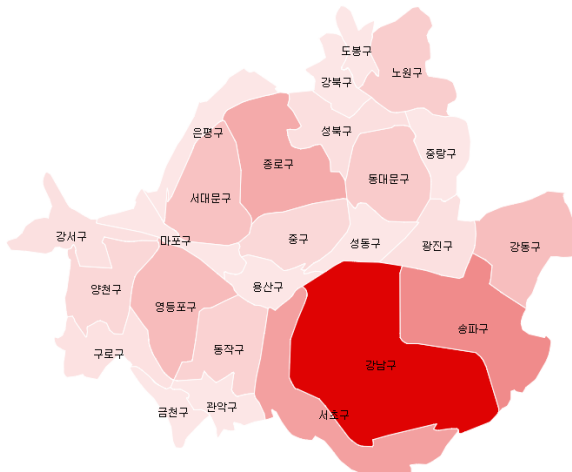


그림3. 의료 인력수 적용

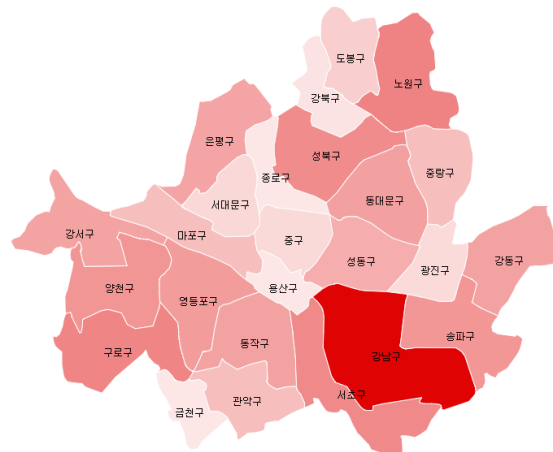


그림4. 교육재정 지원액 적용

그림2, 3, 4을 보면 데이터가 각 구별로 편중되어 있는지 골고루 분포하는지 알 수 있다. 상대적으로 교육재정 지원액은 구별로 골고루 나누어져 있고, 의료 인력수와 기초생활 수급자수는 구별 편차가 크다는 것을 한눈에 알 수 있다.

4. Expected Effect

기존의 그래프로 표현 되는 경우 사용자들은 머릿속에서 한번의 연산을 수행해야 한다. 하지만 Cartogram기법을 이용하여 데이터를 시각화 한다면 좀 더 직관적으로 쉽게 받아들일 수 있다. 시민들은 서울시에서 제공하는 데이터를 이용한 결과를 살펴보면서 여러 사회현상에 대한 문제점을 한눈에 인식하고 개선하기 위한 노력을 이끌 수 있다.