



E  
P  
I  
D  
E  
M  
I  
O  
L  
O  
G  
Y

# R을 이용한 시공간 역학

2015년 6월 26일(금)

황 승 식



역학은 인구집단의 건강을 증진시키기 위해 여러 집단을 비교하는 방식을 활용하여 인구집단 차원의 질병 및 건강상태의 분포와 결정요인을 탐구하는 학문 분야이다.

역학과 철학 분야 전공자가 책을 옮겼지만 원저의 난이도가 매우 높은 관계로 곳곳에 의도하지 않은 오역이 있을 가능성이 다분하다. 눈 밝은 독자들이 읽고 옮긴이나 출판사에 알려주기 바란다. 오는 봄에는 『역학의 철학』을 알리기 위해 분주했으면 좋겠다.

2015년 3월 옮긴이를 대표해 황승식 씀.

# 차례

1. 시공간 역학 개요
2. 시공간 역학 연구 사례
3. R을 이용한 시공간 역학 분석
  1. R을 이용한 지도화
  2. R을 이용한 시공간 군집 탐색
  3. R을 이용한 시공간 회귀 분석

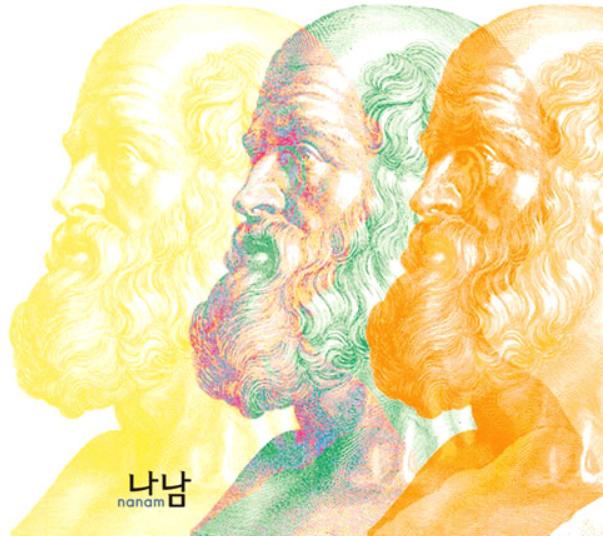
# 1. 시공간 역학 개요

**THE OATH OF HIPPOCRATES** I SWEAR by Apollo the physician and Asclepius, and Health, and All-heal, and all the gods and goddesses, that, according to my ability and judgment, I will keep the Oath and this stipulation — to reckon him who taught me this Art equally dear to me as my parents, to share my substance with him, and relieve his necessities if required; to look upon his offspring in the same footing as my own brothers, and to teach them this art, if they shall wish to learn it, without fee or stipulation; and that by precept, lecture, and every other mode of instruction, I will impart a knowledge of the Art to my own sons, and those of my teachers, and to disciples bound by a stipulation and oath according to the law of medicine, but to none others. I will follow that system of regimen which, according to my ability and judgement, I consider for the benefit of my patients, and abstain from whatever is deleterious and mischievous. I will give no deadly medicine to any one if asked, nor suggest any such counsel; and in like manner I will not give to a woman a pessary to produce abortion. With purity and holiness I will pass my life and practice my Art. I will not cut persons labouring under the stone, but will leave this to be done by men who are practitioners of this work. Into whatever houses I enter, I will go into them for the benefit of the sick, and will abstain from every voluntary act of mischief and corruption; and, further, from the seduction of females or males, of freemen and slaves. Whatever, in connection with my professional service, or not in connection with it, I see or hear, in the life of men, which ought not to be spoken of abroad, I will not divulge, as reckoning that all such should be kept secret. While I continue to keep this Oath unviolated, may it be granted to me to enjoy life and the practice of the art, respected by all men, in all times. But should I trespass and violate this Oath, may the reverse be my lot.

*Corpus Hippocraticum*

# 히포크라테스 선집

히포크라테스 지음 | 여인석 · 이기백 옮김



“의술을 올바로 추구하고자 하는 자는 다음과 같이 행해야 한다. 먼저 한 해의 계절들 각각이 무슨 영향을 미칠 수 있는지를 고려해야 한다. (중략) 또한 찬 바람과 더운 바람, 특히 모든 사람에게 공통적인 바람과 각 지방에 특유한 바람도 고려해야 한다. 물들의 성질에 대해서도 고려해야 한다. (중략) 그래서 어떤 의사가 낯선 나라에 도착하면, 그는 그 곳의 위치에, 다시 말해 그 곳이 바람과 태양의 떠오름에 대해 어떻게 위치하고 있는지에 주의를 기울여야 한다. (하락)”

- 히포크라테스, '공기, 물, 장소에 관하여' 중.

2.45 meters (8 feet 5 inches)



3.20 meters (10 feet 6 inches)

7/5/15 *Lienzo de Quauhquechollan* ca. 1530

Wiley Series in Probability and Statistics

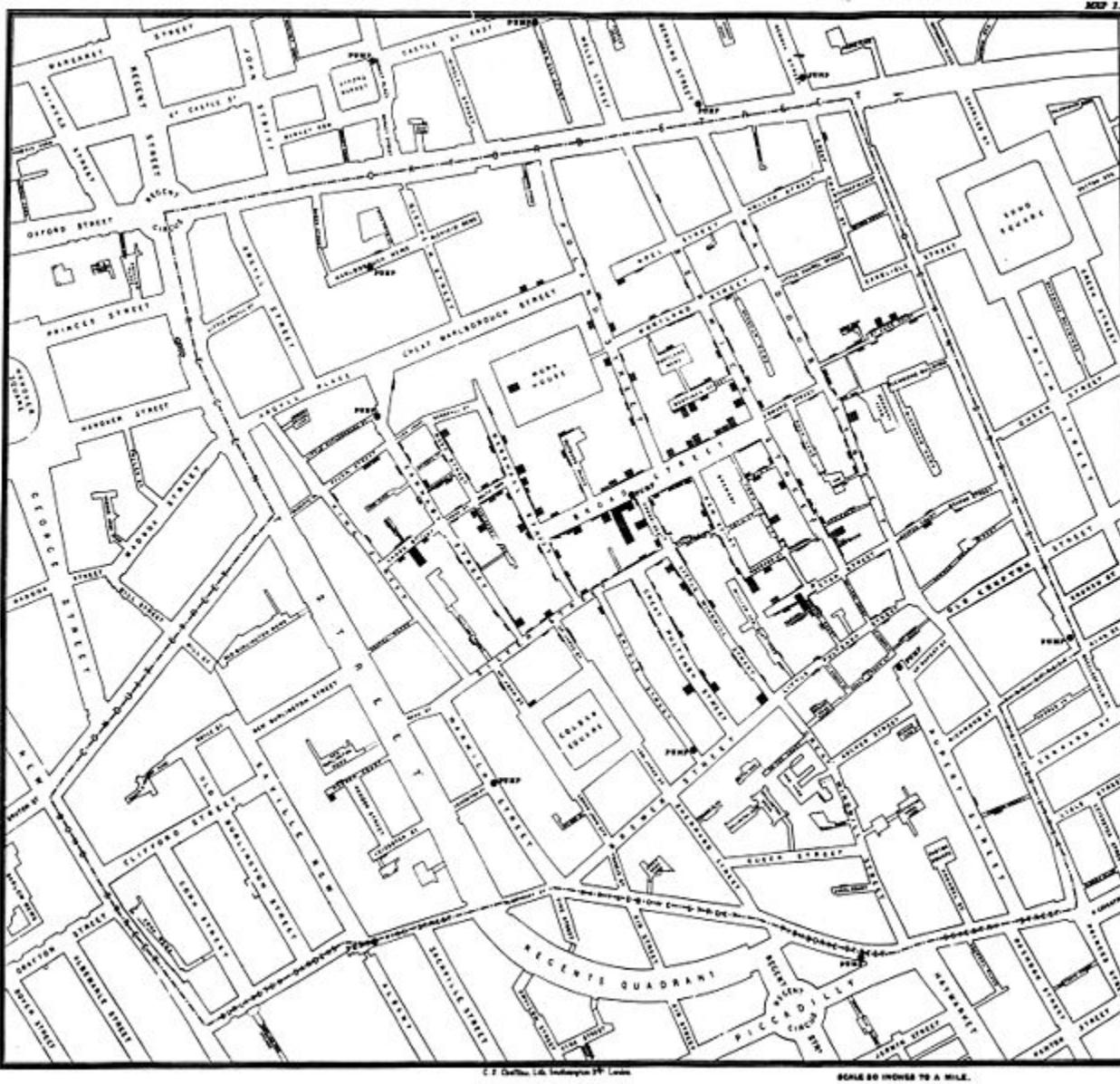
# Statistics for SPATIO-TEMPORAL DATA



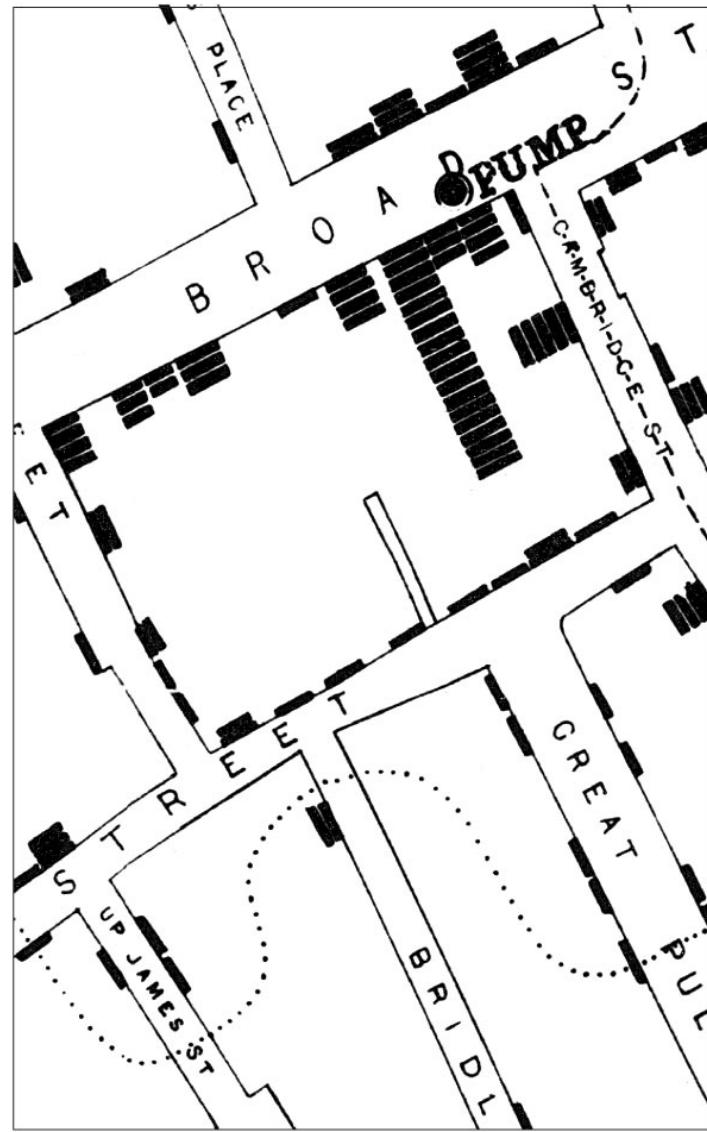
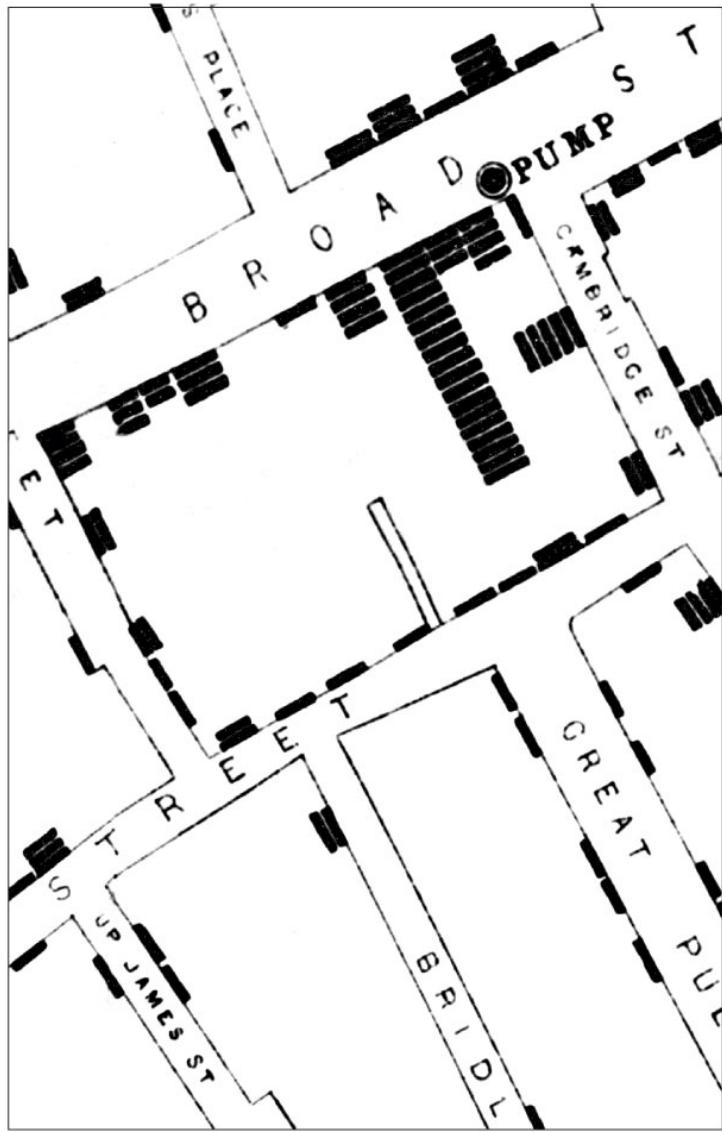
Noel Cressie • Christopher K. Wikle

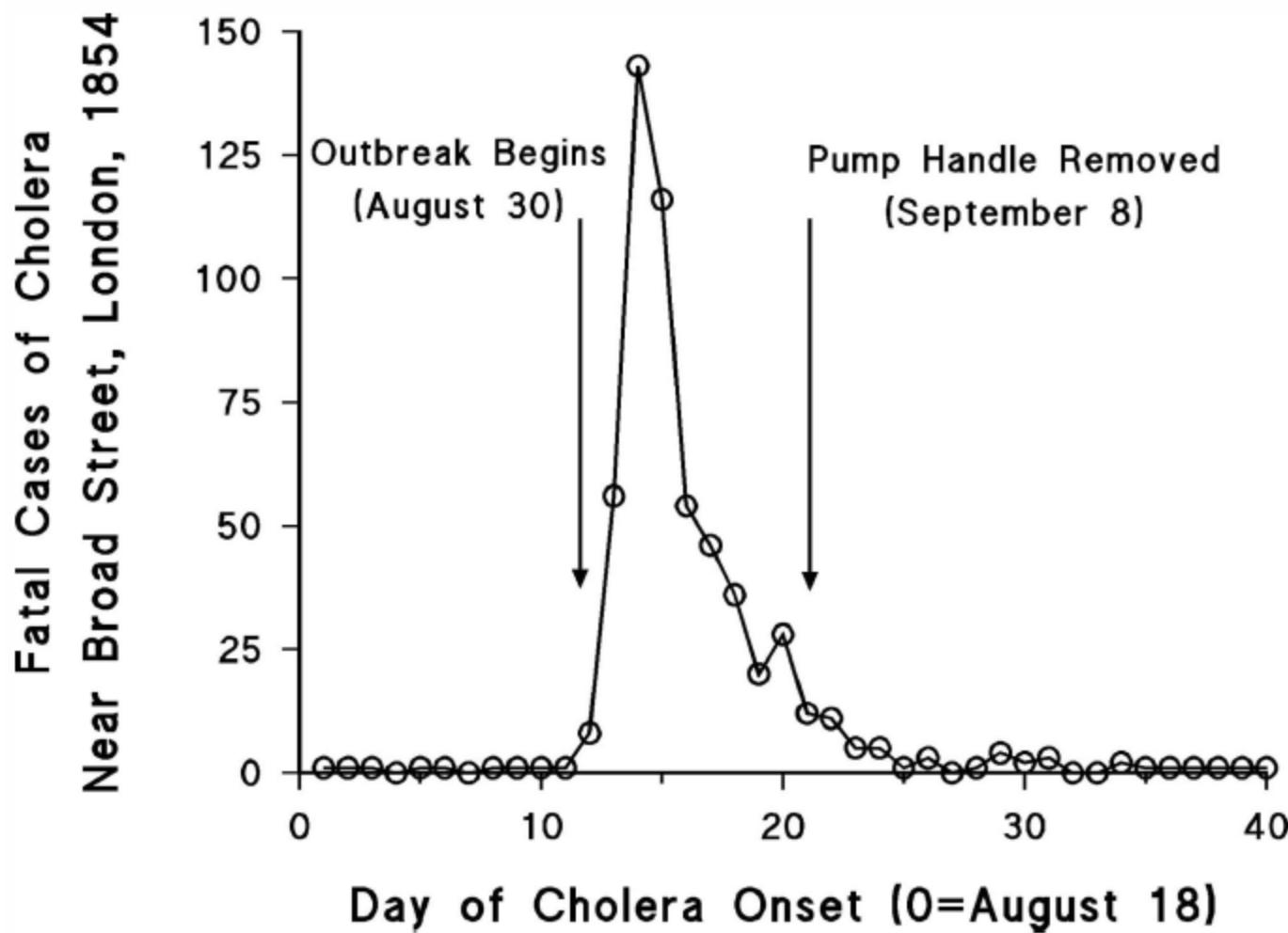
WILEY

ftp://  
wiley.com



Original map by Dr. John Snow showing the clusters of cholera cases in the London epidemic of 1854







Dr. John Snow (1813~1858)  
7/5/15



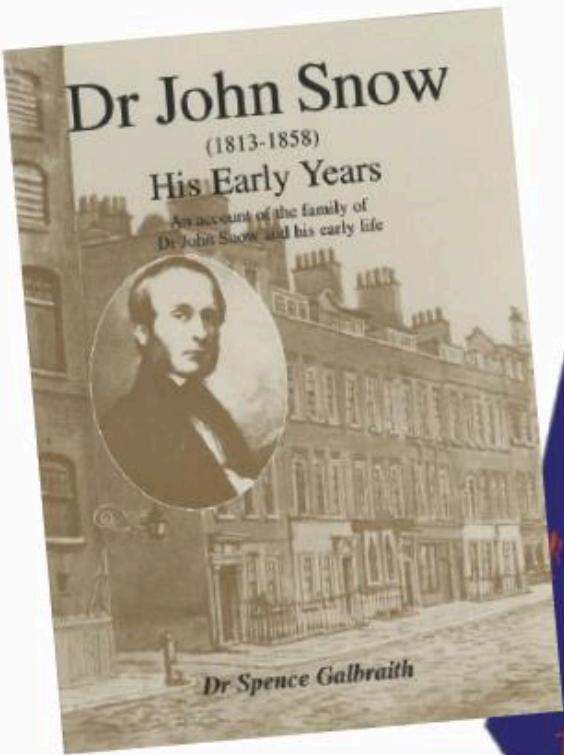
John Snow memorial and public house 11

# Exclusive to JSS Members

The ever popular silk ties are still available to members for £16 (£12 if collected in person). The Society also offers silver pump motif lapel pins, mugs and the book "John Snow: His Early Years" by Dr Spence Galbraith, available at the special price of £5.

Discount is available for bulk purchases.

Please contact [johnsnowsociety@rsph.org.uk](mailto:johnsnowsociety@rsph.org.uk) for more details.



The John Snow Society

www.johnsnowsociety.org

The John Snow Society

The Society | [John Snow](#) | [News](#) | [Lectures](#) | [Broad Sheets](#)

[Home](#) | [Constitution](#) | [Join/Contact](#)

**RSPH**  
ROYAL SOCIETY FOR PUBLIC HEALTH  
VISION, VOICE AND PRACTICE

## The John Snow Society

The John Snow Society aims to promote the life and works of [Dr John Snow](#), the pioneer of epidemiological method and celebrated anaesthetist.

**John Snow Bicentenary Conferences in 2013**

March 15-16: Mapping disease: John Snow and Cholera

April 11-12 : Snow's legacy: epidemiology today and tomorrow

[Programmes and registration details](#)

**The legacy of John Snow (York)**

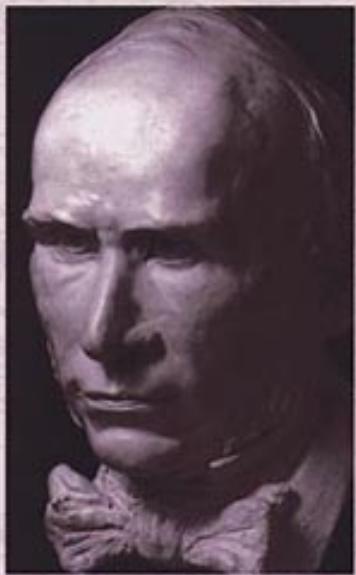
March 15: a one day event to mark the 200th anniversary of the birth of John Snow ([details](#)).

As outlined in the [Constitution](#), the Society has a serious intent, publishing [news](#), collecting facts and dates related to the life and works of John Snow and organising the [Annual Pumphandle Lecture Series](#), but it also aims to provide a communication network for epidemiologists and those trained in the Snow tradition throughout the world.

The Society currently has over 2,000 members worldwide, many of them eminent specialists in their fields.

# Cholera, Chloroform, and the Science of Medicine

A LIFE OF *John Snow*



PETER VINTEN-JOHANSEN  
HOWARD BRODY  
NIGEL PANETH  
STEPHEN RACHMAN  
MICHAEL RIP

# THE **GHOST MAP**

*The Story of London's  
Most Terrifying Epidemic—  
and How It Changed Science,  
Cities, and the Modern World*



A NATIONAL  
BESTSELLER

A NEW YORK TIMES  
NOTABLE BOOK

"By turns a medical thriller, detective story, and paean to city life, Johnson's account of the outbreak and its modern implications is a true page-turner."  
*The Washington Post*

author of

EVERYTHING BAD IS GOOD FOR YOU

**STEVEN  
JOHNSON**



THE GHOST MAP

# 감염지도

스티븐 존슨 | 김명남 옮김

세계화한 도시의 원형 19세기 런던에서 감염된 도시의 출구를 찾는다!

고병원성 AI, 광우병, 말라리아…

전염병의 대반격에 직면한 21세기

타성에서 탈피한 새로운 감염지도는 어떻게 만들어야 하는가?



7/5/15

민영사

## 세상을 바꾼 의학의 10대 발견

의학의 탄생

공중위생

세균 마취

엑스선

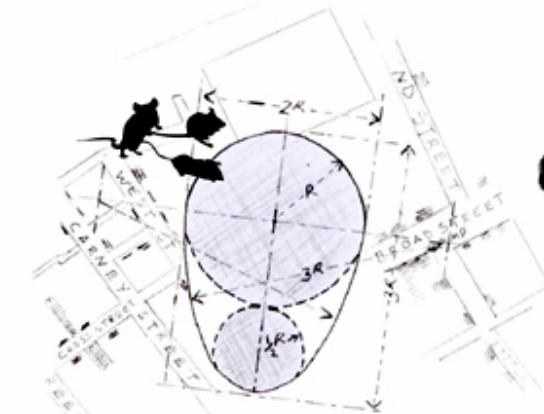
백신

항생제

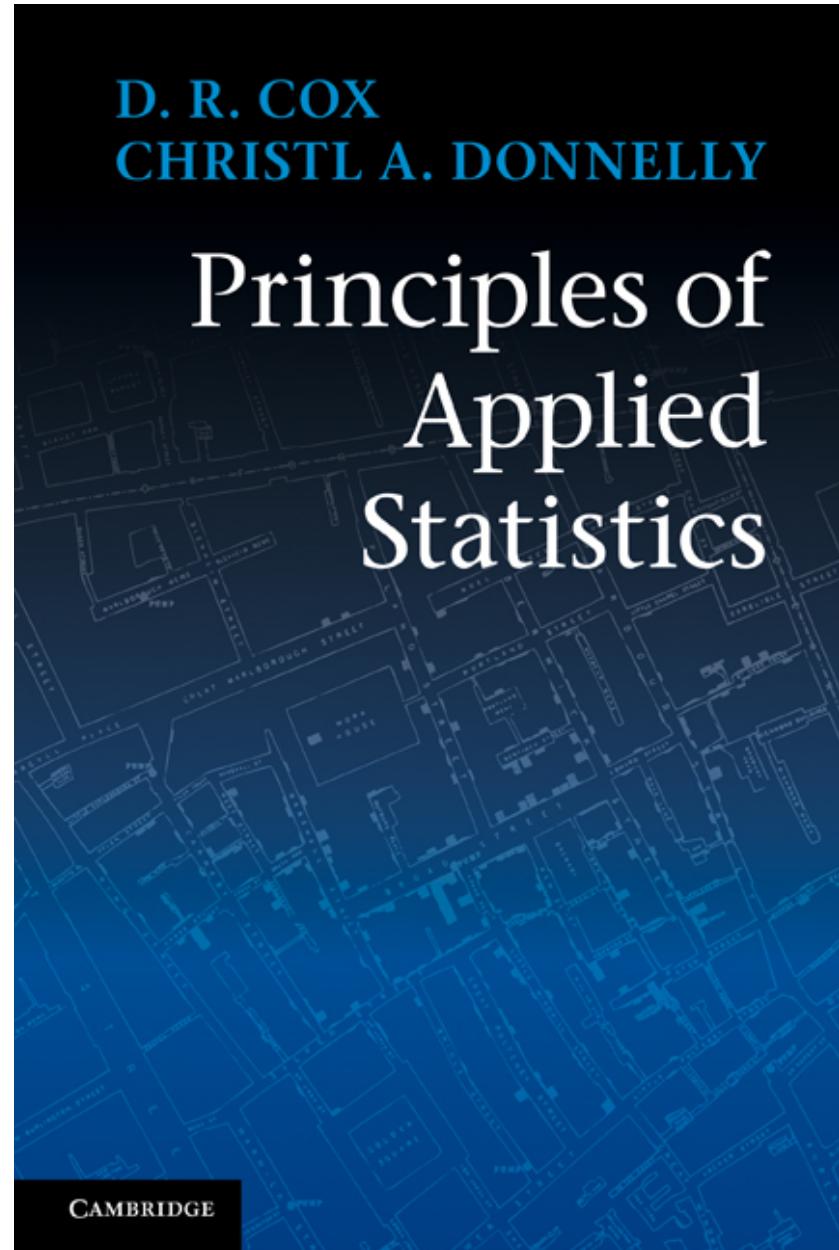
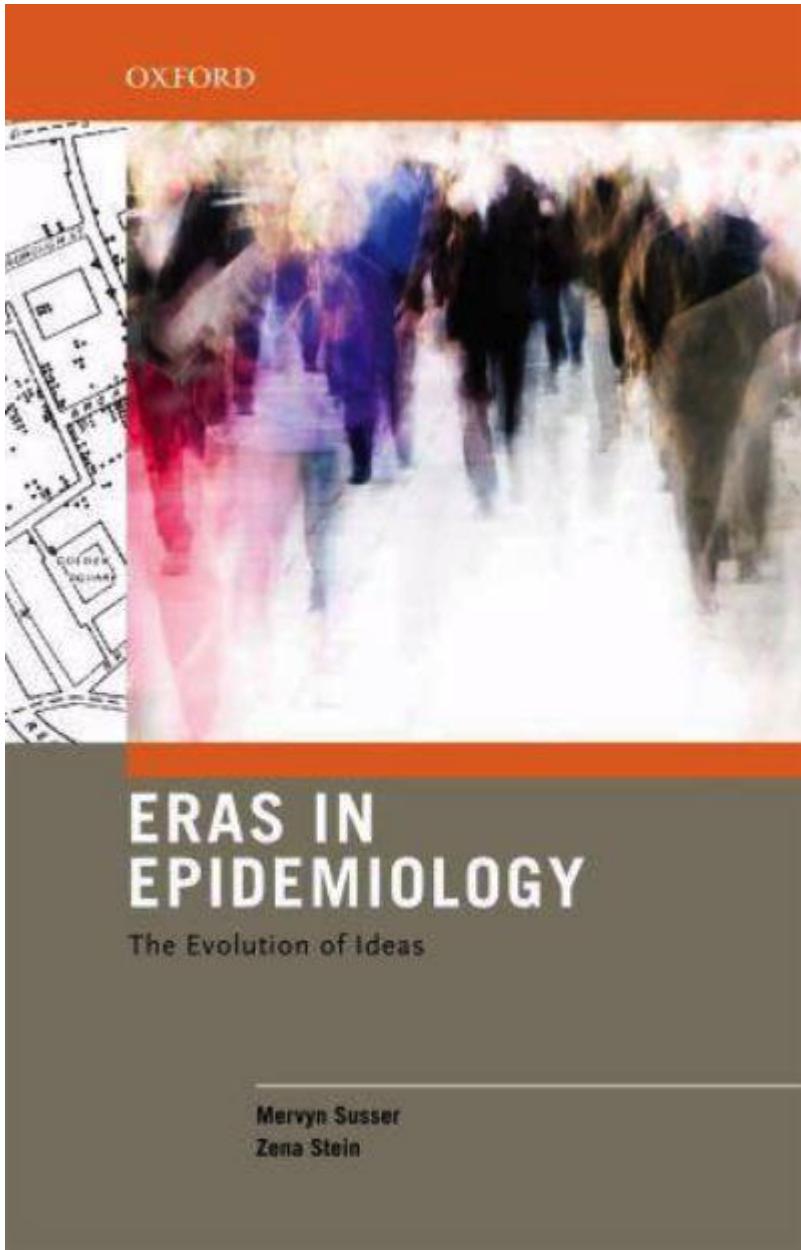
DNA

정신질환 치료제

통합의학



allcl&1



## [김태권의 인간극장] 전염병과 대결한 존 스노 (1813~1858)



한때는 콜레라가 미아스마라는 독기 때문에 생긴다고 했다. 그런데 본디 미아스마란 ‘공동체를 오염시키는 개인의 죄’를 이르는 그리스말이었다. 옛날에는 전염병을 개인 탓으로 돌렸던 것이다. 젊은 의사 존 스노가 역사를 바꾸었다. 19세기 중반 런던에 콜레라가 돌자 한 집 한집 꼼꼼한 조사 끝에 환자 통계지도를 그려 오염된 물이 병의 원인임을 알아냈다. 황승식 교수(인하대 의대)는 “원인균조차 발견되기 전에 정확한 데 이터에 근거해 콜레라 유행의 원인을 밝힌 존 스노는 현대 역학의 아버지라 불린다”며 그의 업적을 기린다. 오늘날 건강은 개인만이 아니라 국가가 함께 책임지는 문제다. 그래야 근대국가다.

John Snow's cholera x

www.theguardian.com/news/datablog/interactive/2013/mar/15/cholera-map-john-snow-

Edition: UK US AU Sign in About us Today's paper Subscribe

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here

the guardian

Home | UK | World | Sport | Football | Opinion | Culture | Economy | Lifestyle | Fashion | Environment | Tech | Travel | Money

News > Datablog

# DATA BLOG

Facts are sacred

Previous Blog home Next

## John Snow's cholera map of London recreated

What would John Snow's famous cholera map look like on a modern map of London, using modern mapping tools? The map changed what we know about germs and disease – and created a new way of looking at the world. With the help of mapping tool CartoDB and using the Stamen style maps, this is how it looks with larger circles representing more deaths. What do you think?

- Debate and download the data behind this map

Simon Rogers  
theguardian.com, Friday 15 March 2013 09.29 GMT

Water pumps Cholera deaths

Technology Mapping technologies

Society Cholera

World news

Science Microbiology

Culture Maps

More from Datablog on

Technology Mapping technologies

7/5/15 18

# 연구 설계

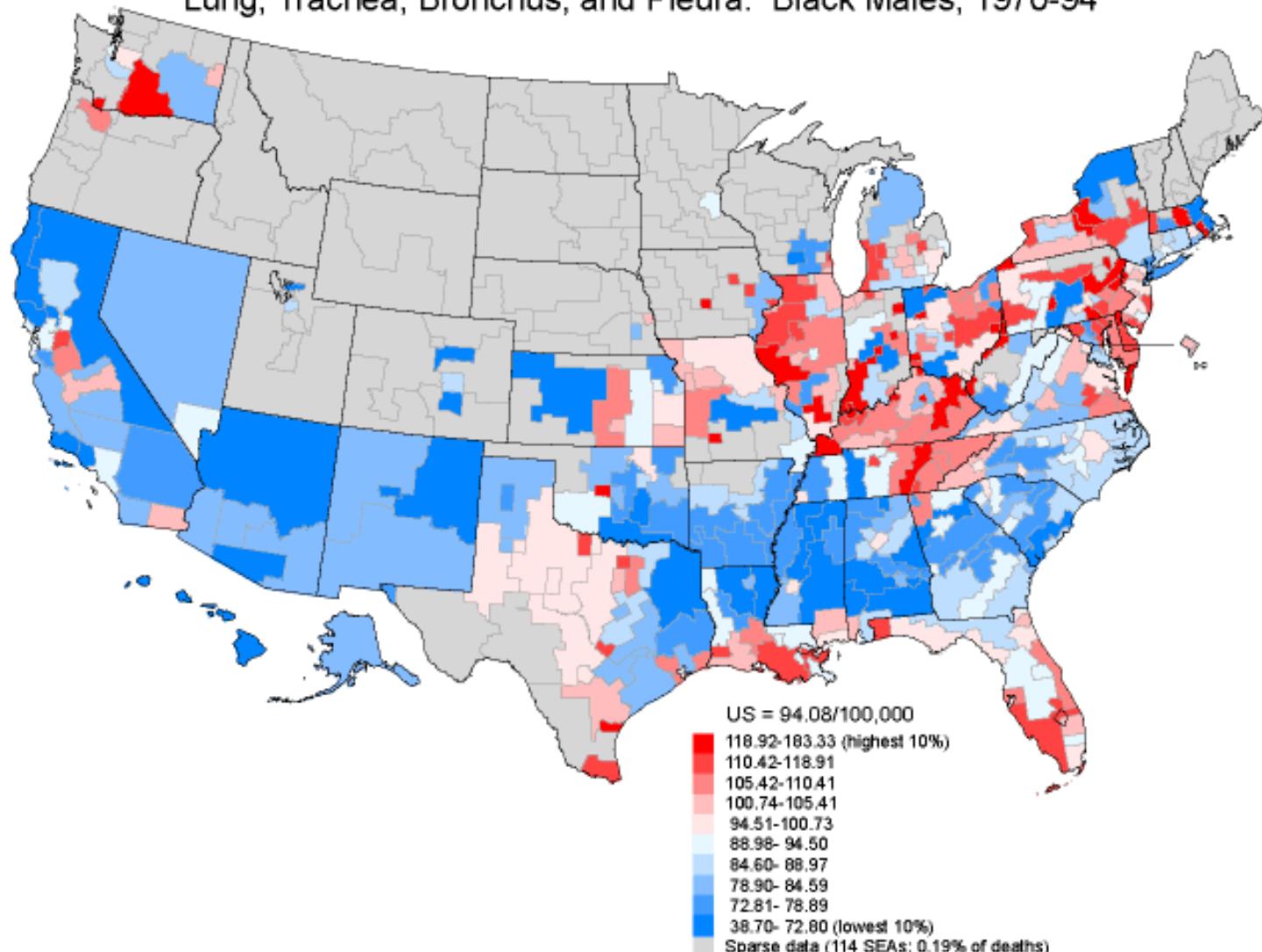
- 공간 역학 주요 연구 유형
  - 질병 지도화(Disease mapping) : 기술 연구
  - 지리적 상관성 연구(Geographical correlation studies) : 지리적 단위 비교 생태학적 연구
  - 군집 탐색(Cluster detection)
  - ‘오염원’ 연구(‘Point source’ studies) : 인과성 입증 근거 추가

# 질병 지도화 연구

- 질병 지도화에 근거한 기술 연구로 시간에 따른 질병 패턴의 변화를 파악 가능
- 질병 지도화는 노출과 질병 관련성에 대한 최초 가설 탐색에 유용
  - IARC 암 발생 및 암 사망 지도
  - US NCI 암 사망 지도(  
<http://www3.cancer.gov/atlasplus/type.html> )
    - 암 발생 패턴과 주요 제조업 산업단지 위치와의 생태학적 연구에 유용

Cancer Mortality Rates by State Economic Area (Age-adjusted 1970 US Population)

Lung, Trachea, Bronchus, and Pleura: Black Males, 1970-94





Thomas Mason, Robert Hoover, and Joseph Fraumeni examine a U.S. cancer mortality map in 1977.

# 지리적 상관성 연구

- 개인 수준보다 집단 수준으로 묶인 인구집단 정보를 이용한 연구로, 노출과 질병의 요약 통계량의 관련성 분석

$$y_i \sim \text{Poisson}(r_i E_i)$$
$$\log r_i = b_0 + \sum_k b_k X_{ik}$$

- 인과성 추론 전 단계에서 가설 생성 목적 이용

# 그림 5 역학 연구의 위상학

## Figure 5 Typology of Studies

		노출	
		X (개인 수준 측정)	X (집단 수준 측정)
결과	y (개인 수준 측정)	(y, x) 전통적 위험 요인 연구	(y, X) 다수준 연구
	Y (집단 수준 측정)	(Y, x)*	(Y, X) 생태학적 연구

\* 이 유형의 연구를 현 상태로 특징짓기는 불가능하다. 실제적으로는 X가 단순히 x의 중심 경향이 되므로 생태학적 연구는 (Y, X)의 형태를 띠게 될 것이다. 만일 Y의 분해가 가능하여, y를 관찰할 수 있다면 이 때는 (y, x)와 같게 될 것이다.

출처: Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: The perils of individualistic and ecologic fallacy. *Int J Epidemiol* 2009;38(2):342-360.

# 오염원 연구

- SAHSU 프로그램에서 공장과 같은 오염원과 건강 위험 탐색 연구 수행
  - 확산 모형 등으로 노출 범위 정의
  - 현재 노출 상태만 이용 가능한 경우, 만성 질환 평가에 어렵다는 제한점.
  - 확산 모형으로 수은 노출과 신장병 사망 연구 결과, 모형화한 대기중 수은 농도가 높을수록 신장병 사망 위험이 높음을 보고.

Rapid Inquiry Facility | S... X

www.sahsu.org/content/rapid-inquiry-facility

## Small Area Health Statistics Unit

Search this site...

About Us Our Research Publications News Events Rapid Inquiry Facility Environment and Health Atlas Environmental Exposures Group Contact

### Our Research

#### Rapid Inquiry Facility

The Rapid Inquiry Facility (RIF) is a freely available software application that supports two types of environmental health activities: disease mapping studies and risk analysis studies. It was designed to help epidemiologists and public health researchers to rapidly investigate potential environmental hazards, especially those related to industrial sites. The tool uses health, environmental, socio-economic, population and geographic data to calculate risks in relation to sources of exposure and to generate maps.

Disease mapping studies are used to visualise mortality or morbidity rates and risks across an area. They are used to explore spatial patterns of health outcomes; identify potential issues regarding data quality by geographical area; and identify areas which need additional study.

Risk analysis studies are used to provide an initial investigation into whether a suspected source of some particular exposure is having an impact on health in a local population. The tool can generate standardised rates and relative risks for a set of health outcomes. The RIF allows for full flexibility in the selection of a range of ages; a time frame specifying an era for data set collections; and a geographic area.

The software application was originally developed in the late 1990s by SAHSU staff. The RIF was further developed as part of the EUROHEIS and EUROHEIS2 <http://www.euroheis.org/> projects, funded by the European Commission's Directorate for General Health and Consumer Protection. The aim of these projects was to improve health information and analysis in order to assess relationships between environmental pollution and disease.

The RIF also received funding from the US Center for Disease Control (CDC) as part of the [US] Environment and Health Public Tracking Program (<http://www.cdc.gov/nceh/tracking/>). As an example, the State of Utah Department of Health used the RIF to investigate the perceived excess of leukaemia in relation to oil refineries in Utah [3]. The CDC support enabled modifications to the RIF to increase its functionality, ease of use and versatility.

# 질병 군집 탐색

- 질병 군집은 특정 인구집단에서 비슷한 또는 관련 있는 질병의 결집(aggregation)을 의미
  - 해당 시기, 지역, 작업장이나 클럽과 같은 하위 인구집단, 심지어 같은 약을 복용한 집단에서 발생하는 환자 숫자로 시작
  - 미국에서는 매년 1000건 이상 군집 보고
  - 군집 탐색은 비용과 시간이 많이 들며, 군집 발생과 원인이 명확하지 않은 경우 많은 고려가 필요

# 군집 탐색 설계

- 일반적 단계
  1. 환례 및 가능한 위험 요인 관련 군집 보고서의 원천 자료에 대한 정보 수집
  2. 보고 자료의 평가
    1. 환례와 진단명 검증,
    2. 위험 인구집단과 군집의 시공간 범위 특성화 및 정의, 적절한 참조 인구집단 선정,
    3. 질병 군집 평가에 특화된 기법을 이용하여 생물학적 및 통계적 초과 위험 가능성 결정
  3. 군집 생성 가설에 기초한 역학 연구 수행 가능성 평가
  4. 원인 연구(etiological) 수행

*Centers for Disease Control and Prevention*



Recommendations and Reports / Vol. 62 / No. 8

Morbidity and Mortality Weekly Report

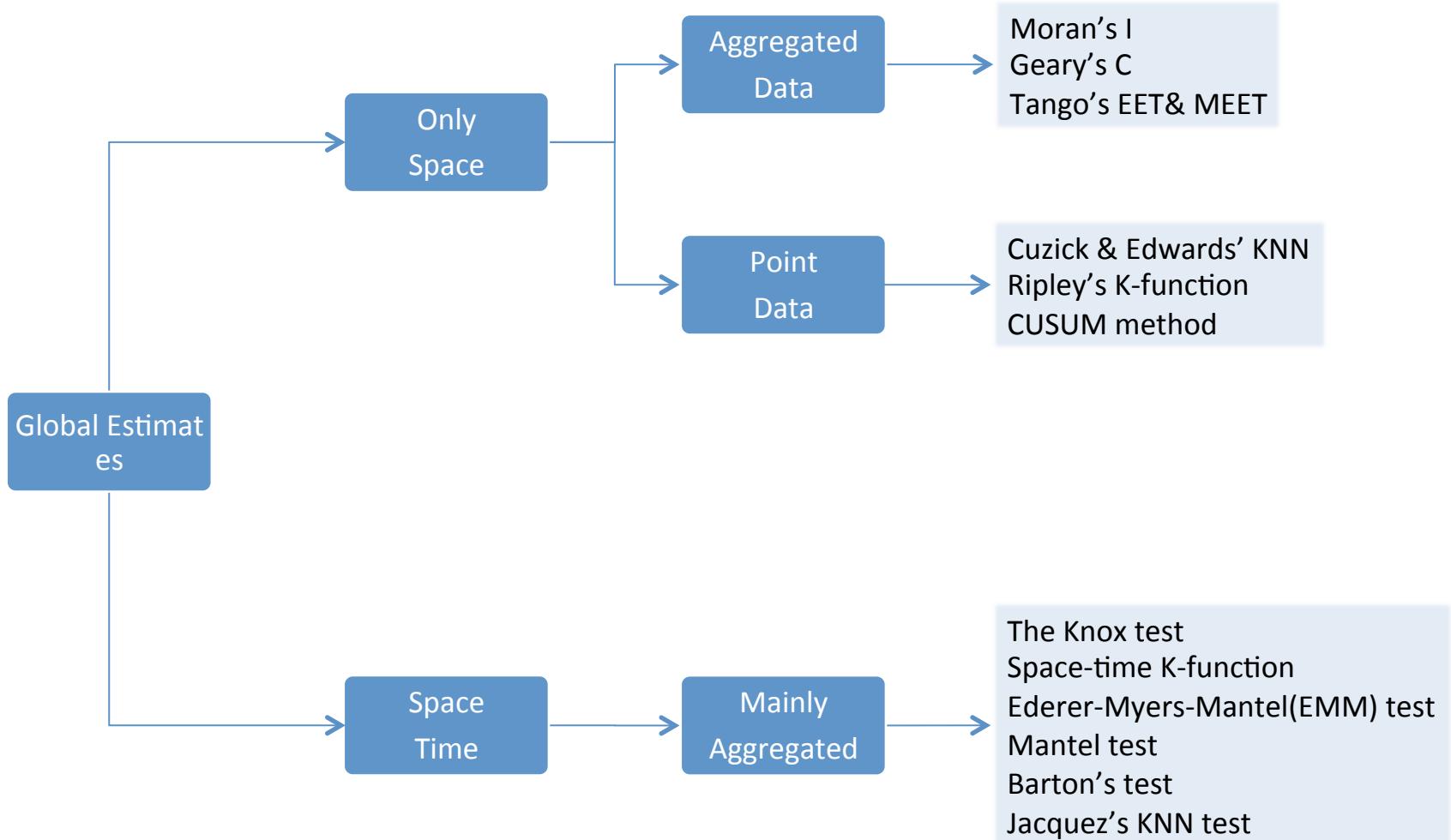
September 27, 2013

## Investigating Suspected Cancer Clusters and Responding to Community Concerns

Guidelines from CDC and  
the Council of State and Territorial Epidemiologists

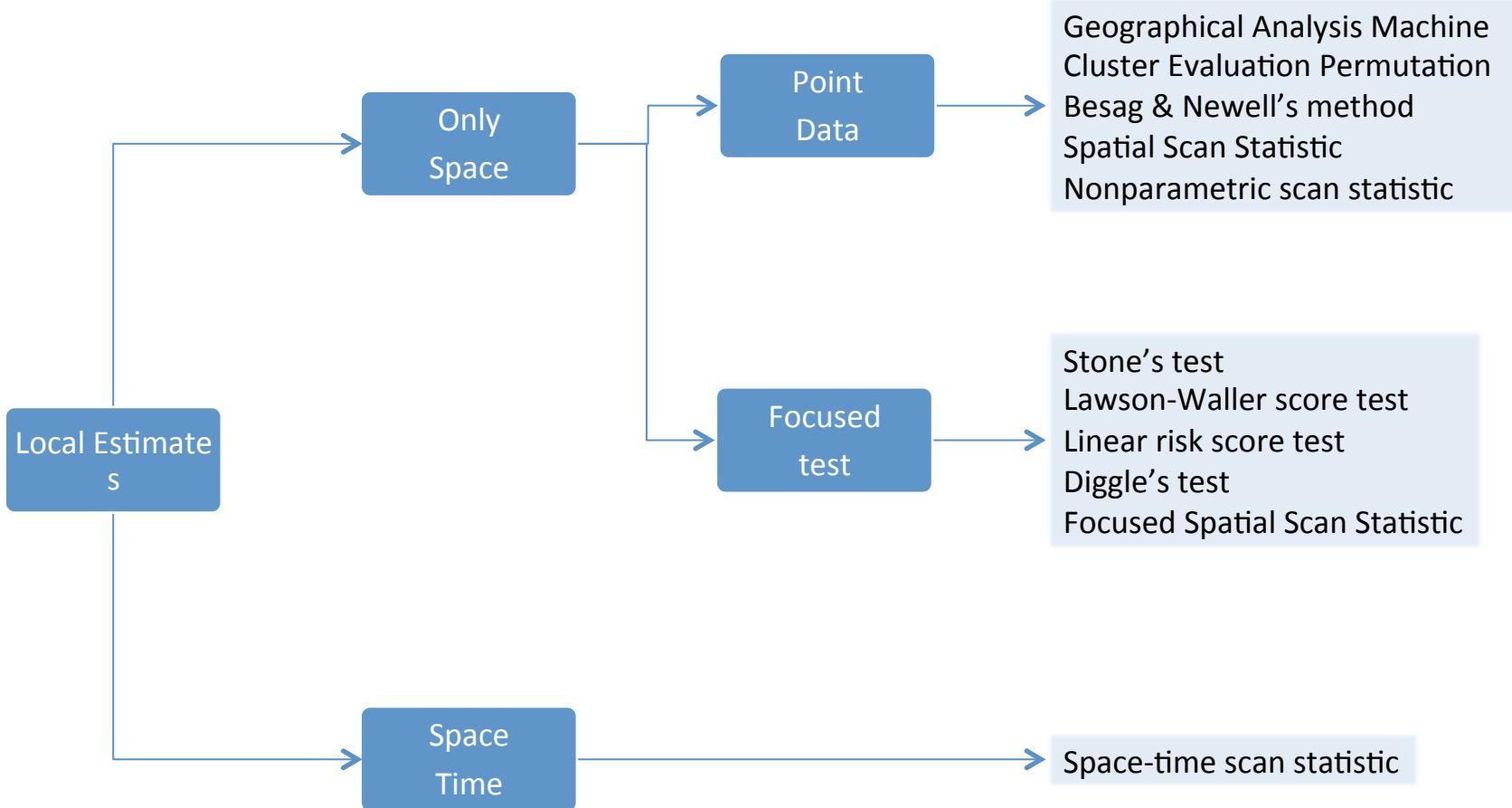
# 공간군집 종류(1)

- Global Estimates:
  - 조사지역 전체를 놓고 볼 때 군집이 존재하는지 검정
  - 구체적으로 어느 지역인지는 구하지 않음
  - 공간분석의 초기 단계, 공간적 상관관계를 고려한 분석을 할 지 결정하는 기준



# 공간군집 종류(2)

- Local Estimates:
  - Global estimates에 비해 특정 지역을 찾는다는 점에서 보다 구체적임
  - 역학에서의 질병 군집, 환경 오염원 탐지 등



## **2. 시공간 역학 연구 사례**

# Impact of outdoor air pollution on the incidence of tuberculosis in the Seoul metropolitan area, South Korea

Seung-sik Hwang<sup>1</sup>, Sungchan Kang<sup>1</sup>, Ji-Young Lee<sup>1</sup>, Ji Sun Lee<sup>2</sup>, Hee Jin Kim<sup>3</sup>, Sung Koo Han<sup>2</sup>, and Jae-Joon Yim<sup>2</sup>

<sup>1</sup>Department of Social and Preventive Medicine, Inha University School of Medicine, Incheon; <sup>2</sup>Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul; <sup>3</sup>Korean Institute of Tuberculosis, Osong, Korea

**Background/Aims:** Although indoor air pollution is a well-known risk factor for tuberculosis (TB), the possible link between outdoor air pollution and TB development has not been examined fully. We assessed the impact of outdoor air pollution on TB development in the Seoul metropolitan area, South Korea.

**Methods:** The mean concentrations of ambient particulate matter (PM) with an aerodynamic diameter  $\leq 10 \mu\text{m}$  ( $\text{PM}_{10}$ ),  $\text{O}_3$ ,  $\text{CO}$ ,  $\text{NO}_2$ , and  $\text{SO}_2$  levels in Seoul, between January 1, 1997 and December 31, 2006, were determined. Furthermore, their association with the risk of developing TB after adjusting for socioeconomic

# Spatial clustering method via generalized lasso

Eunjung Song · Hosik Choi · Seungsik Hwang · Woojoo Lee<sup>1</sup>

Department of Statistics, Inha University

(Received 0000; accepted 0000)

---

## 요약

In this paper, we propose a penalized likelihood method for detecting spatial cluster. The basic idea is that we can find areas having similar characteristic because the fused lasso , a special case of generalized lasso, encourages mean between the neighboring areas to be similarly estimated. To optimize the objective function, genlasso is exploited. But solutions of the objective function depend on a tuning parameter, we select it by using Wang's BIC. Also this method can easily add not only categorical variables but also continuous variables to regression model unlike existing methods. Lastly, we apply our proposed method to the Seoul tuberculosis data.

**Keywords:** spatial clustering, generalized lasso, lasso, fused lasso

---

# 결과

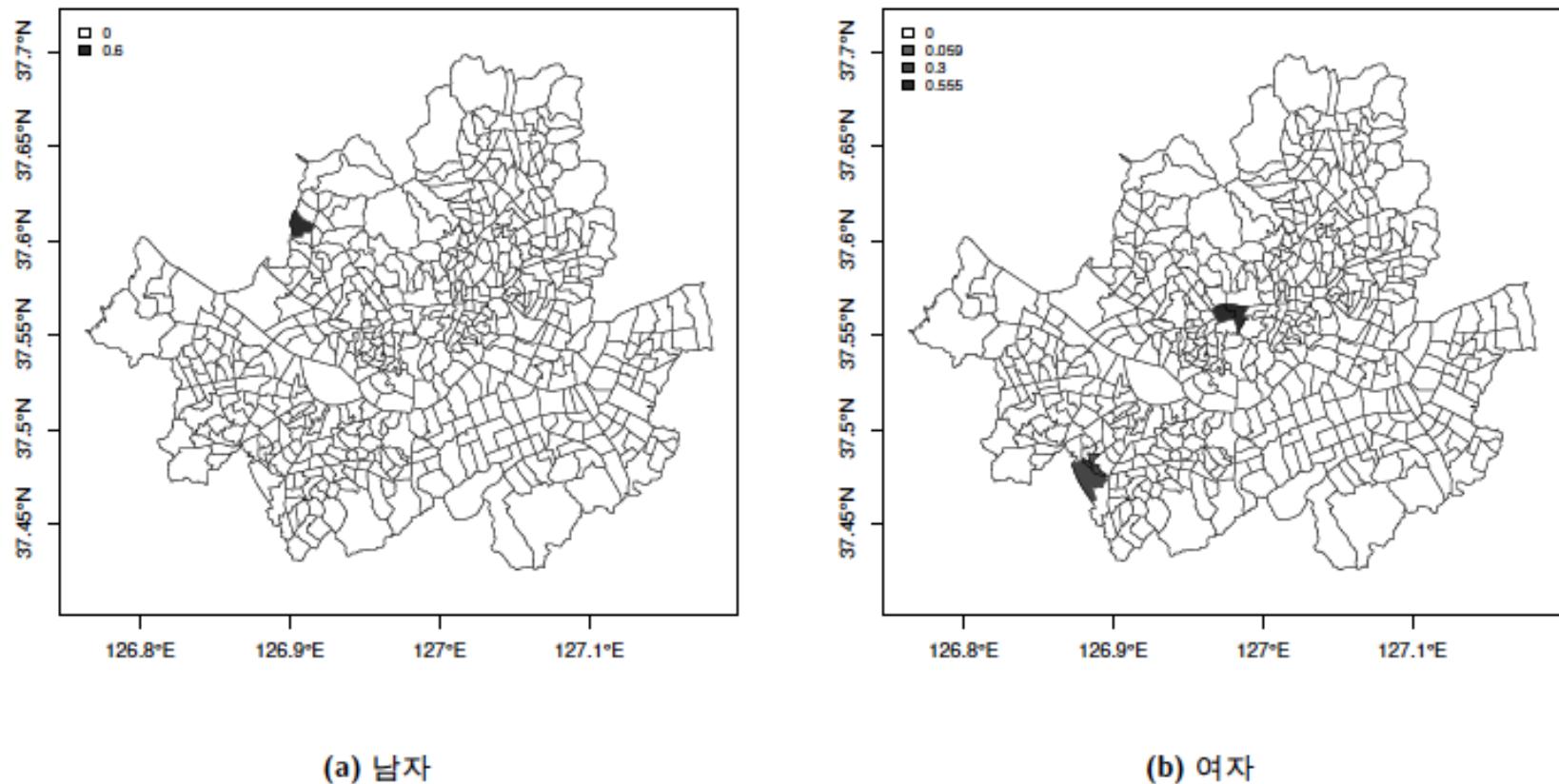
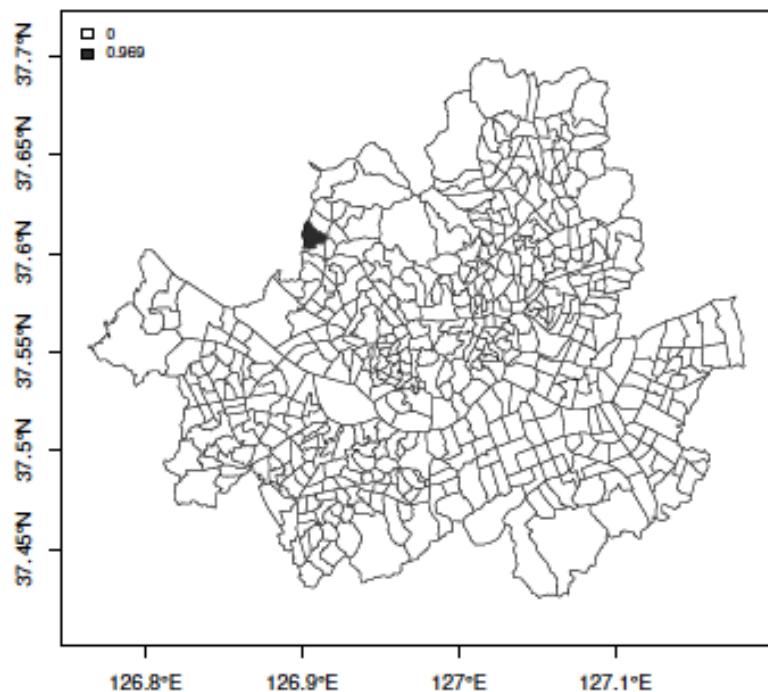
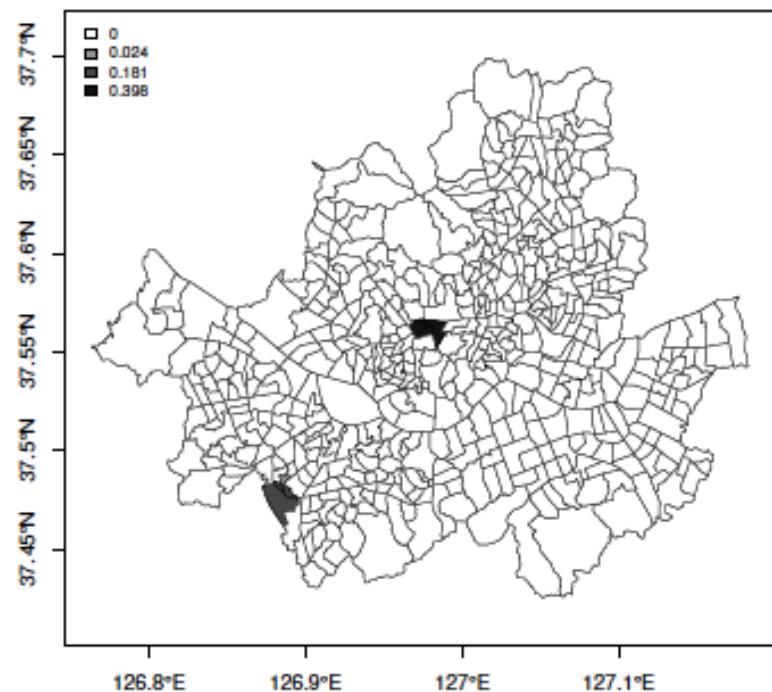


그림 4.2. 설명변수를 이용하지 않았을 때 genlasso 분석 결과

# 결과

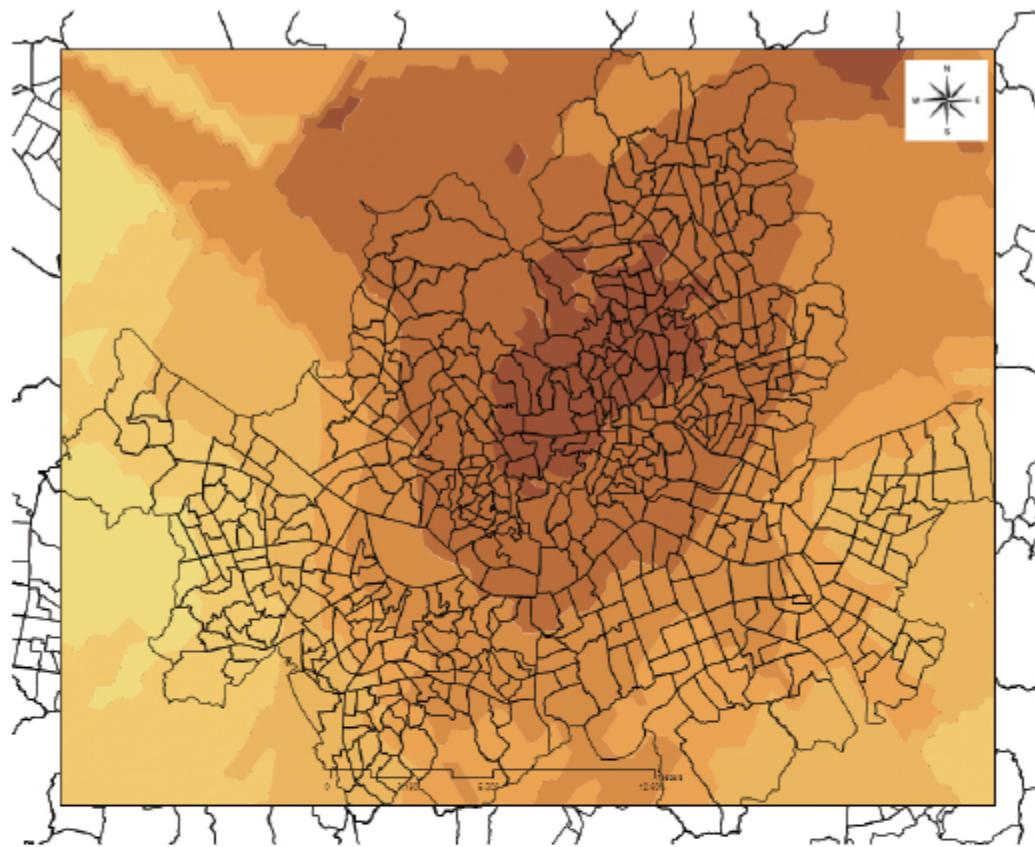


(a) 남자



(b) 여자

그림 4.3. 설명변수로 보정한 후 결핵군집



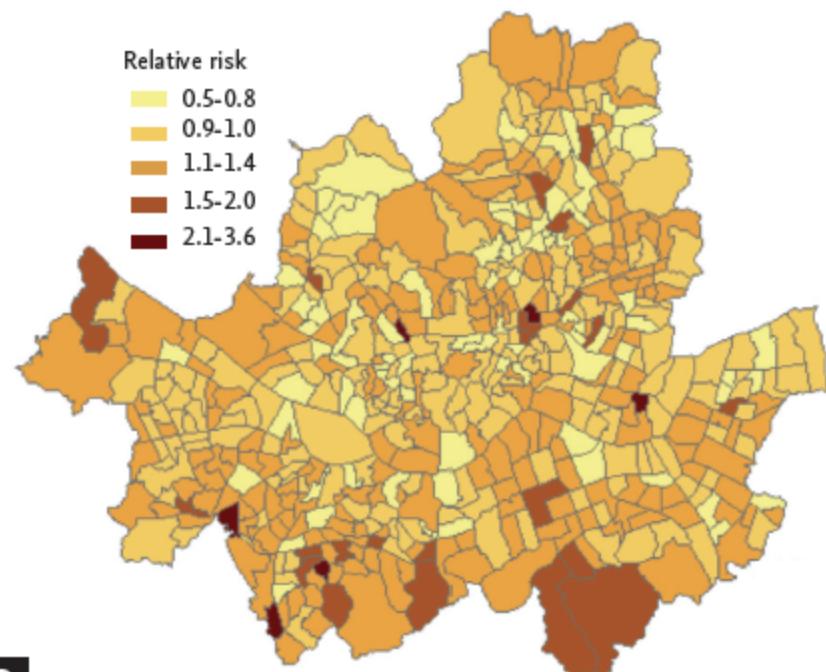
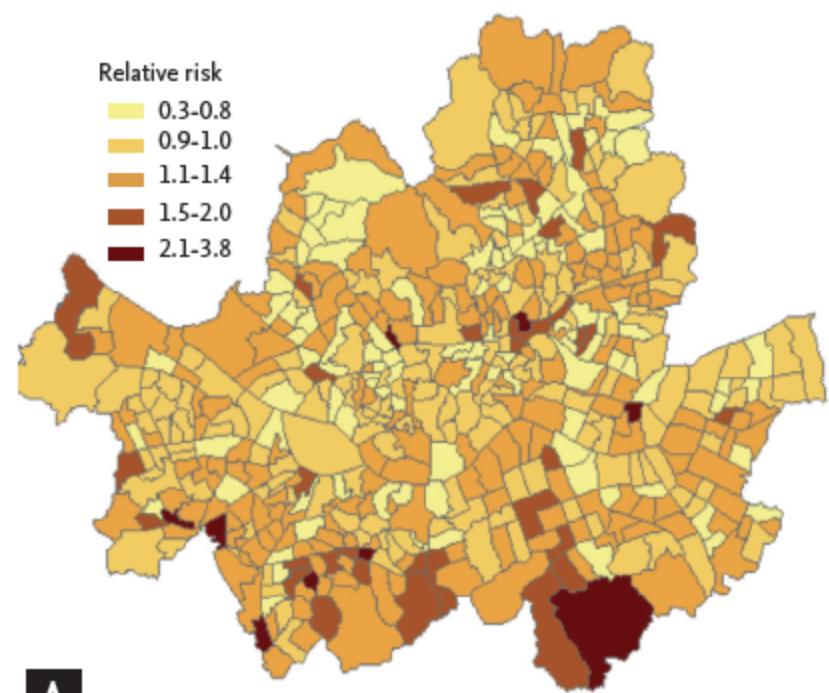
■ Seoul township  
Ordinary kriging prediction map

SO<sub>2</sub> concentration (ppb)

4.41-4.95	4.95-5.32	5.32-5.57	5.57-5.75
5.75-5.87	5.87-6.04	6.04-6.29	6.29-6.66
6.66-7.21	7.21-8.00		

**Figure 1.** Estimated SO<sub>2</sub> levels using air-monitoring data and kriging between January 1, 1997 and December 31, 2006, in the Seoul metropolitan area.

# 결과



**Figure 2.** Unsmoothed (A) and smoothed (B) map of the standardized incidence ratios of tuberculosis in the Seoul metropolitan area between January 1, 2002 and December 31, 2006 (smoothing of risk estimates were calculated based on Bayesian inference methods).

# 결과

Table 4. Impact of an interquartile increase in pollutant concentration on the incidence of tuberculosis

Model	Male RR <sup>a</sup> (95% CrI)	Female RR <sup>a</sup> (95% CrI)
PM <sub>10</sub> , µg/m <sup>3</sup>	0.98 (0.94–1.02)	1.01 (0.97–1.06)
O <sub>3</sub> , ppb	0.99 (0.94–1.03)	1.01 (0.97–1.05)
CO, ppb	0.99 (0.95–1.03)	1.01 (0.98–1.04)
NO <sub>2</sub> , ppb	1.00 (0.96–1.05)	1.01 (0.98–1.05)
SO <sub>2</sub> , ppb	1.07 (1.03–1.12)	1.02 (0.98–1.07)

RR, relative risk; CrI, credible interval; PM<sub>10</sub>, particulate matter with an aerodynamic diameter ≤ 10 µm.

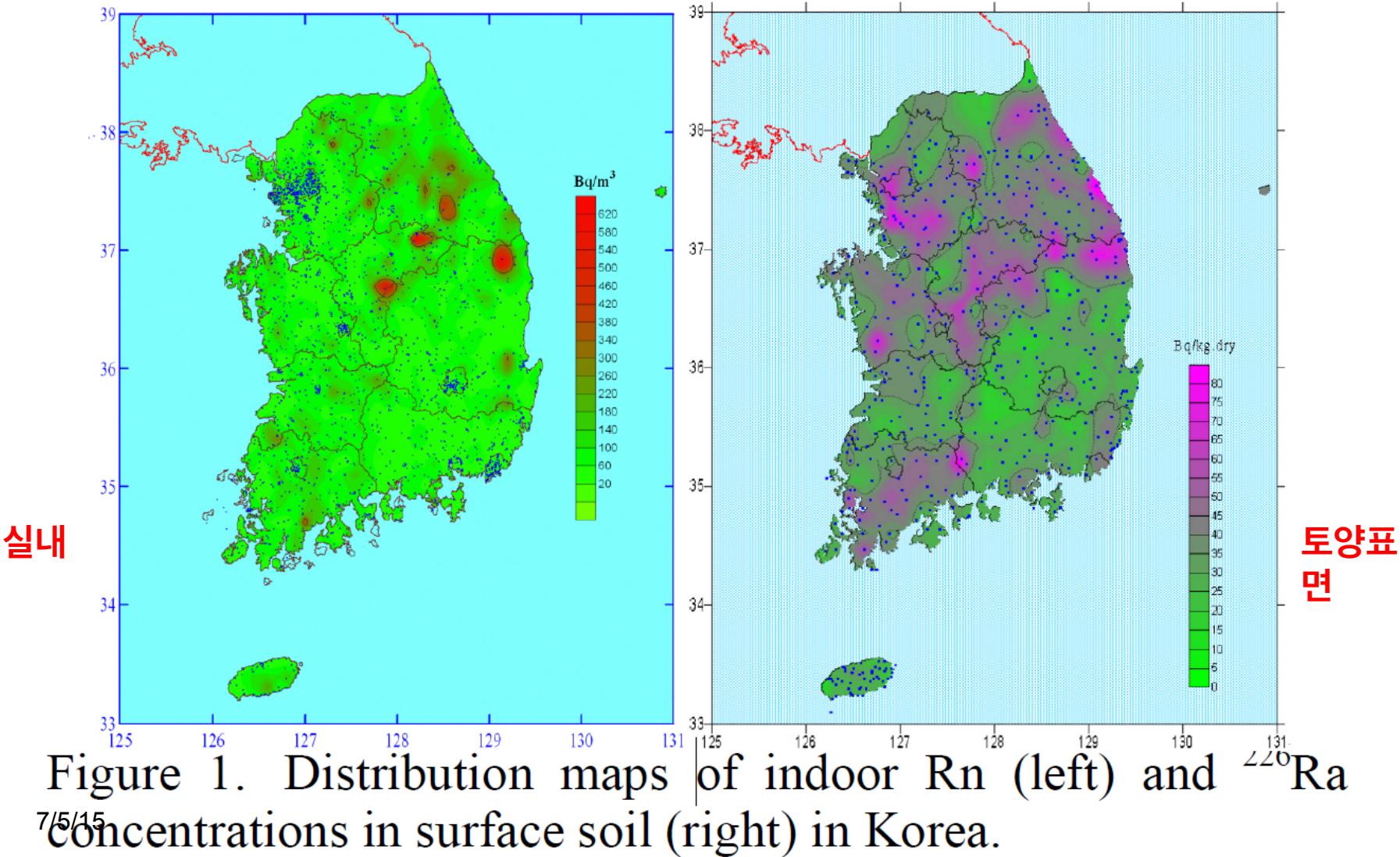
<sup>a</sup>RRs were adjusted for the quintiles of the Carstairs index as an indicator variable.



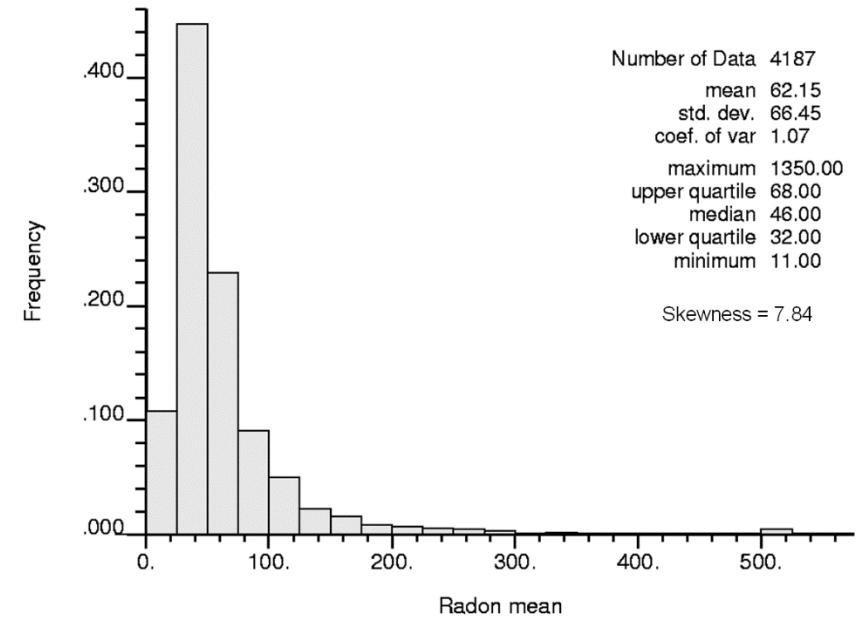
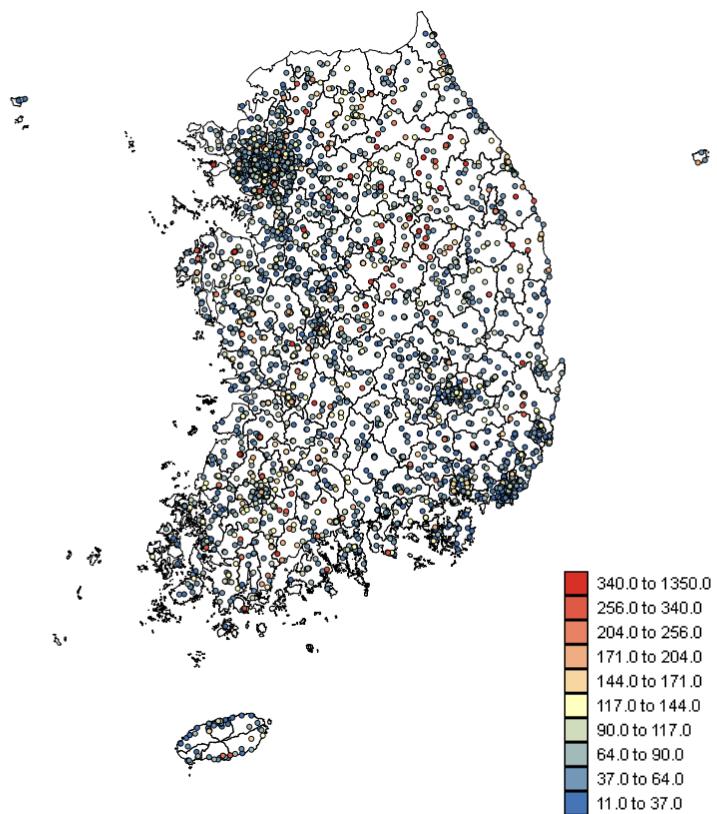
# 라돈가스 노출 수준과 조혈기계 암 발 생에 관한 지리상관성 연구

# 우리나라의 라돈 농도 분포

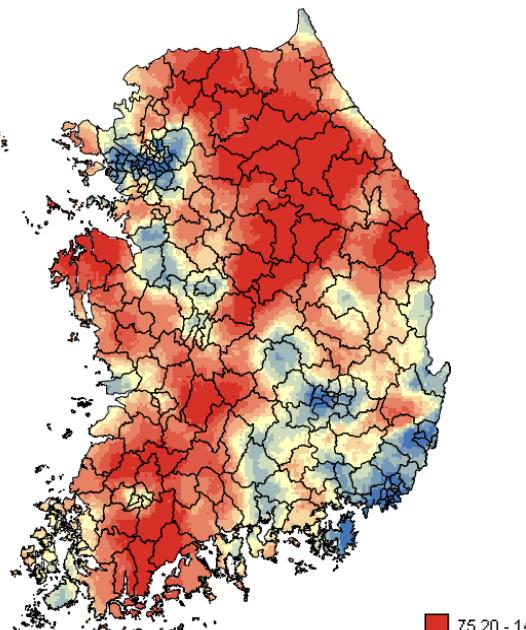
(Kim Yongjae 등, 2011, Rad Prot Dosim in press)



# 라돈 관측 자료 크리깅 (1)

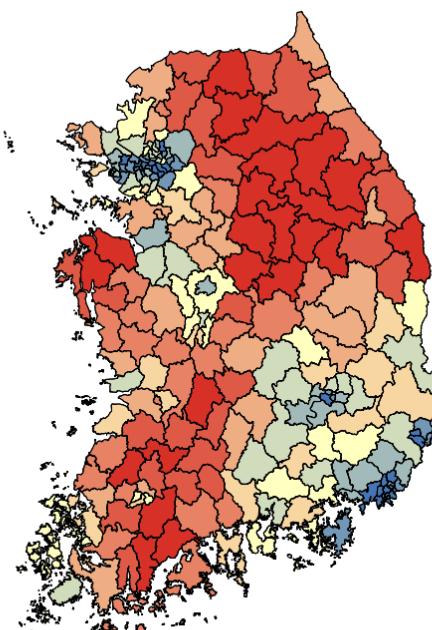


# 라돈 관측 자료 크리깅 (2)



75.20 - 144.70
67.38 - 75.20
59.31 - 67.38
54.51 - 59.31
51.09 - 54.51
47.16 - 51.09
43.70 - 47.16
39.55 - 43.70
35.62 - 39.55
28.00 - 35.62

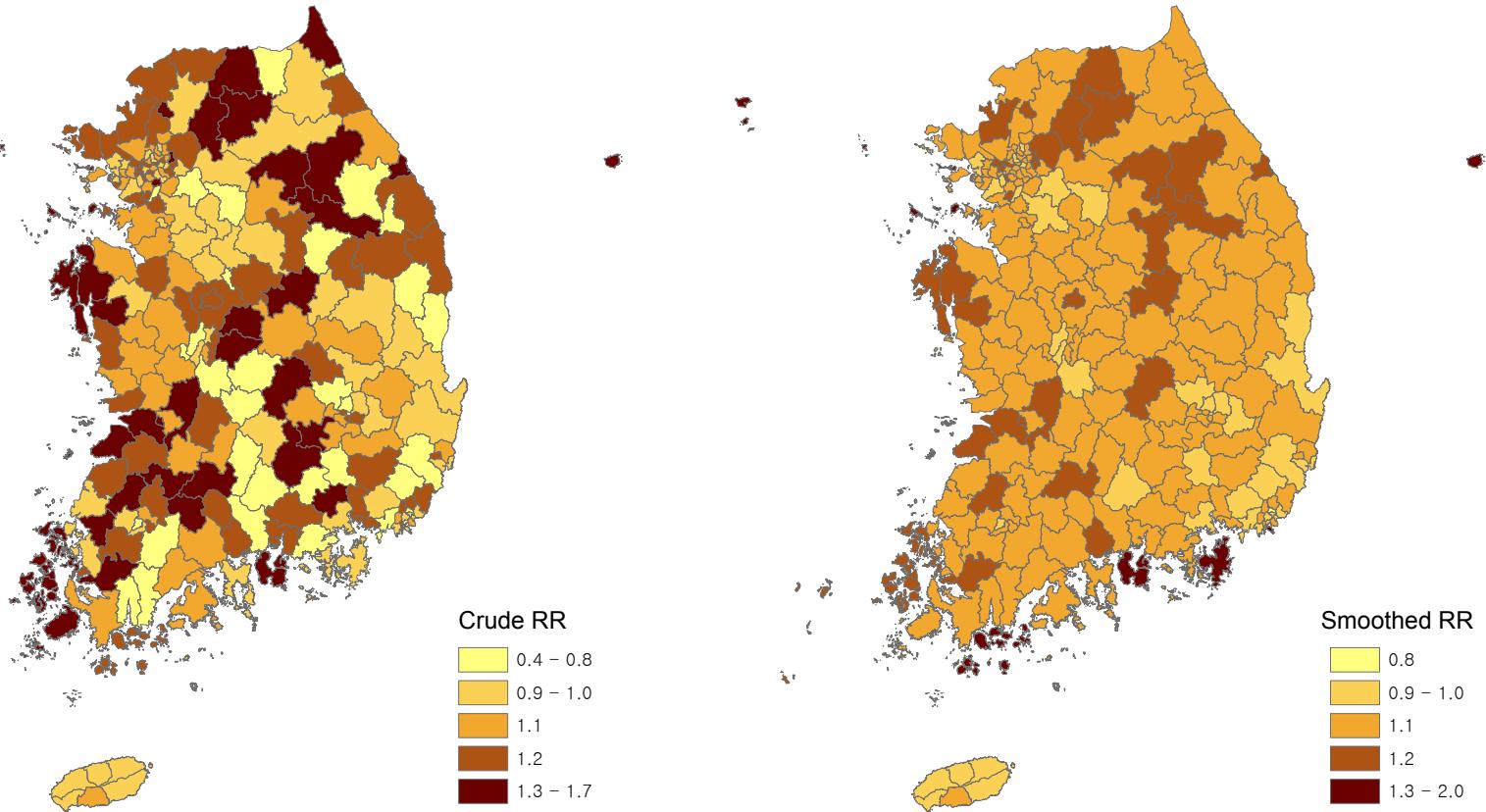
정규 크리깅 적용 결과



75.20 - 119.61
67.38 - 75.20
59.31 - 67.38
54.51 - 59.31
51.09 - 54.51
47.16 - 51.09
43.70 - 47.16
39.55 - 43.70
35.62 - 39.55
28.00 - 35.62

행정구역별 평균 라돈 농도 분포도

# 남자 비호지킨 림프종 지도화



# 계층적 베이지언 모형 구조

- 계층적 베이지언 방법

$$O_i \sim Poisson(\mu_i)$$

$$\mu_i = \theta_i E_i$$

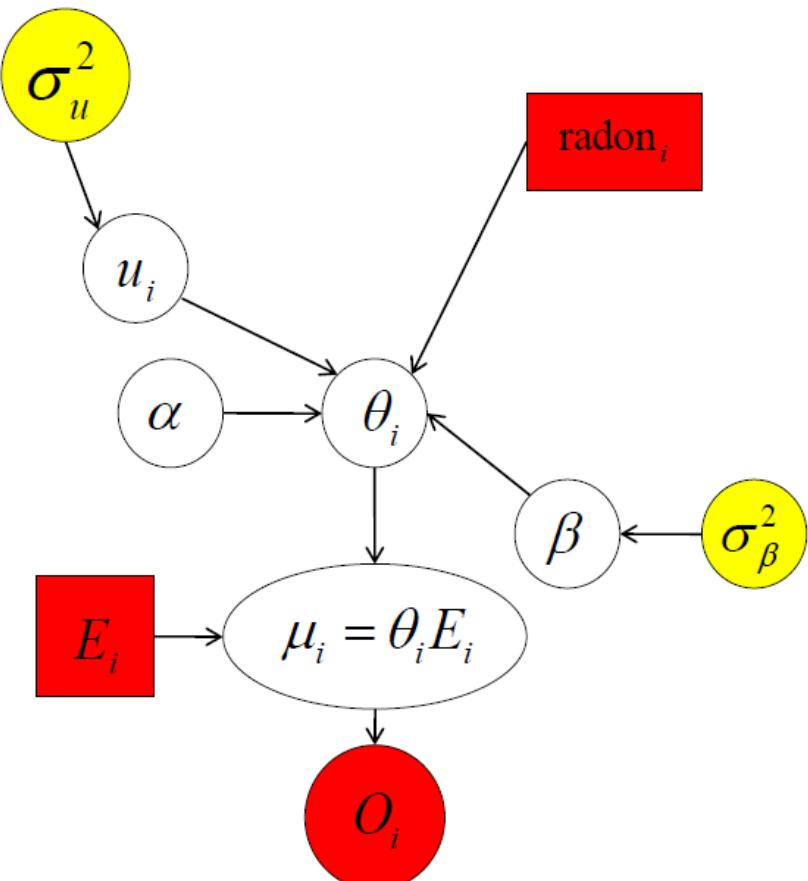
$$\log(\theta_i) = \alpha + \beta \times \text{radon}_i + u_i$$

$$u_i \sim Normal(0, \sigma_u^2)$$

$$\alpha \sim flat$$

$$\beta \sim Normal(0, 100000)$$

$$\sigma_u^2 \sim Gamma(0.5, 0.0005)$$



# 계층적 베이지언 모형 결과

라돈 농도 1 Bq/m<sup>3</sup>증가 당

암종	남자		여자	
	비교위험도 Relative risk	95% 신용구간 95% Credible interval	비교위험도 Relative risk	95% 신용구간 95% Credible interval
백혈병	0.999	(0.997, 1.000)	0.999	(0.997, 1.001)
비호지킨 림프종	<b>1.002</b>	<b>(1.000, 1.004)</b>	<b>1.004</b>	<b>(1.002, 1.006)</b>
폐암	<b>1.003</b>	<b>(1.002, 1.004)</b>	1.000	(0.999, 1.002)

# 여자 폐암 INLA 모형 결과

## (라돈농도, 흡연율, 박탈지수)

```
> summary(result_lungf4)

Call:
c("inla(formula = mod4_f, family = \"poisson\", data = lung_f, E = E_LUNGF, ",  "

Time used:
Pre-processing    Running inla Post-processing          Total
0.3730221        7.9194529      0.6680379        8.9605129

Fixed effects:
               mean        sd  0.025quant  0.5quant  0.975quant      kld
(Intercept) 0.00312695 0.006215104 -0.009091999 0.003138219  0.0152847 0.03160175

Random effects:
Name      Model      Max KLD
region.struct Besags ICAR model
region     IID model
radon     RW2 model
smoke     RW2 model
dep       RW2 model

Model hyperparameters:
                     mean        sd  0.025quant  0.5quant  0.975quant
Precision for region.struct 103.351    30.272    57.139    98.900   174.864
Precision for region      20624.077 19244.775 1899.242 15163.575 71548.924
Precision for radon       208.966    59.815   114.502   201.381   347.507
Precision for smoke        88.592    37.621    34.445    82.392   179.243
Precision for dep         33.901    24.664     5.864    27.861   97.737

Expected number of effective parameters(std dev): 108.31(8.493)
Number of equivalent replicates : 2.161

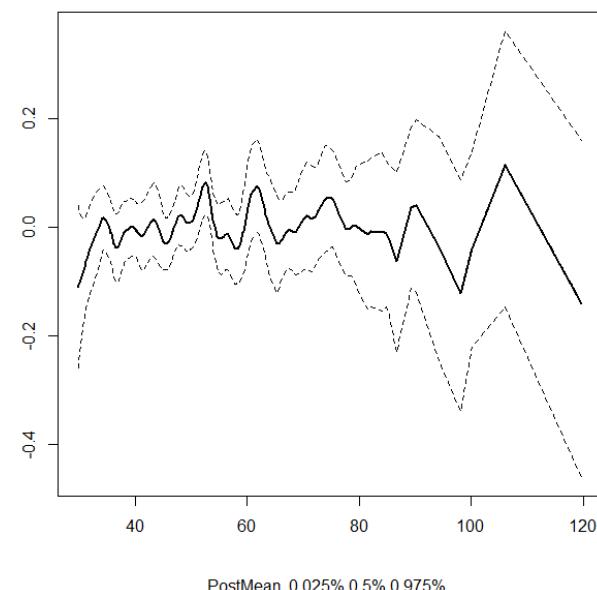
Deviance Information Criterion: 1939.41
Effective number of parameters: 109.23

Marginal Likelihood: -3536.35
CPO and PIT are computed

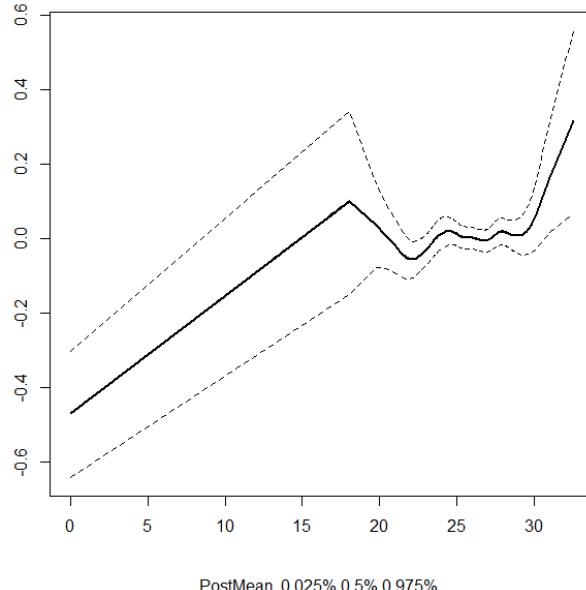
Posterior marginals for linear predictor and fitted values computed
```

# 앞의 모형에서 라돈 농도, 흡연율, 박탈지수가 여자 폐암 발생 위험도에 미치는 효과

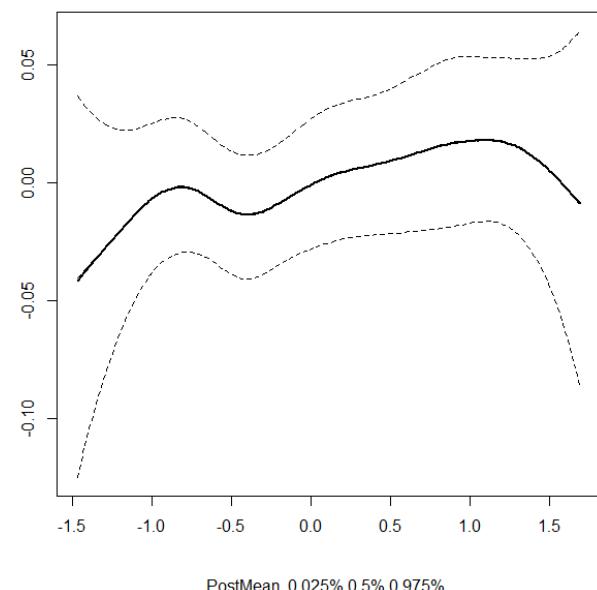
radon



smoke



dep



라돈농도

흡연율

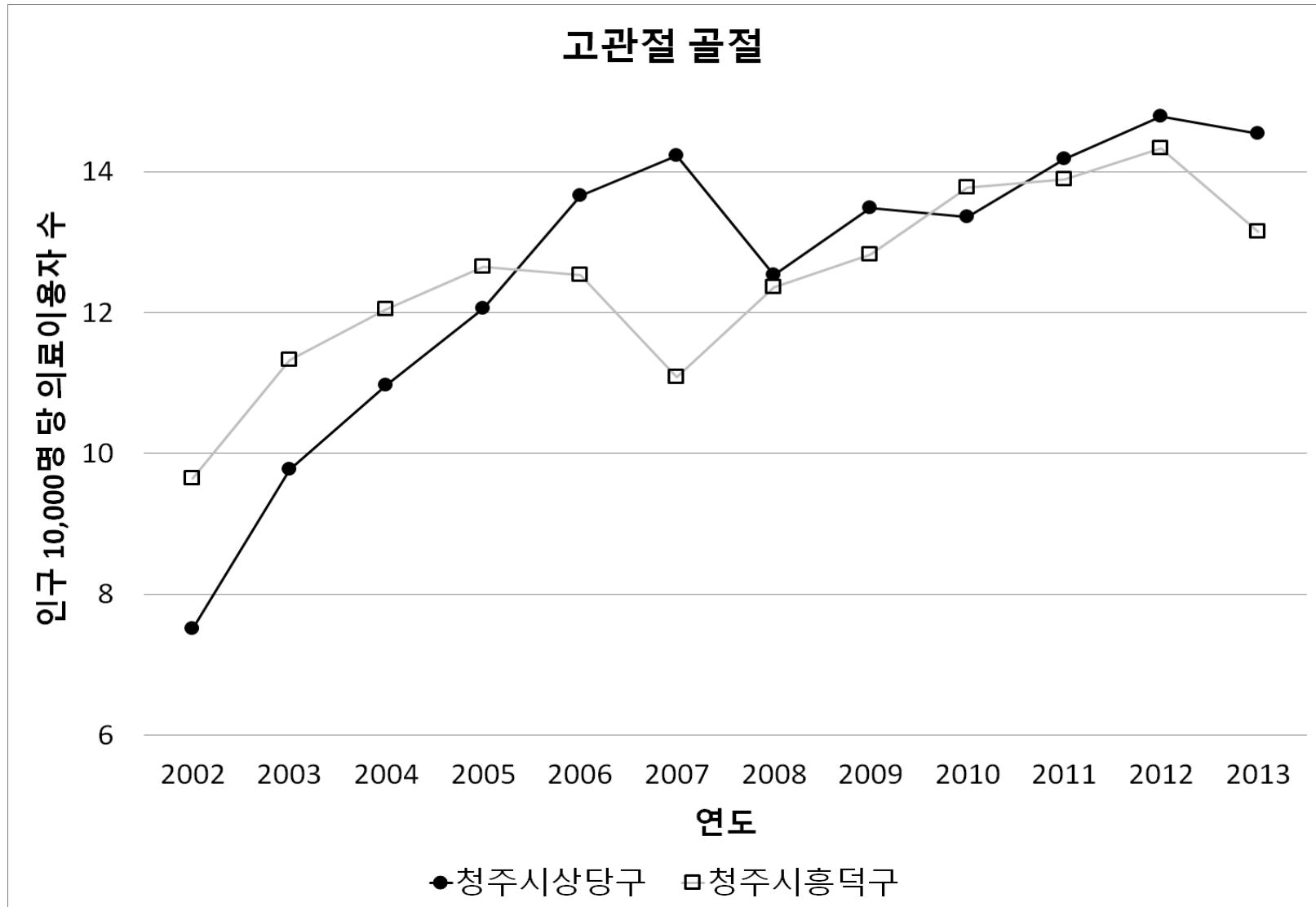
박탈지수

E  
P  
I  
D  
E  
M  
I  
O  
L  
O  
G  
Y



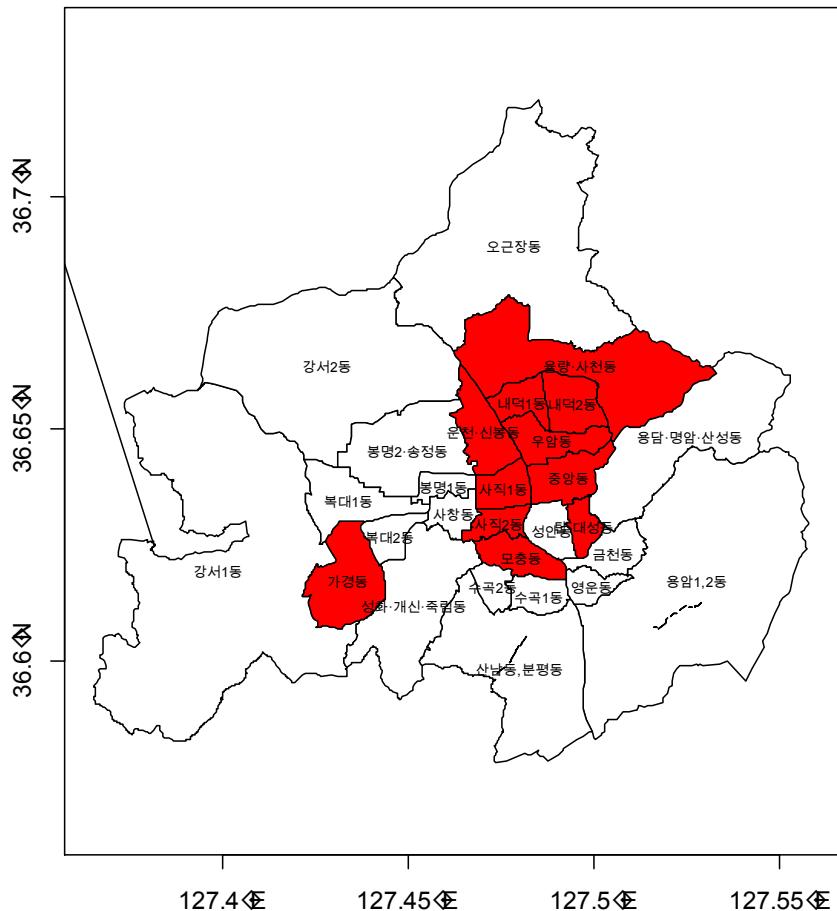
# 수돗물 불소농도조정사업의 건강영향 역학 연구

## 고관절 골절

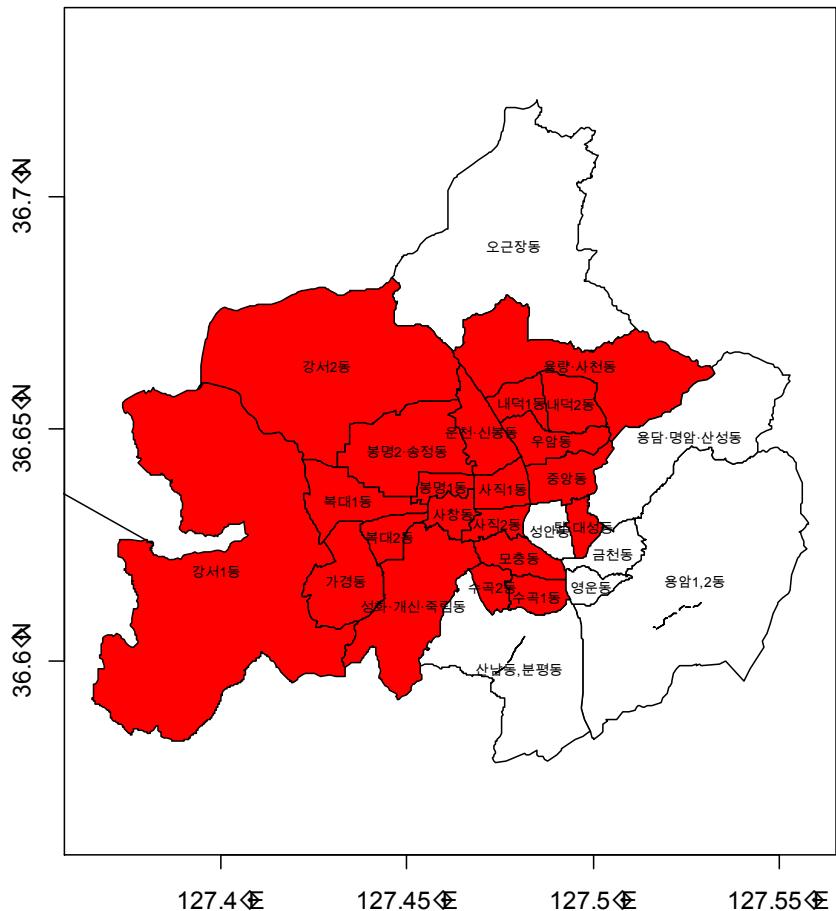


# 청주 상당구-흥덕구 불소화 시행 지역 읍면동\*

시행기간 1982-2004

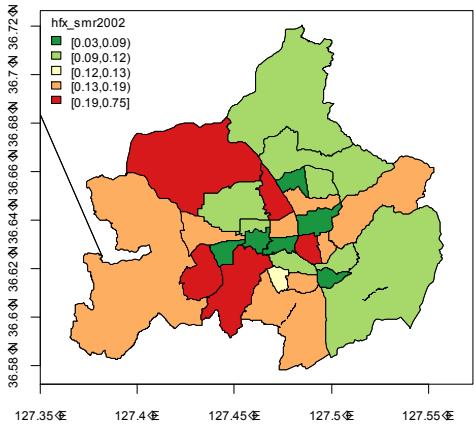


시행기간 1997-2004

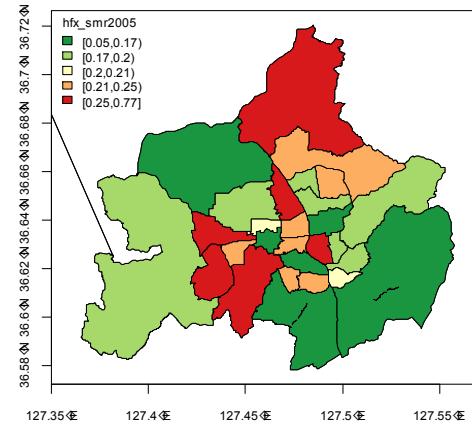


# 청주시 읍면동별 고관절 골절 표준화비

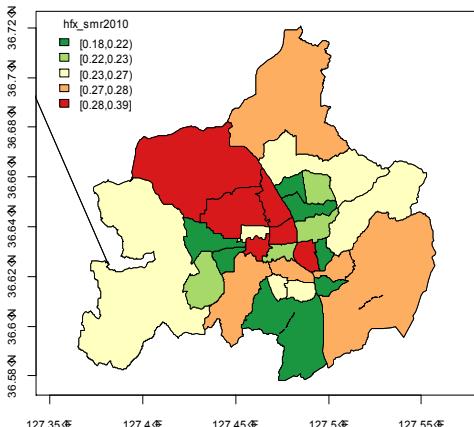
2002년



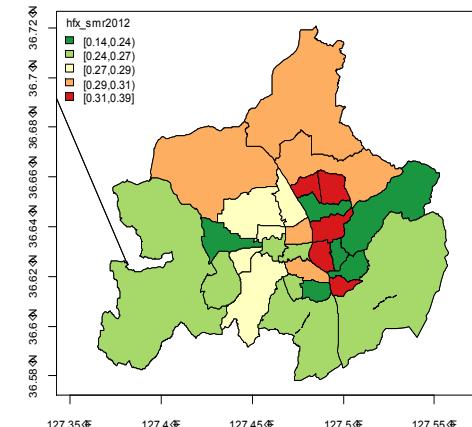
2005년



2010년



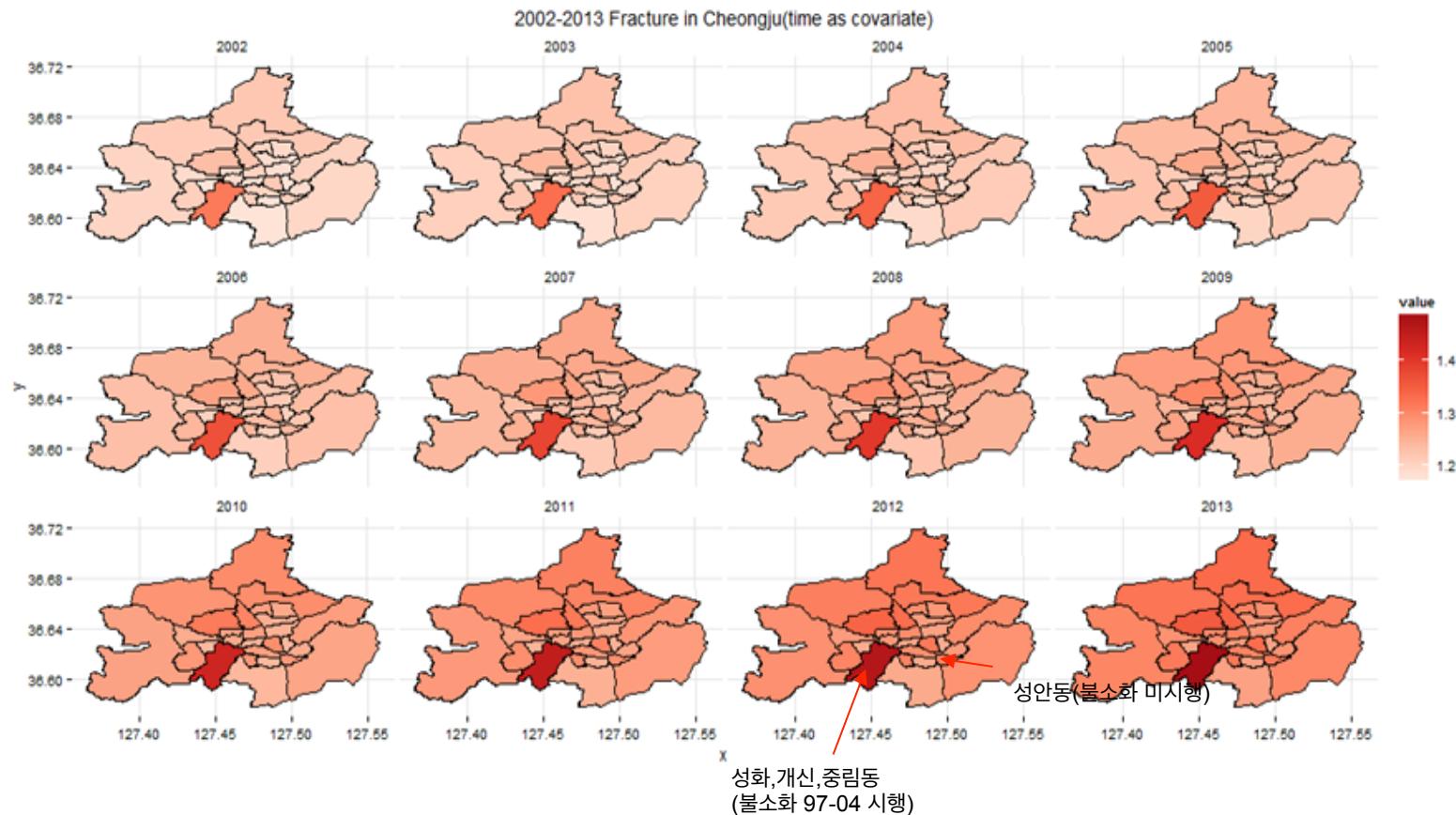
2012년



# 시공간 회귀분석

	시간 공변량			시간 평활화		
	RR	95% CrI		RR	95% CrI	
고관절골절	0.969	0.874	1.074	0.971	0.876	1.075
골다공증	0.953	0.872	1.041	0.954	0.874	1.043
골암	1.267	0.891	1.824	1.267	0.891	1.824
알츠하이머치매	1.007	0.889	1.141	1.007	0.889	1.141
정신지체	1.071	0.828	1.385	1.073	0.828	1.391

# 연도별 읍면동별 추정 표준화비(고관절 골절)

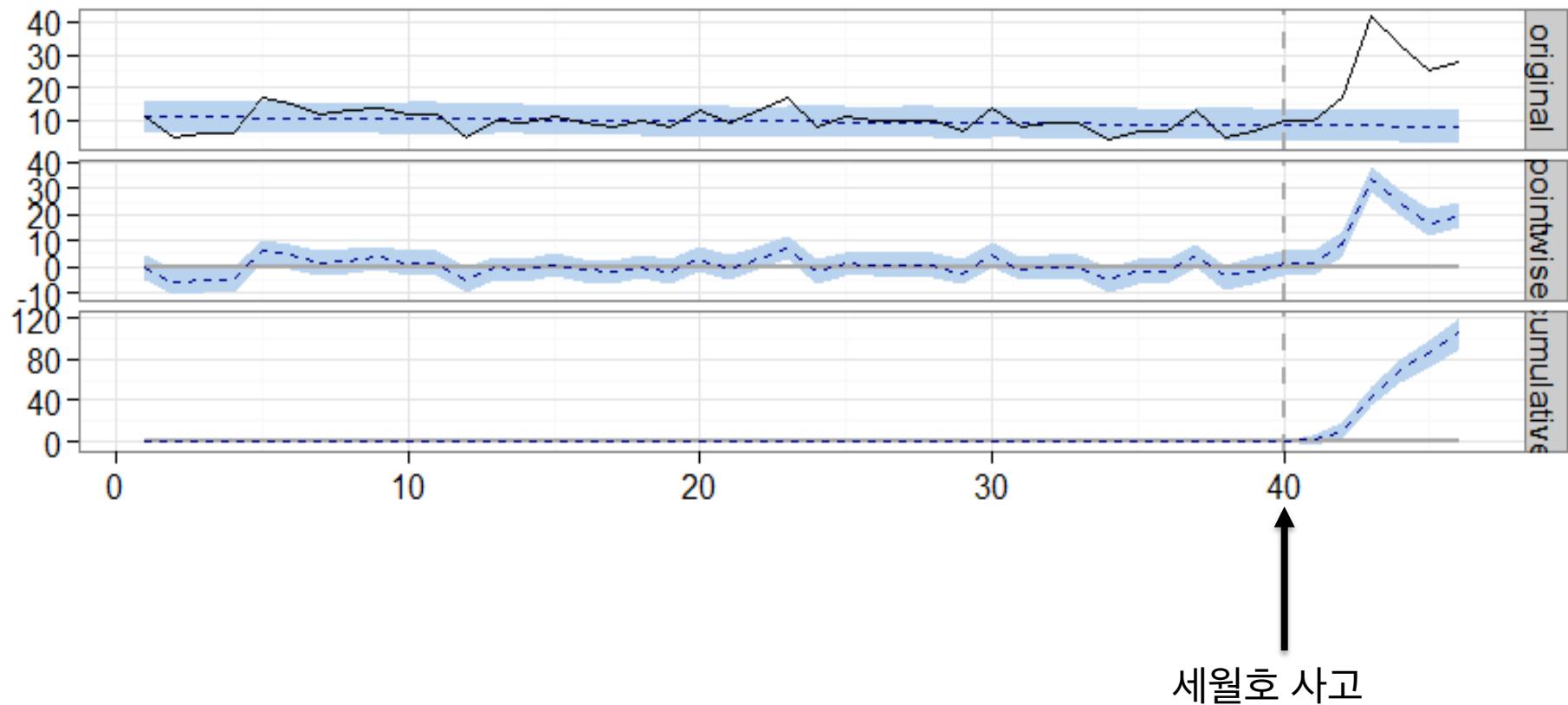




세월호 사고 이후  
안산시 의료이용 분석

# 연구결과

## -시계열: 우울증 응급(단원구)



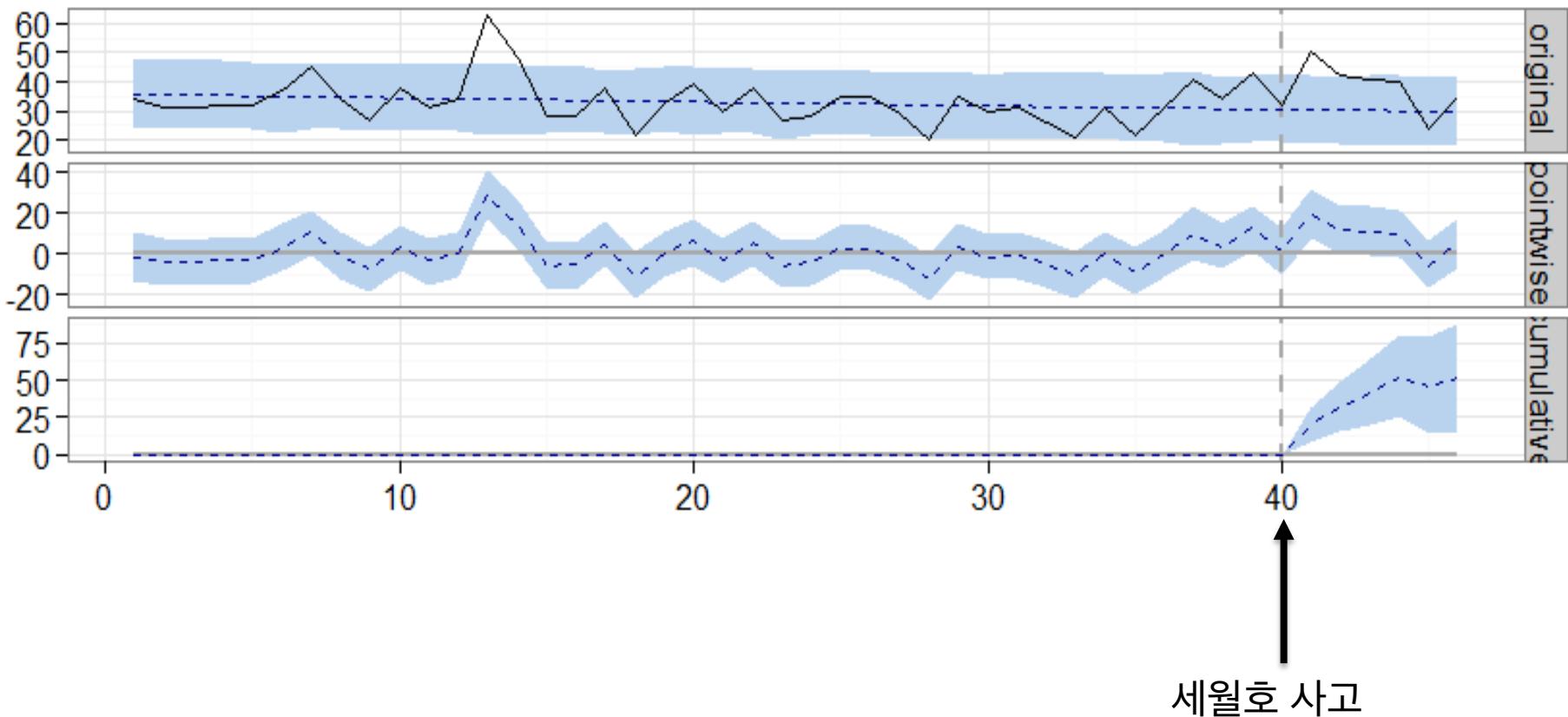
# 연구결과

## -시계열: 우울증 응급(단원구)

```
## Posterior inference {CausalImpact}
##
##                                     Average          Cumulative
## Actual                           26              155
## Prediction (s.d.)                8.3 (1.3)      49.9 (7.8)
## 95% CI                          [5.7, 11]       [34.3, 65]
##
## Absolute effect (s.d.)           18 (1.3)       105 (7.8)
## 95% CI                          [15, 20]        [90, 121]
##
## Relative effect (s.d.)           211% (16%)     211% (16%)
## 95% CI                          [180%, 242%]   [180%, 242%]
##
## Posterior tail-area probability p: 0.00111
## Posterior prob. of a causal effect: 99.88901%
```

# 연구결과

## -시계열: 협심심근경색 응급(단원구)

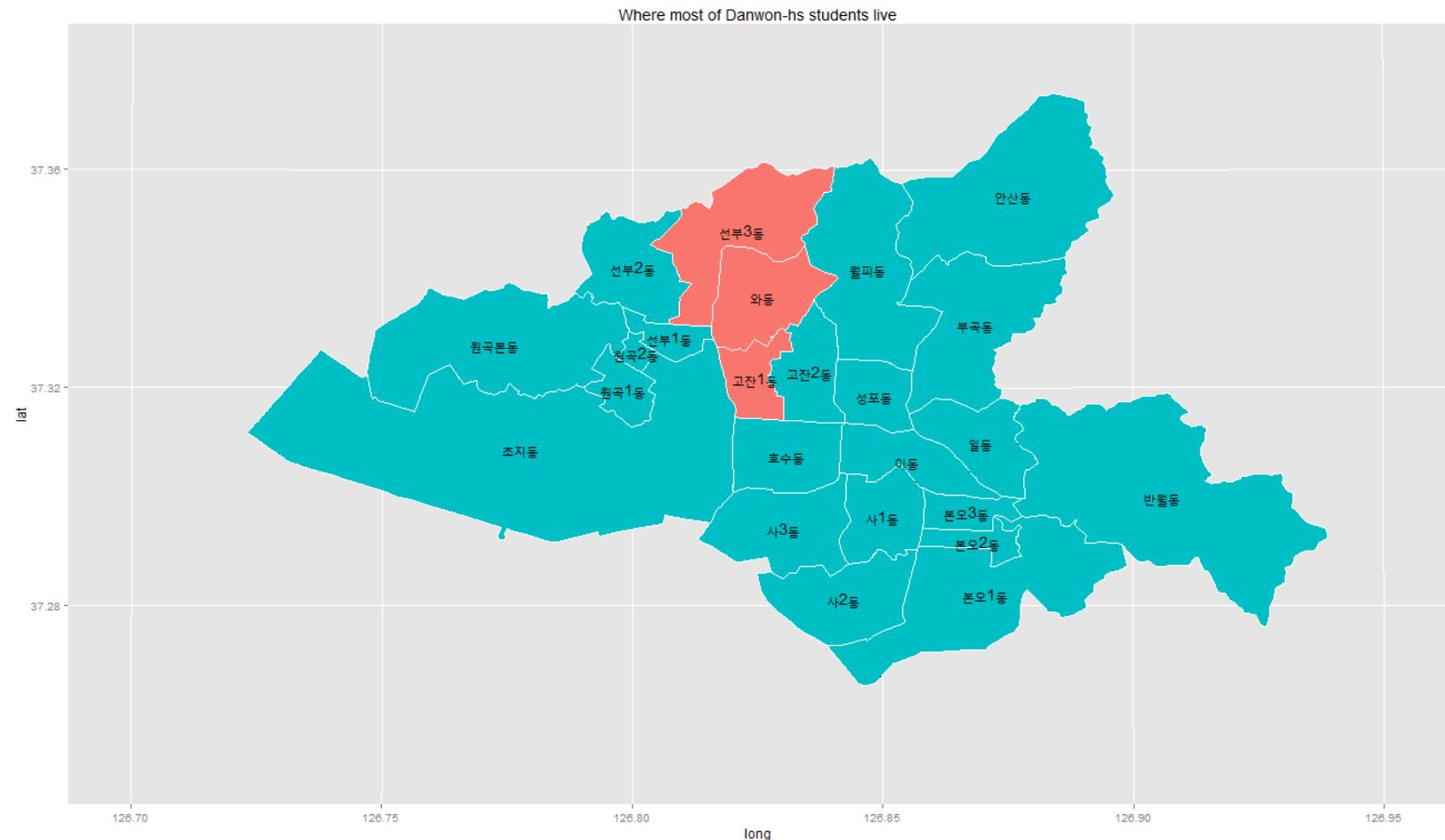


# 연구결과

## -시계열: 협심심근경색 응급(단원구)

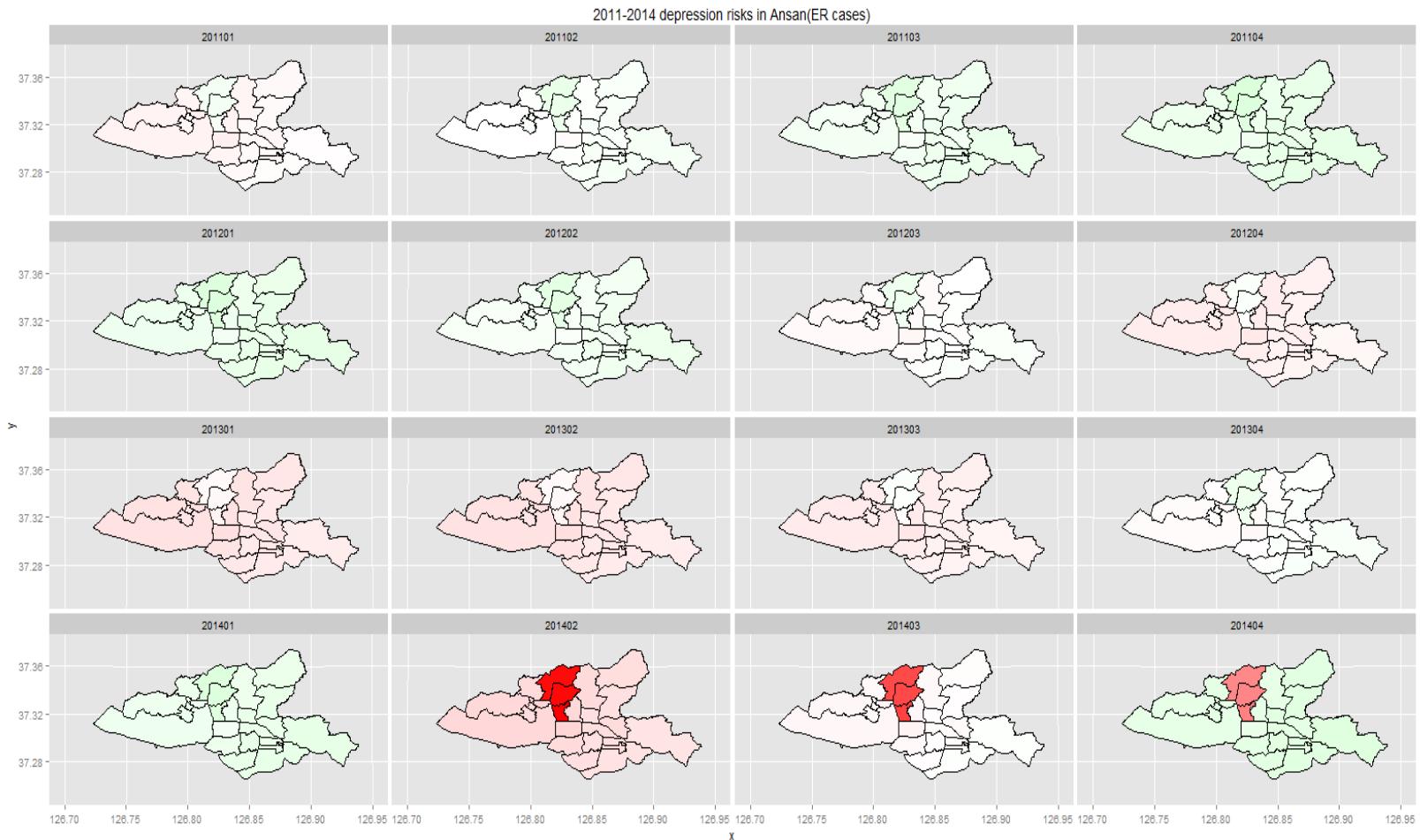
```
## Posterior inference {CausalImpact}
##
##                                     Average      Cumulative
## Actual                           39          232
## Prediction (s.d.)                30 (3.1)    181 (18.3)
## 95% CI                            [24, 36]    [144, 218]
##
## Absolute effect (s.d.)           8.6 (3.1)   51.4 (18.3)
## 95% CI                            [2.3, 15]   [13.7, 88]
##
## Relative effect (s.d.)           28% (10%)   28% (10%)
## 95% CI                            [7.6%, 49%] [7.6%, 49%]
##
## Posterior tail-area probability p: 0.00444
## Posterior prob. of a causal effect: 99.55605%
##
## For more details, type: summary(impact, "report")
```

# 안산시 동별 지도



# 연구결과

-2011~2014년 분기별 안산시 우울증 응급 상대위험도 추정 지도



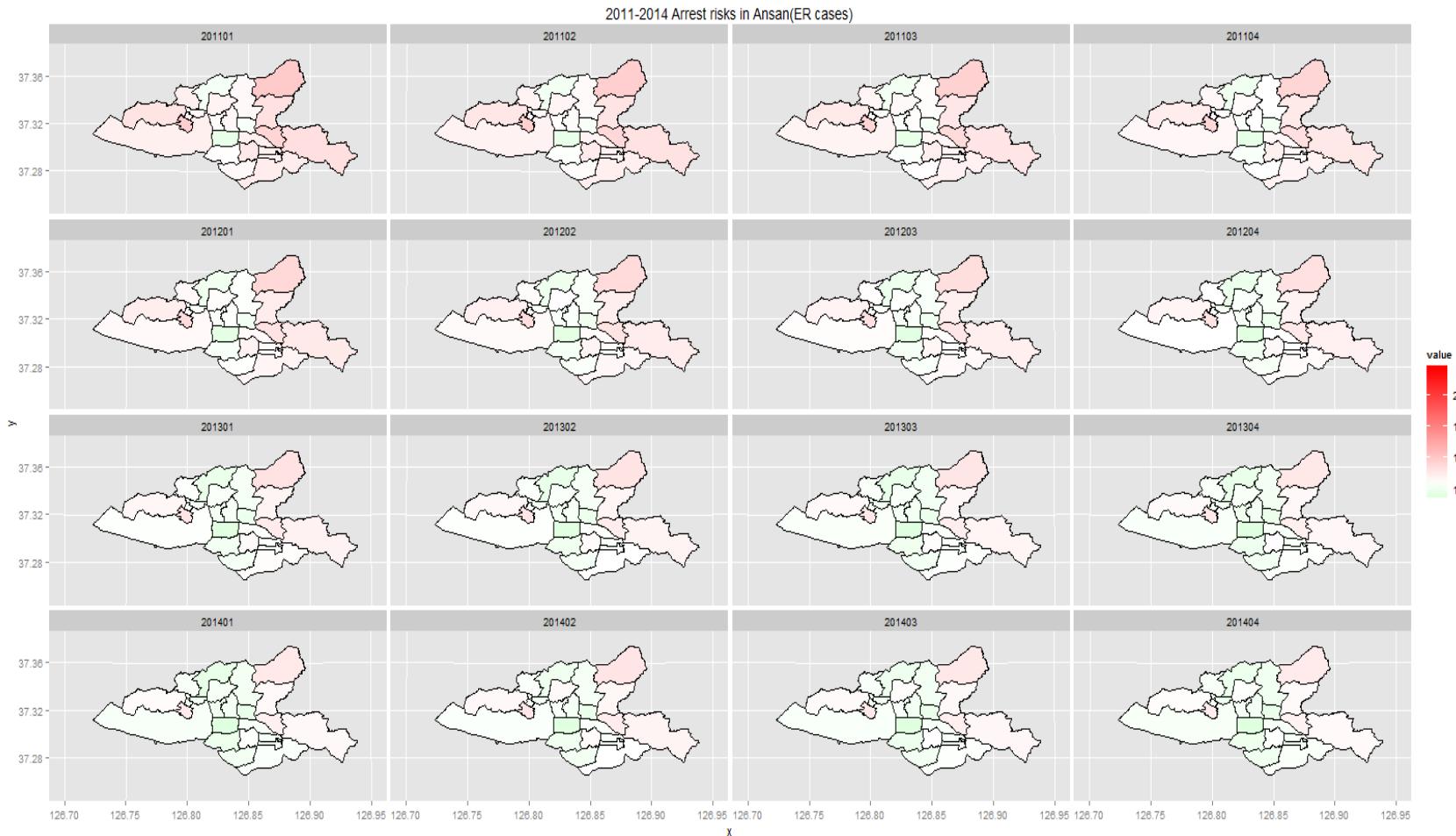
# 연구결과

-2011~2014년 분기별 안산시 우울증 응급 시공간 회귀분석 결과

우울증	mean	sd	0.025quant	0.5quant	0.975quant
(절편)	-1.4819	0.0745	-1.6315	-1.4797	-1.3445
단원고 지역 3개동	-0.0922	0.1307	-0.3568	-0.0895	0.1572
세월호 사고 이후	0.3481	0.3001	-0.1776	0.3356	0.9462
상호작용항	0.6393	0.2344	0.1725	0.6415	1.0934

# 연구결과

-2011~2014년 분기별 안산시 협심심근경색증 응급 상대위험도 추정 지도



# 연구결과

-2011~2014년 분기별 안산시 협심심근경색증 응급 시공간 회귀분석 결과

협심심근경색증	mean	sd	0.025quant	0.5quant	0.975quant
(절편)	-1.4050	0.0494	-1.5025	-1.4050	-1.3073
단원고 지역 3개동	-0.1461	0.1375	-0.4178	-0.1463	0.1267
세월호 사고 이후	0.0628	0.0815	-0.0977	0.0626	0.2237
상호작용항	0.1229	0.1279	-0.1322	0.1242	0.3703

### **3. R을 이용한 시공간 역학 분석**

### 3. R을 이용한 지도화

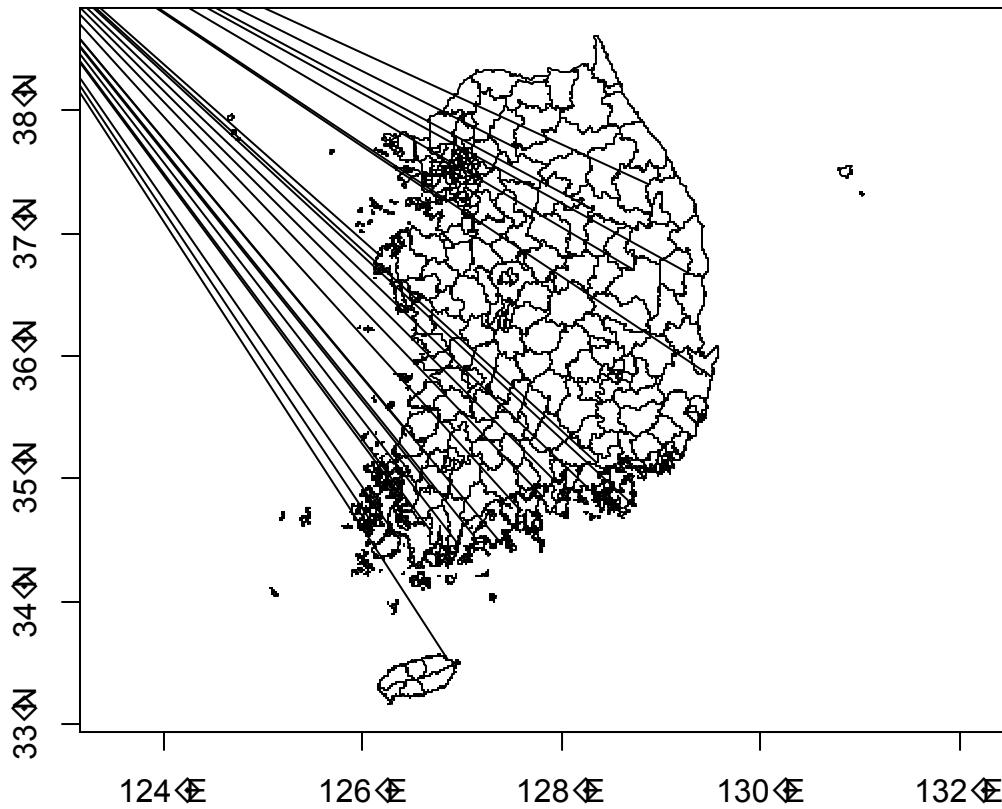
1. 지도 불러오기
2. 데이터 불러오기
3. mapping할 변수 선정 및 채색을 위한 작업
4. Mapping
5. 세부설정

# 사전 작업

- R에서 지도 관련 작업을 하기 전에 다음 패키지들을 사전에 설치해야 함.
  - > `install.packages(c("rgdal", "maptools", 'RColo  
rBrewer', 'classInt', 'spdep', 'DCluster', 'Spatial  
Epi', 'epiR', 'spgwr', 'sp', 'Hmisc', 'ggmap', 'ggp  
lot2', 'plyr', 'reshape2', 'RColorBrewer', 'scales'  
, 'geos'))`

# 지도 .shp 파일 불러오기

```
# 작업폴더 지정  
setwd("C:\\prevent2014")  
  
# reading shapefiles  
map <- readShapePoly("sigungu2009.shp", proj4string=CRS("+proj=t  
merc +lat_0=38 +lon_0=127 +k=1 +x_0=200000 +y_0=500000 +ellps  
=GRS80 +units=m +no_defs"))  
  
# TM Mercator 단위 -> 위-경도 단위  
map <- spTransform(map, CRS=CRS("+proj=longlat +lat_0=38 +lon_  
0=127 +k=1 +x_0=200000 +y_0=500000 +ellps=WGS84 +units=m +  
no_defs"))  
  
# 지도 로딩  
plot(map, axes=T)  
  
# 데이터와 연동시키기 전, 변수 형태를 똑같이 숫자변수로  
map@data$PHC_NUM =as.numeric(map@data$PHC_NUM)
```



# 데이터 불러오기 및 합치기

```
# 데이터 불러오기  
dat <- read.csv("reg_data.csv")  
dat = data.frame(dat)  
  
# 데이터 체크  
names(dat)  
str(dat)  
head(dat)  
  
# 데이터 merge  
map@data <- data.frame(map@data, dat[match(map@data[, "PHC_NUM"],  
                                         dat[, "PHC_NUM"])] )  
  
# merge()함수는 지도 폴리곤 순서를 무시하는 문제가 있음  
# 데이터 자체를 붙이는 데는 문제없음.  
#map@data<-merge(x=map@data, y=dat, by="PHC_NUM")
```

# Mapping 1

#### 그림그릴 데이터 plotvar라는 이름으로 지정

```
plotvar <- map@data$heart
```

#### 5등분.

```
nclr <- 5
```

# 각 변수를 위에서 설정된 5분위(quantile)로 나눔

```
class <- classIntervals(plotvar, nclr, style="quantile")
```

#### 지도 Legend에 나오는 자릿수를 소수점 2자리로 정리

```
brks <- round(class$brks, 2)
```

```
class <- classIntervals(plotvar, nclr, style = "fixed", fixedBreaks = brks)
```

# 위의 5분위 각각에 대응하는 색 지정하기.

```
plotclr <- brewer.pal(nclr, "RdYIGn")
```

# 색 순서를 역순으로

```
plotclr=plotclr[nclr : 1]
```

# Mapping 2

```
#### 각 지도 폴리곤별 색 코드 지정
```

```
colcode <- findColours(class, plotclr, digits=0)
```

```
png('map.png', width = 960, height = 960, units = "px", pointsize = 20 )
```

```
# 지도를 직접 안보고 그림파일로 저장하고 싶으면 plot 앞뒤에 png(), dev.off()
```

```
# 이때, 경로는 앞에서 setwd(' ')로 지정한 경로
```

```
# 백지 상태 지도
```

```
plot(map, density=16, col="grey", axes=T, cex.axis=.75)
```

```
# 색 입히기
```

```
plot(map, col=colcode, axes=T, add=T)
```

```
# legend에서 지도 모양에 따라 'topleft', 'topright', 'bottomleft', 'bottomright', 'center'
```

```
legend("bottomright", legend=names(attr(colcode, "table")), title='heart'
```

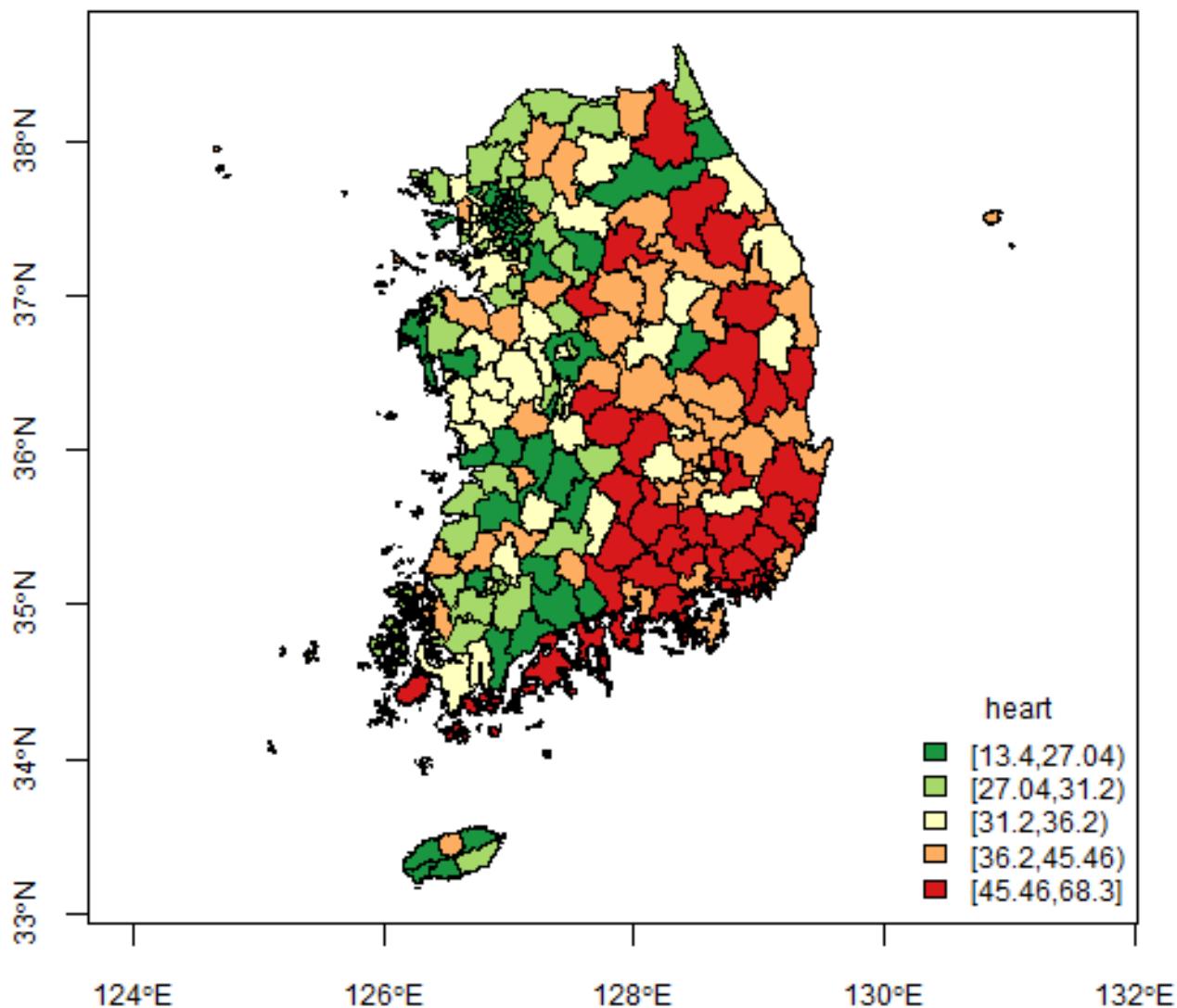
```
fill=attr(colcode, "palette"), cex=0.75, bty="n")
```

```
title("심장질환 사망률")
```

**dev.off()**

**# 지도 완성된 후 해당 경로에 저장완료**

## 심장질환 사망률



# ggplot2를 이용한 mapping

# 앞에서 map이란 이름으로 shapefile 불러왔음을 가정함.

# fortify는 지도 shapefile을 ggplot2에 맞게 그리기 위해 폴리곤 각 점의 경, 위도를 계산함.

```
map_ft<- fortify(map, region="PHC_NUM")
```

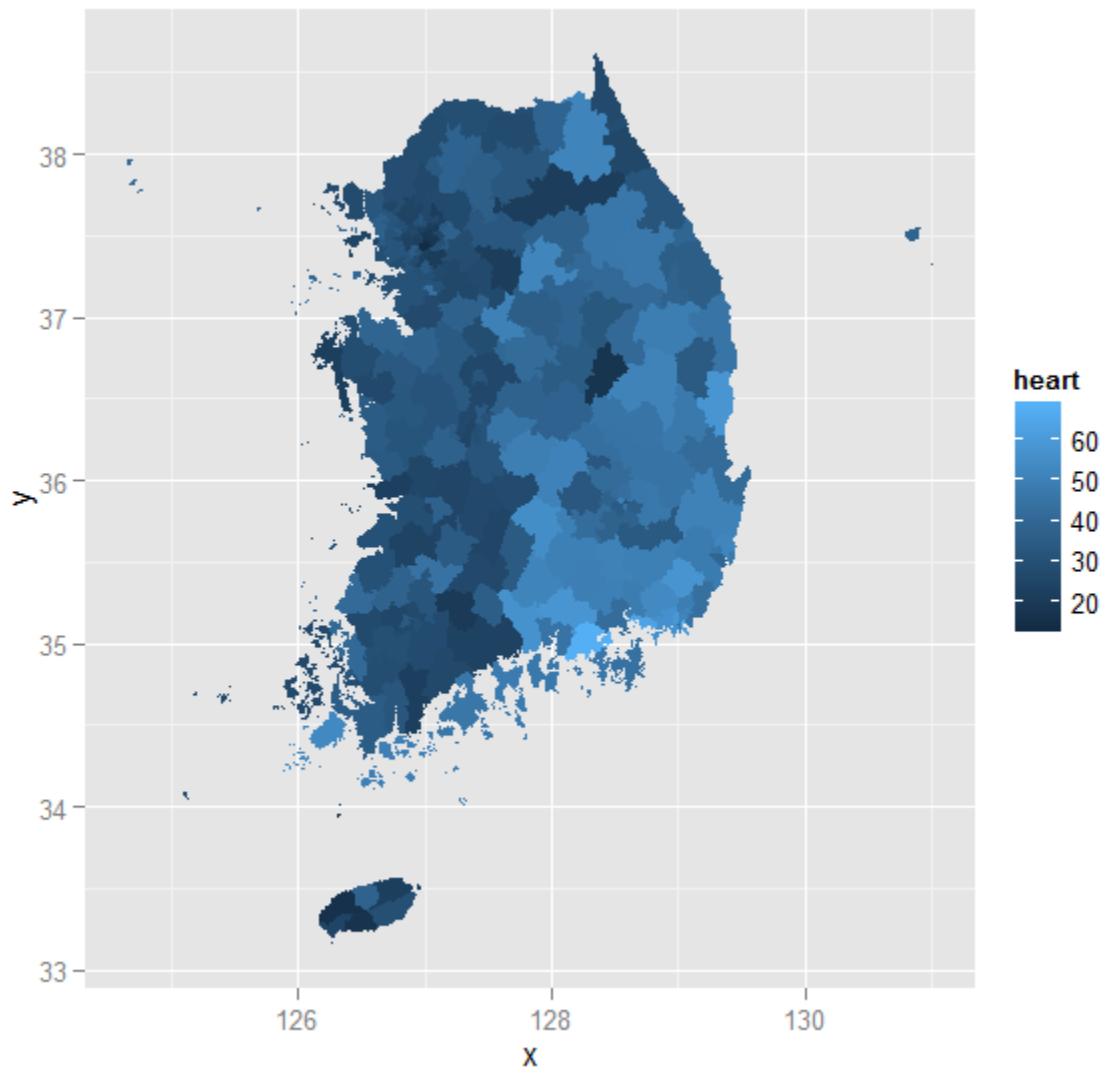
# fortify 명령어 다음부터 plot(map\_ft) 쓰지 말 것. 컴퓨터가 느려짐.

```
ggplot() +  
  geom_map(data = map@data, aes(map_id = PHC_NUM, fill = heart)  
, map = map_ft) +  
  expand_limits(x = map$long, y = map$lat)
```

# 지도id(시군구)=데이터id(시군구) 일치한다는 가정에서

# geom\_map(data=데이터, aes(map\_id=ID변수, fill=매핑변수), map=지도)

# expand\_limits는 지도가 나타날 범위 지정.



# 옵션 조정을 통한 그림 색 바꿈

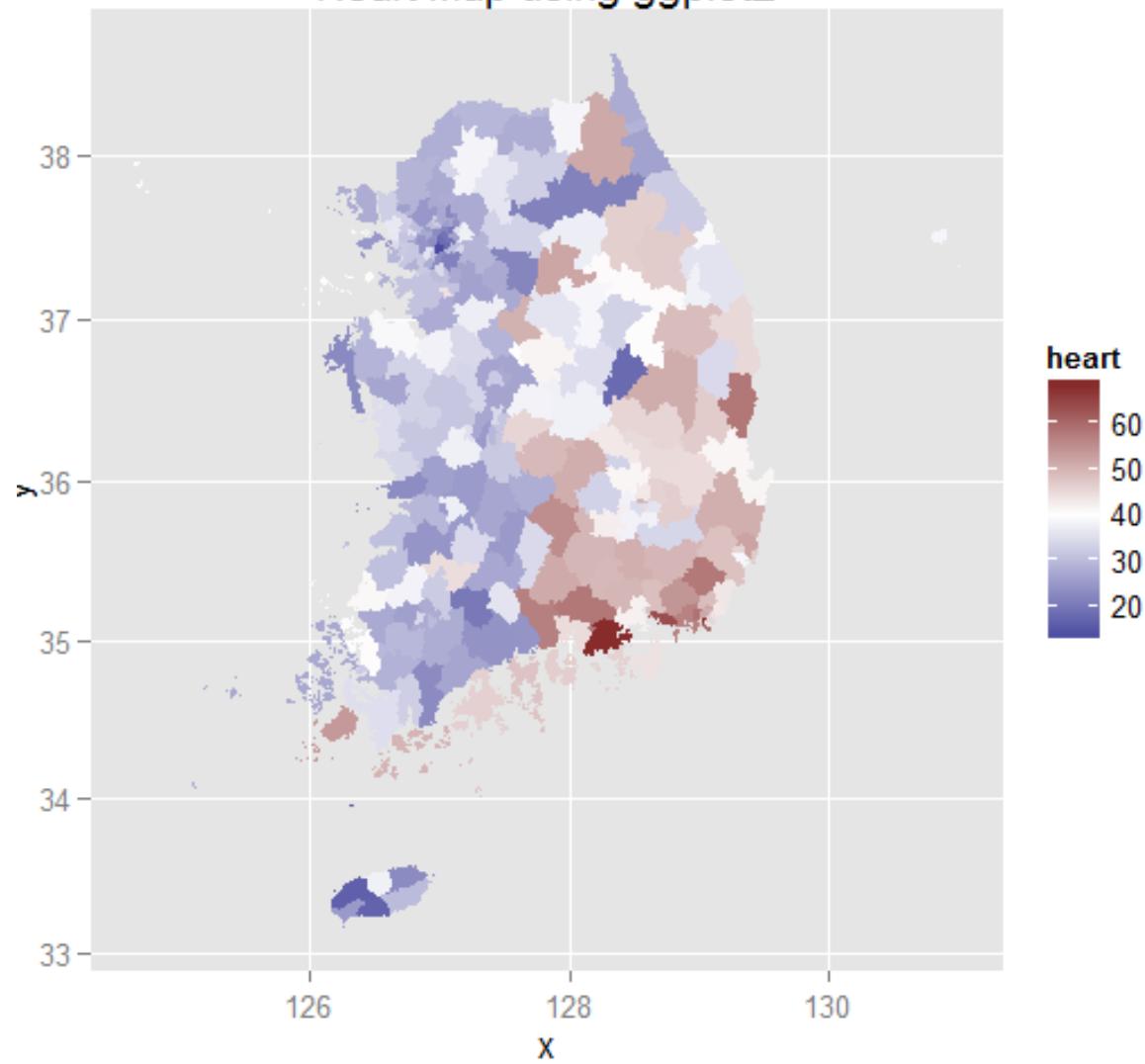
```
ggplot() +  
  geom_map(data = map@data, aes(map_id = PHC_NUM, fill = heart), map =  
    map_ft) +  
  expand_limits(x = map$long, y = map$lat) +  
  scale_fill_gradient2(low = muted("green"), mid = "white", midpoint = 24, high  
    = muted("red"), limits = c(15, 32)) +  
  coord_map(projection="mercator") +  
  labs(title = "Heart Map using ggplot2")
```

```
# scale_fill_gradient2(low=낮은값 색, mid=중간색, midpoint=중간값, high=높은값, li  
mit=c(최소값, 최대값) )
```

# muted("색")을 쓰면, 원래 색보다 약간 탁해짐.

# coord\_map(projection="mercator")은 지도를 확대해도 가로세로 비율 일정.

## Heart Map using ggplot2



# 경계선 그리기

```
ggplot() +  
  geom_map(data=map@data, aes(map_id=PHC_NUM, fill=heart),  
map=map_ft) +  
  geom_path(data=map_ft, aes(x=long, y=lat, group=group), colour=  
"black", size=0.25) +  
  coord_map(projection="mercator") + scale_fill_gradient2(limits = c(  
15, 32), low = muted("green"), mid = "white", midpoint = 25, high =mu  
ted("red"))  
  
# geom_path(data=지도, aes(x=long, y=lat, group=group), colour="경  
계선 색", size=0.25)
```

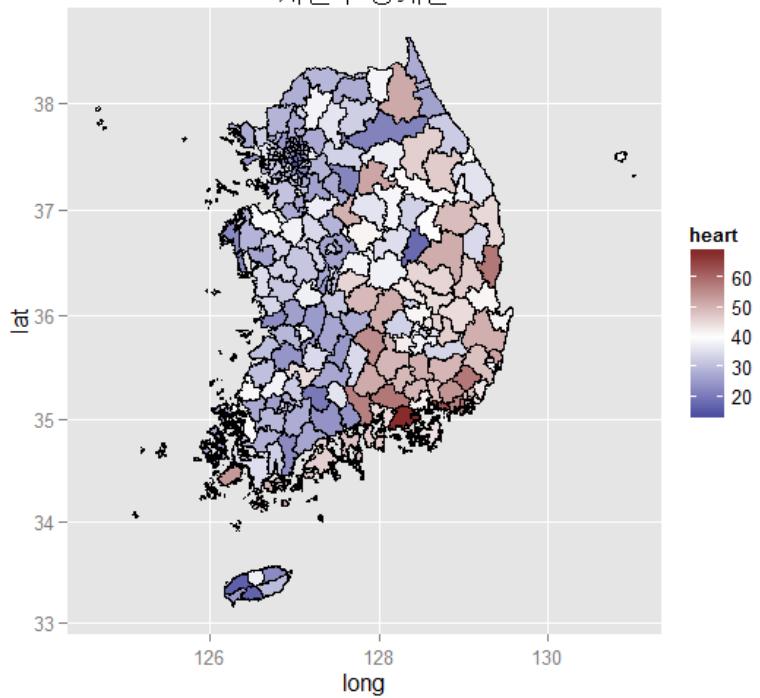
# 시군구 경계가 아닌 시도 경계

```
map1 <- readShapePoly("sigungu2009.shp", proj4string=CRS("+proj=tmerc +la_t_0=38 +lon_0=127 +k=1 +x_0=200000 +y_0=500000 +ellps=GRS80 +units=m +no_defs"))
map1 <- spTransform(map1, CRS=CRS("+proj=longlat +lat_0=38 +lon_0=127 +k=1 +x_0=200000 +y_0=500000 +ellps=WGS84 +units=m +no_defs"))
map1@data$PHC_NUM =as.numeric(map1@data$PHC_NUM)
colnames(map1@data)=c('sido_2009','gungu_2009','PHC_NUM')
map1_ft <- fortify(map1, region="sido_2009")
```

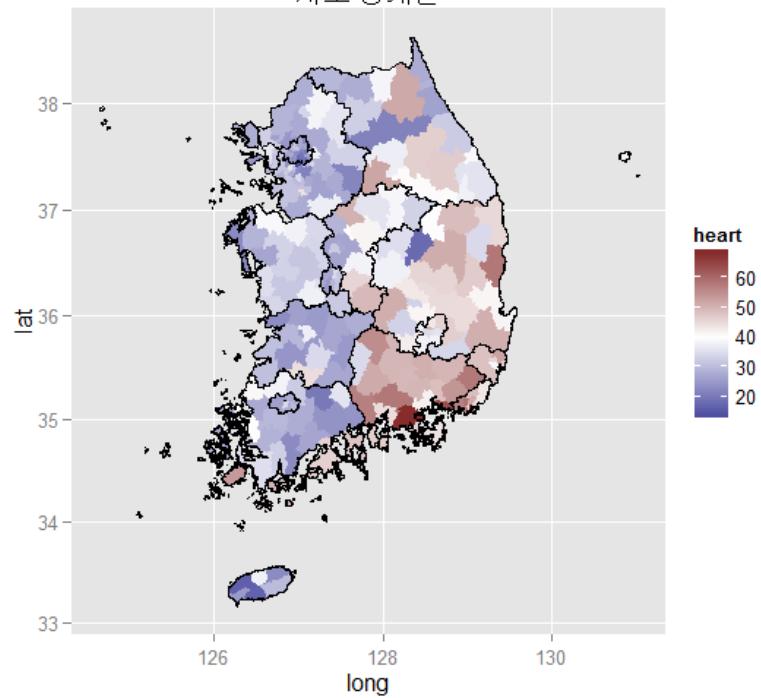
# 시도를 한번 더 불러오고 fortify에서 폴리곤을 시도단위로 인식.

```
ggplot() +
  geom_map(data=map@data, aes(map_id=PHC_NUM, fill=heart), map=map)
  + geom_path(data=map1_ft, aes(x=long, y=lat, group=group), colour="black", size=0.25) +
    coord_map(projection="mercator") + scale_fill_gradient2(limits = c(15, 32), low = muted("green"), mid = "white", midpoint = 25, high =muted("red"))
```

시군구 경계선



시도 경계선



## 시군구 경계선과 시도 경계선

# ggmap에서 실제 지도를 불러오기

# 앞에서 지도를 불러와서 map이란 이름으로 저장했다는 전제.  
# 또다시 할 필요는 없음.

```
map <- readShapePoly("sigungu2009.shp", proj4string=CRS("+proj=tm  
erc +lat_0=38 +lon_0=127 +k=1 +x_0=200000 +y_0=500000 +ellps=GR  
S80 +units=m +no_defs"))  
map <- spTransform(map, CRS=CRS("+proj=longlat +lat_0=38 +lon_0=  
127 +k=1 +x_0=200000 +y_0=500000 +ellps=WGS84 +units=m +no_d  
efs"))  
  
map@data$PHC_NUM =as.numeric(map@data$PHC_NUM)
```

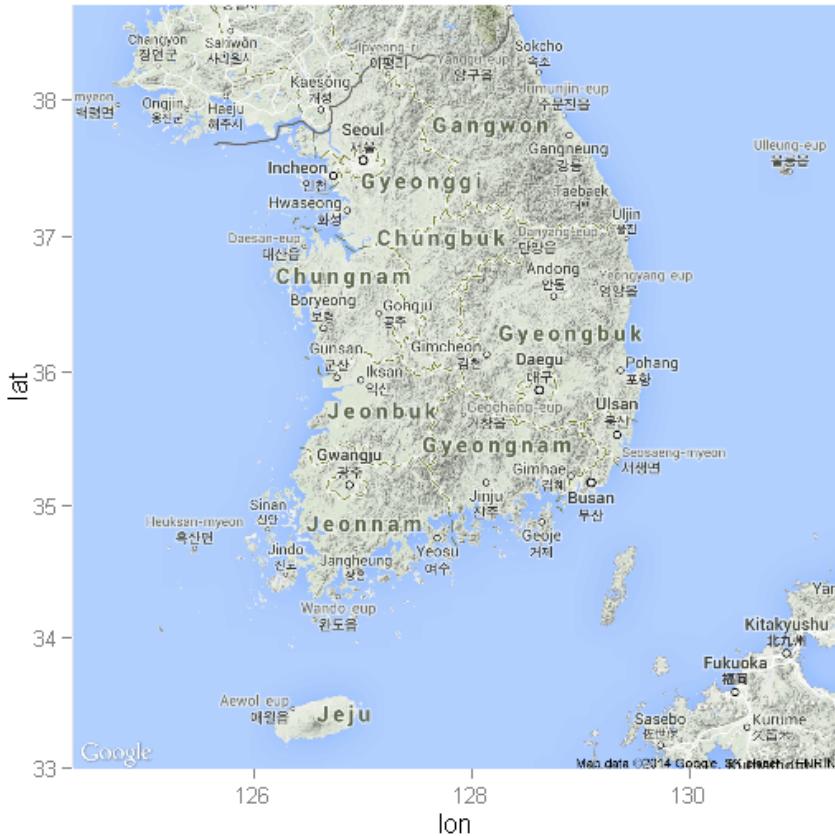
```
krmap = get_map(location=as.vector(bbox(map)))
```

```
krmap2 = get_map(location=as.vector(bbox(map)), source="stamen", maptype="toner", zoom=7)
```

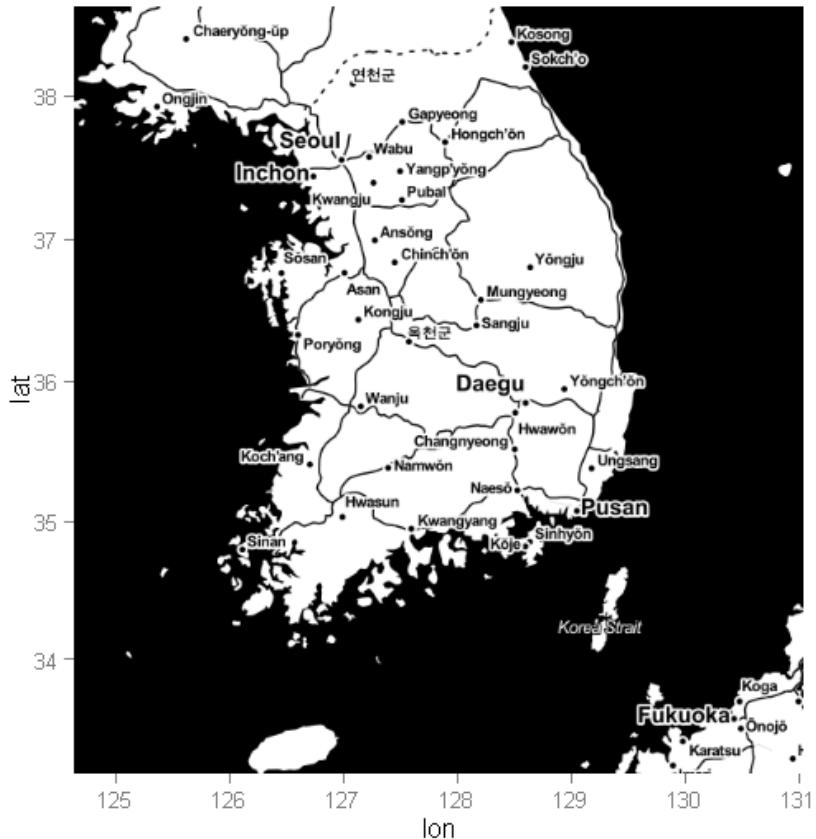
```
krmap3 = get_map(location=as.vector(bbox(map)), maptype="satellite")
```

```
ggmap(krmap)  
ggmap(krmap2)
```

# get\_map(location=앞 지도에 해당하는 지역좌표,  
# bbox(map)은 지도 구역에 대한 경위도 좌표값을 2차원 행렬로 저장.  
#,source="stamen", maptype="toner"로 하면 지도가 흑백으로 나옴.  
# maptype="satellite"은 위성사진, 그 외 'terrain', 'roadmap', 'hybrid'



# krmap



## krmap2

# 데이터값을 점으로 찍기

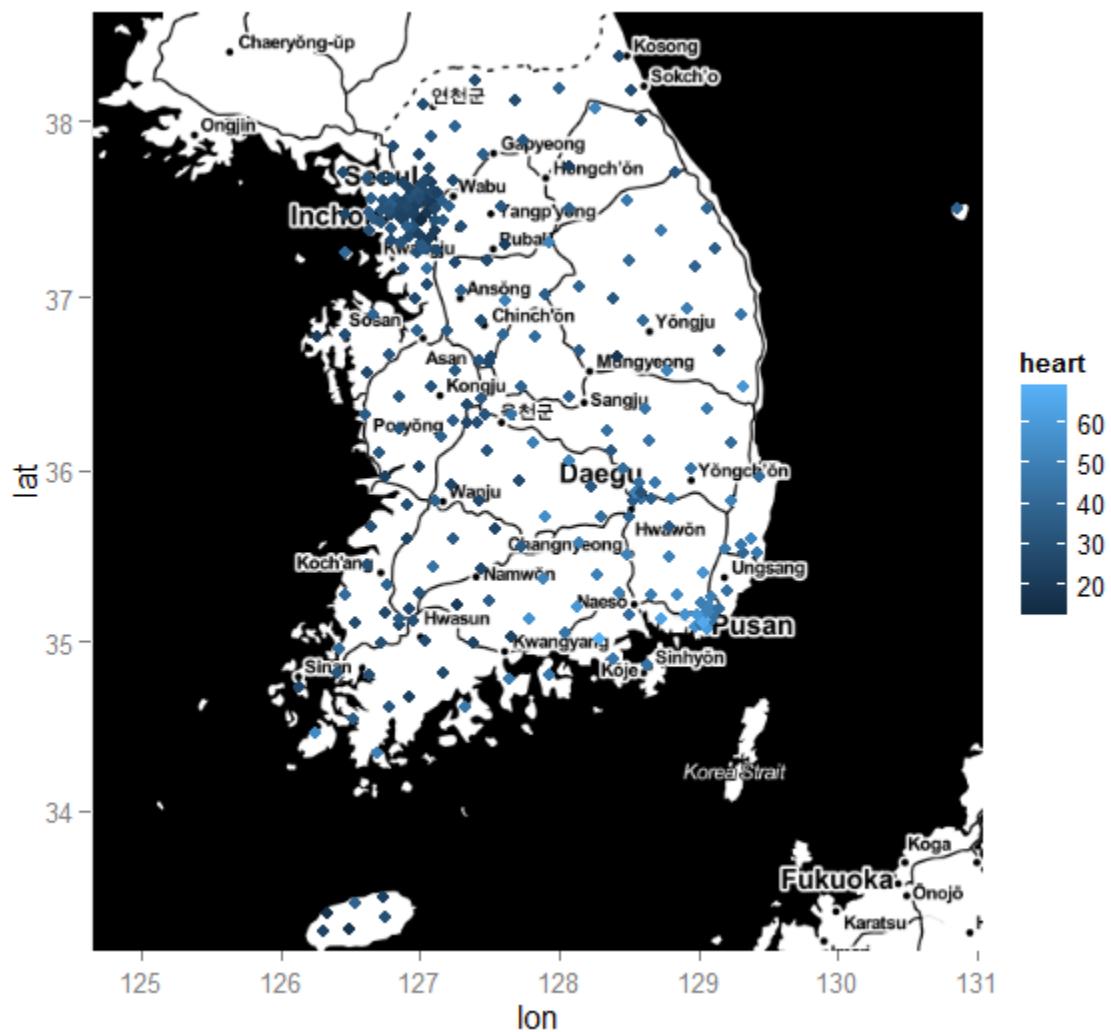
```
# 각 시군구 경, 위도 좌표와 id, 매팅변수로 이루어진 데이터셋 만들기
hrt = data.frame(map@data$PHC_NUM, coordinates(map), map@data$sido_2009, map@data$gungu_2009, map@data$heart)

head(hrt)
# 데이터셋 이름 지정
colnames(hrt)=c("PHC_NUM", "long", "lat", "sido_2009", "gungu_2009","heart")

# ggmap과 ggplot 비교

ggmap(krmap)+geom_point(aes(x=long,y=lat,col=heart), data=hrt)

ggplot()+geom_point(aes(x=long, y=lat, col=heart),data=hrt)
```



# 데이터값 mapping

```
myMap = geom_map(inherit.aes = FALSE, aes(fill=heart, map_id=PHC_NUM),  
, map=krFort, data=hrt)
```

```
Limits = expand_limits(x=krFort$long,y=krFort$lat)
```

```
# 아까 그림과 동일
```

```
ggplot() + myMap + Limits
```

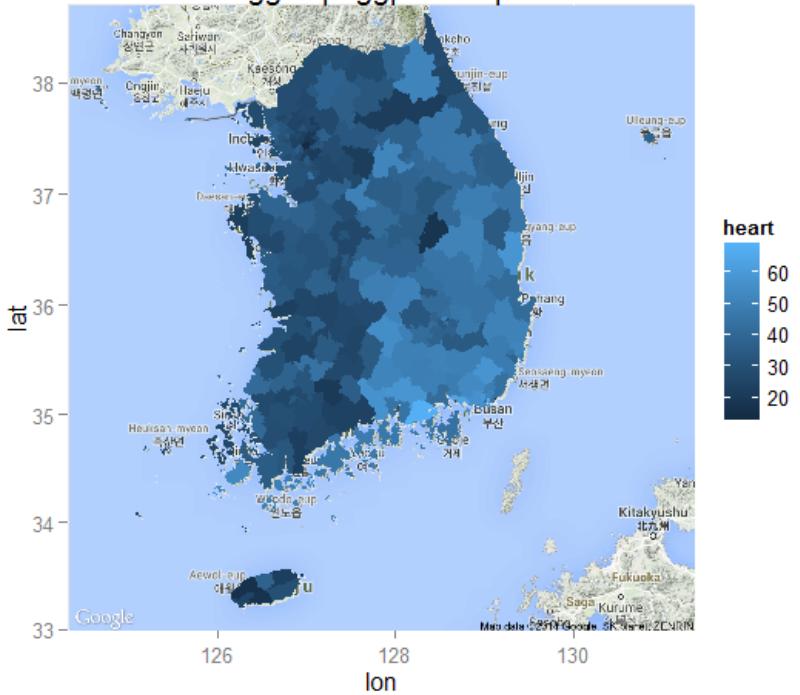
```
# ggmap에 써운 것
```

```
ggmap(krmap) + myMap + Limits + labs(title='ggmap+ggplot2map')
```

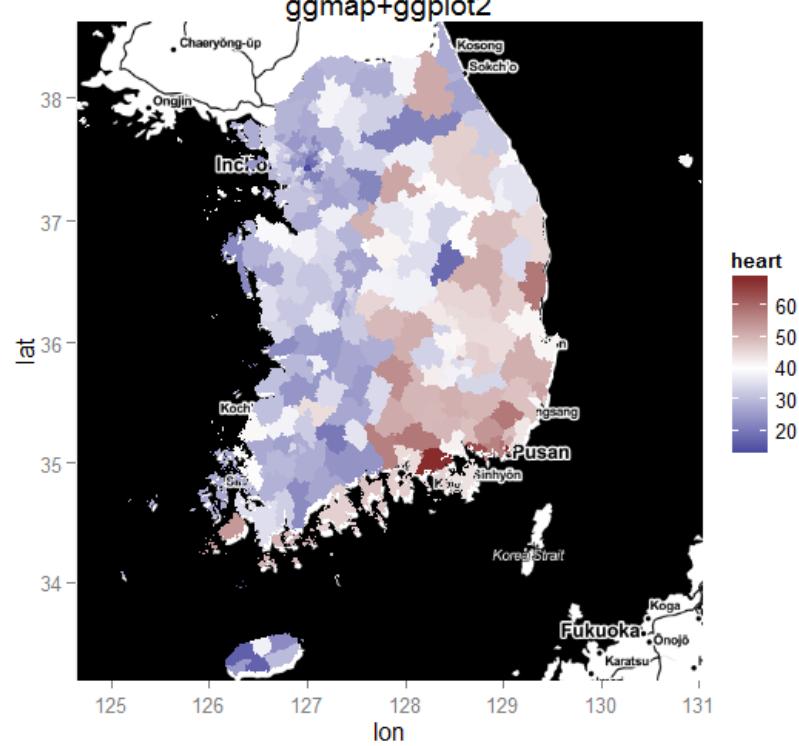
```
# 그림 옵션 추가
```

```
ggmap(krmap2) + myMap +  
coord_map(projection="mercator") + scale_fill_gradient2(limits = c(14,  
40), low = muted("red"), mid = "white", midpoint = 25, high = muted("blue"))  
+ Limits + labs(title='ggmap+ggplot2')
```

**ggmap+ggplot2map**



**ggmap+ggplot2**



# 3. R을 이용한 공간 군집 탐색

- Moran's I
- Getis-Ord global G statistic
- Mantel's permutation test

# 공간 가중치 파일 만들기

```
## 지도에서 gal 파일 만들기(파일명: test.gal)
```

```
temp=poly2nb(map)
```

```
write.nb.gal(temp, file="test.gal")
```

```
## geoda에서 만든 GAL 파일 불러오기#
```

```
# 파일 위치를 직접 찾아 마우스로 선택하고자 할 때 file.choose()
```

```
## poly.nlist <- read.gal(file.choose())
```

```
poly.nlist <- read.gal("test.gal")
```

```
summary(poly.nlist)
```

```
## spatial weight 만들기
```

```
## listw <- nb2listw(poly.nlist)
```

```
## 안되면?(섬 지역은 다른 곳과 떨어져 있어서 빈 공간이 존재)
```

```
listw <- nb2listw(poly.nlist, zero.policy=TRUE)
```

```
# map에 아까 만들었던 hrt(경, 위도, 데이터값 )과 merge
```

```
hrt1 = subset(hrt, select=c(PHC_NUM, heart))
```

```
map@data<-merge(x=map@data, y=hrt, by="PHC_NUM")
```

# 여러 종류 Moran 검정과 상관성 도표 출력

```
# Moran test
```

```
moran.test(map$heart, listw, zero.policy=T)
```

```
# Monte Carlo Moran Test
```

```
bperm=moran.mc(map$heart, listw, nsim=999 ,  
zero.policy=T)
```

```
bperm
```

R Gui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> moran.test(map@data$heart, listw, zero.policy=T)

  Moran's I test under randomisation

data: map@data$heart
weights: listw

Moran I statistic standard deviate = 13.6456, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
  0.560242025     -0.004081633     0.001710285

>
>
>
>
>
>
>
> |
```

R Gui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
>
> moran.mc(map@data$heart, listw, nsim=999 ,zero.policy=T)

Monte-Carlo simulation of Moran's I

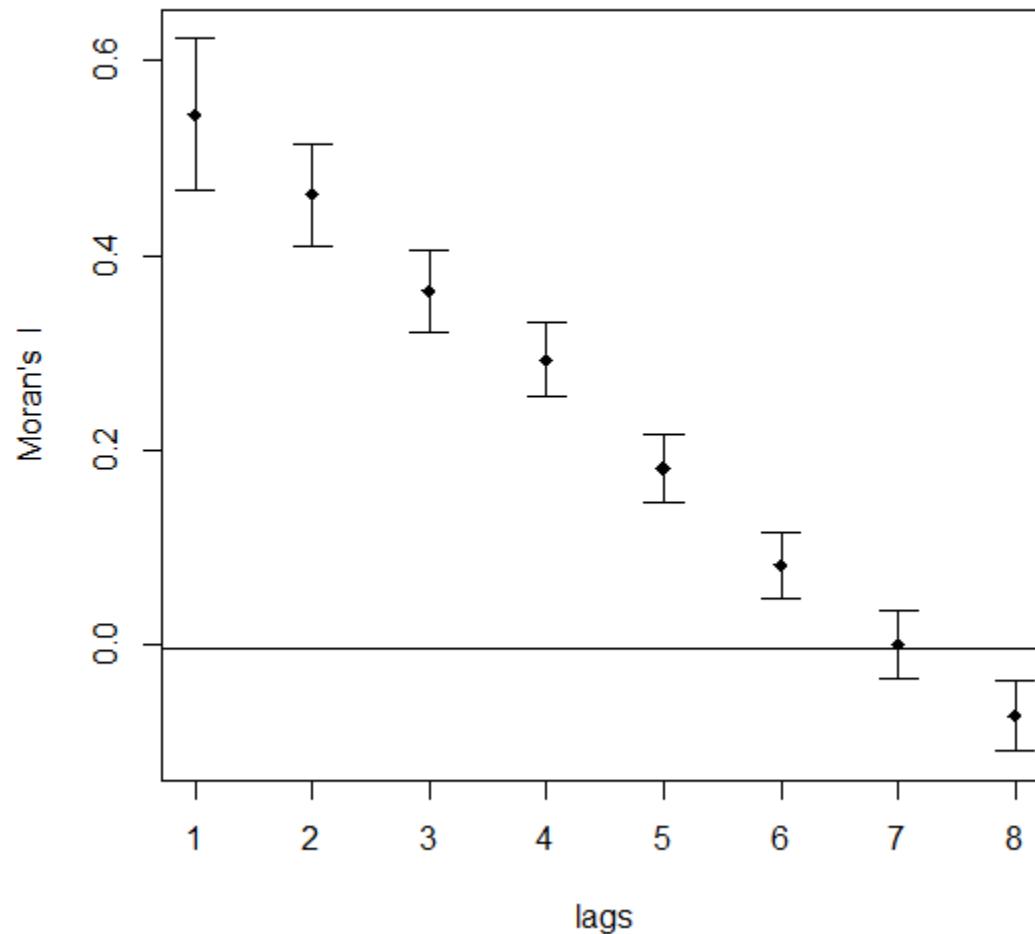
data: map@data$heart
weights: listw
number of simulations + 1: 1000

statistic = 0.5762, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater

>
>
>
>
>
>
>
> |
```

```
# Correlogram 출력  
# poly.nlist <- poly2nb(map, queen = FALSE)  
  
cor8 <- sp.correlogram(poly.nlist, var=map@data$heart, order=8, method="I", style= "C", zero.policy=T)  
  
plot(cor8)
```

map@data\$heart



# 그 외의 통계량

```
# Getis-Ord G-statistic
```

```
globalG.test(map$heart, listw = listw, zero.policy=T)
```

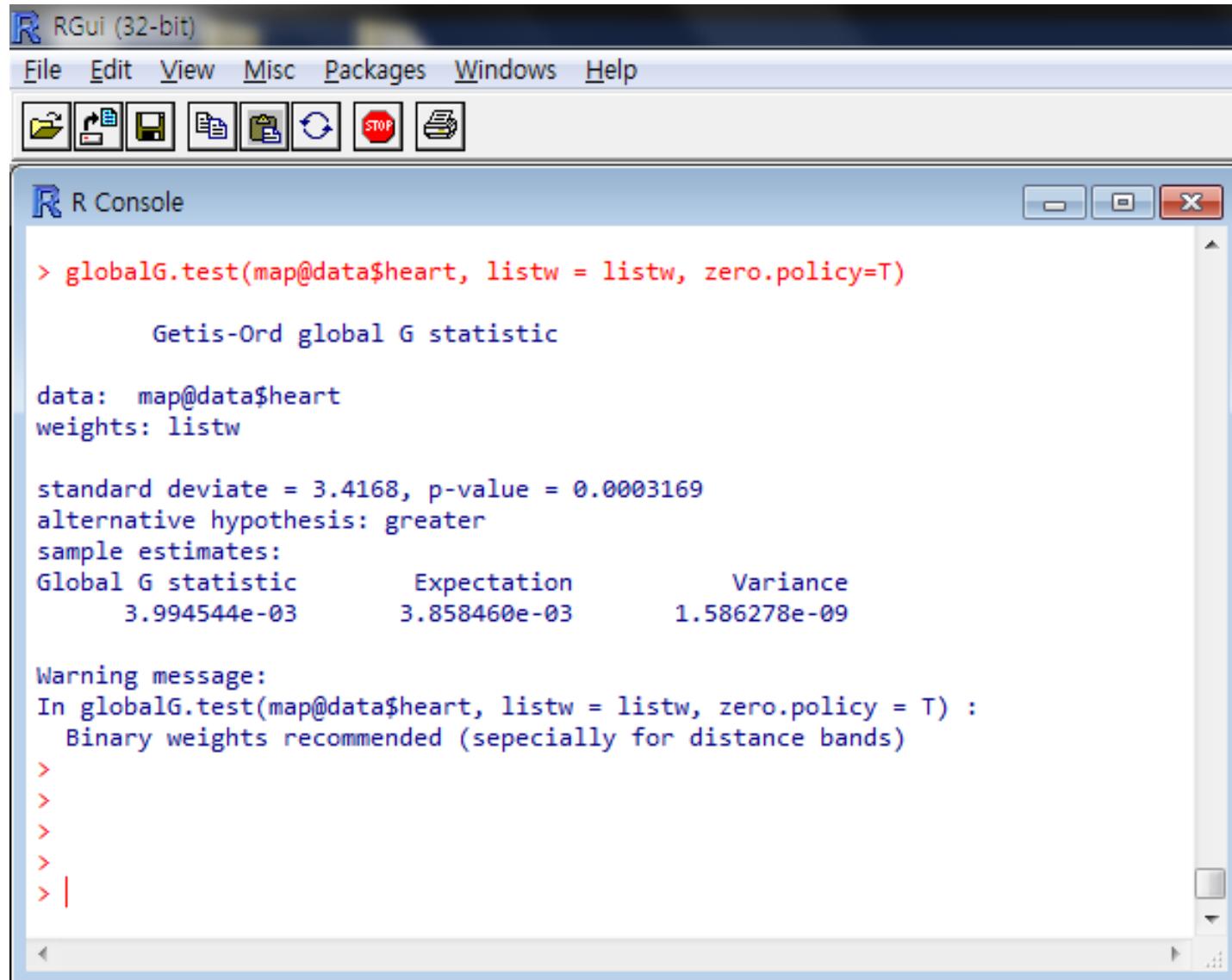
```
# Mantel's permutation test
```

```
sp.mantel.mc(map$heart, listw = listw, nsim = 999,  
zero.policy=T)
```

```
# Spatial lag plot
```

```
moran.plot(map$heart, listw = listw, zero.policy=T)
```

# Getis-Ord G statistic



The screenshot shows the RGui (32-bit) application window. The title bar reads "RGui (32-bit)". Below it is a menu bar with "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". A toolbar with various icons follows. The main area is the "R Console" window, which displays the following R session:

```
> globalG.test(map@data$heart, listw = listw, zero.policy=T)

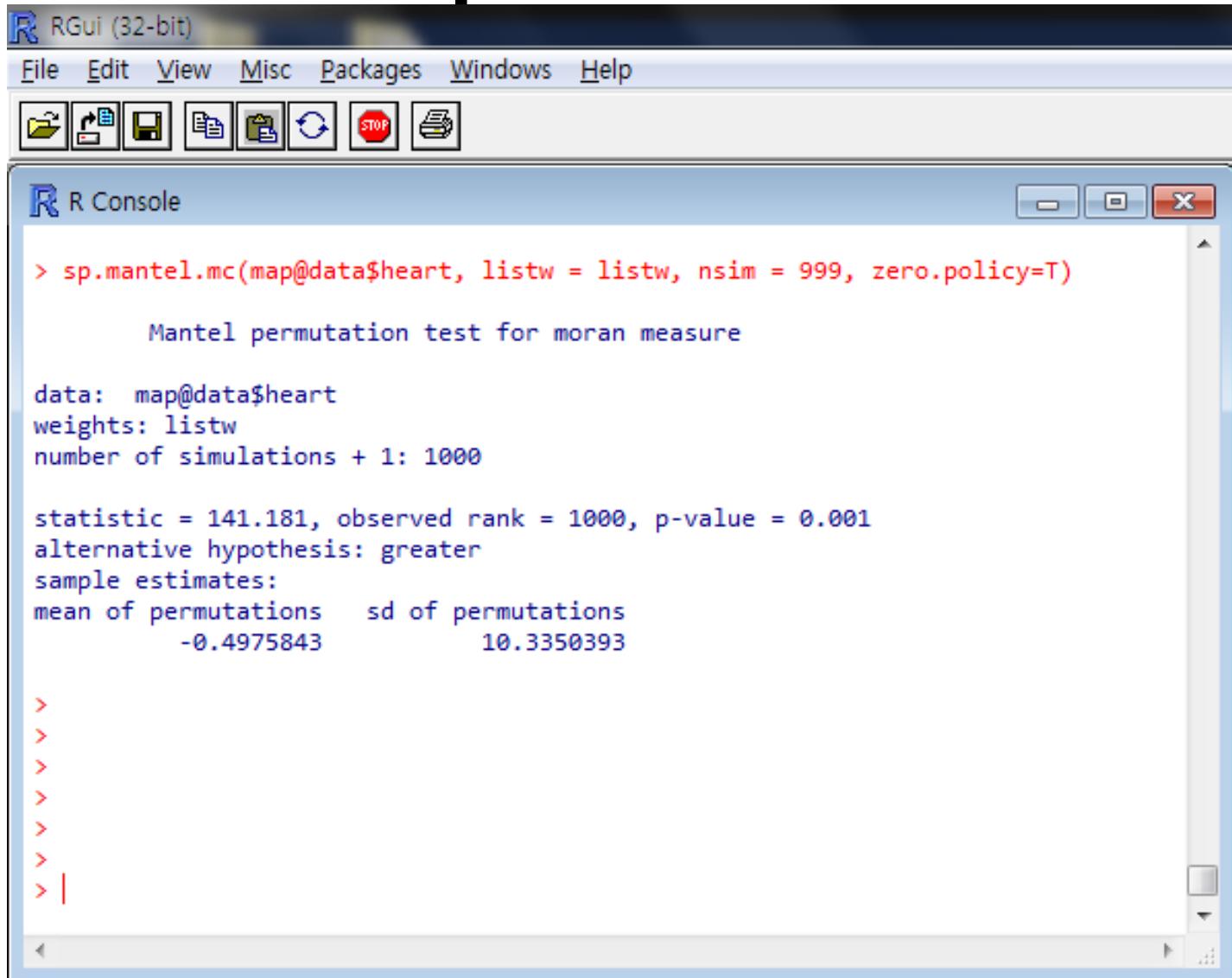
    Getis-Ord global G statistic

data: map@data$heart
weights: listw

standard deviate = 3.4168, p-value = 0.0003169
alternative hypothesis: greater
sample estimates:
Global G statistic      Expectation      Variance
            3.994544e-03     3.858460e-03     1.586278e-09

Warning message:
In globalG.test(map@data$heart, listw = listw, zero.policy = T) :
  Binary weights recommended (sepecially for distance bands)
>
>
>
>
> |
```

# Mantel's permutation test



The screenshot shows the RGui (32-bit) interface with the R Console window open. The console displays the output of a Mantel permutation test. The command run is:

```
> sp.mantel.mc(map@data$heart, listw = listw, nsim = 999, zero.policy=T)
```

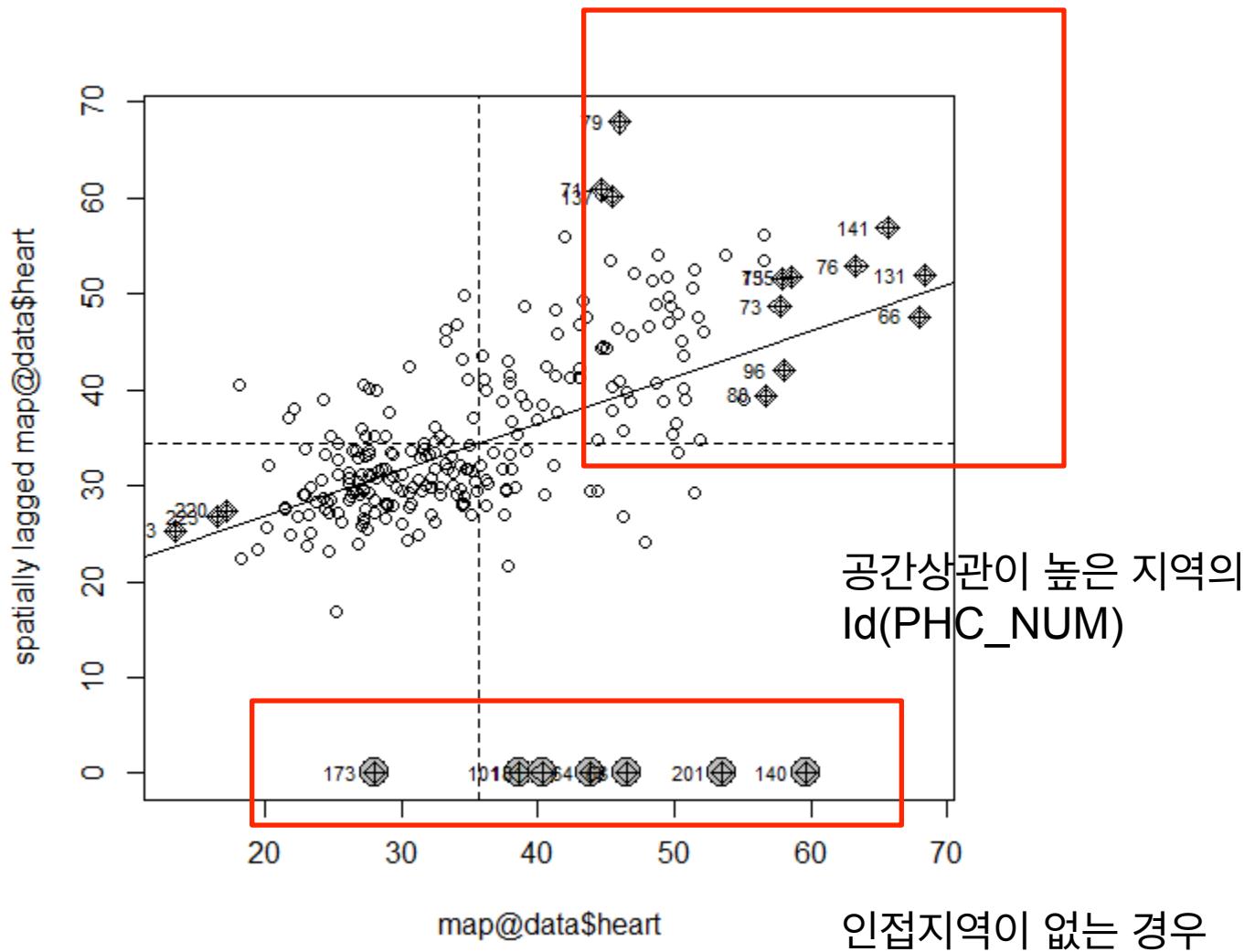
The output shows:

```
Mantel permutation test for moran measure

data: map@data$heart
weights: listw
number of simulations + 1: 1000

statistic = 141.181, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater
sample estimates:
mean of permutations    sd of permutations
-0.4975843              10.3350393

>
>
>
>
>
>
> |
```



# 4. R을 이용한 공간 회귀분석

- Spatial Regression Methods
- GAM, GWR,
- 데이터에 공간상관이 존재하는 걸로 판명되면 공간 회귀분석, 그렇지 않으면 보통의 회귀분석 기법을 사용해도 무방함.

# 회귀모형 잔차의 공간상관 검정

## # 공간상관을 고려하지 않은 단순회귀분석

```
#library(spdep)
lm0 <- lm(heart~cr, data=map@data)
summary(lm0)
```

## #앞선 lm0 분석의 잔차에 대한 Moran's test

```
lm.morantest(lm0, listw, zero.policy=T)
#or
moran.test(lm0$residuals, listw, zero.policy=T)
```

R GUI (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> lm0 <- lm(heart~cr, data=map@data)
> summary(lm0)

Call:
lm(formula = heart ~ cr, data = map@data)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.873 -8.021 -1.975  6.872 30.899 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 21.7089    5.4485   3.984 8.87e-05 ***
cr           0.5912    0.2282   2.590   0.0102 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.34 on 251 degrees of freedom
Multiple R-squared:  0.02604, Adjusted R-squared:  0.02215 
F-statistic:  6.71 on 1 and 251 DF,  p-value: 0.01015

> moran.test(lm0$residuals, listw, zero.policy=T)

        Moran's I test under randomisation
data: lm0$residuals
weights: listw
Moran I statistic standard deviate = 13.8397, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
      0.568471288     -0.004081633     0.001711508

> |
```

회귀분석 후 잔차에 공간상관이 있어  
공간회귀분석 필요

# CAR 공간회귀분석

## family ="CAR" 부분은 SAR로 전환 가능

```
lm1 <- spautolm(heart~cr, data=map, family="CAR", listw=listw)
```

```
summary(lm1)
```

```
moran.test(lm1$fit$residuals, listw, zero.policy=T)
```

R GUI (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
Call: spautolm(formula = heart ~ cr, data = map@data, listw = listw,
family = "CAR")

Residuals:
    Min      1Q  Median      3Q     Max 
-23.720384 -4.008993 -0.052756  4.003544 21.745759

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 29.34984   4.65828  6.3006 2.966e-10  
cr           0.60228   0.17351  3.4711 0.0005184 

Lambda: 0.99488 LR test value: 167.29 p-value: < 2.22e-16
Numerical Hessian standard error of lambda: 0.0056002

Log likelihood: -865.2344
ML residual variance (sigma squared): 44.074, (sigma: 6.6388)
Number of observations: 253
Number of parameters estimated: 4
AIC: 1738.5

>
> moran.test(lm1$fit$residuals, listw, zero.policy=T)

    Moran's I test under randomisation

data: lm1$fit$residuals
weights: listw

Moran I statistic standard deviate = -7.0535, p-value = 1
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
-0.295421519     -0.004081633     0.001706027
```

# gwr은 시간이 걸리고 약간 불안정

```
# library(spgwr)
```

```
col.bw <- gwr.sel(heart ~ cr, data=map, coords=cbind(map@data$long, map@data$lat), gweight=gwr.bisquare)
```

```
col.gauss <- gwr(heart ~ cr, data=map, coords=coordinates(map), bandwidth=col.bw, hatmatrix=TRUE)
```

R Gui (32-bit)

File Edit View Misc Packages Windows Help

R Console

```
> col.gauss
Call:
gwr(formula = heart ~ cr, data = map@data, coords = coordinates(map),
    bandwidth = col.bw, hatmatrix = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 1.752484
Summary of GWR coefficient estimates at data points:
      Min. 1st Qu. Median 3rd Qu. Max. Global
X.Intercept. 15.0400 17.1100 20.6400 22.0300 32.2000 21.7089
cr           0.3048  0.6266  0.6643  0.6916  0.7887  0.5912
Number of data points: 253
Effective number of parameters (residual: 2traceS - traceS'S): 4.53346
Effective degrees of freedom (residual: 2traceS - traceS'S): 248.4665
Sigma (residual: 2traceS - traceS'S): 8.825111
Effective number of parameters (model: traceS): 3.484467
Effective degrees of freedom (model: traceS): 249.5155
Sigma (model: traceS): 8.806541
Sigma (ML): 8.745686
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 1824.442
AIC (GWR p. 96, eq. 4.22): 1818.759
Residual sum of squares: 19351.22
Quasi-global R2: 0.2969985
>
> moran.test(col.gauss$lm$residuals, listw, zero.policy=T)

      Moran's I test under randomisation

data: col.gauss$lm$residuals
weights: listw

Moran I statistic standard deviate = 13.8397, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
          0.568471288     -0.004081633     0.001711508
```

# GAM을 이용한 공간회귀

- # Generalized Additive Model을 이용한 분석
  - X,Y 좌표를 smoothing term으로 하여 공간상관성 보정

```
# library(mgcv)
gmod <- gam(heart~ cr + s(long, lat), data=map@data)
summary(gmod)
```

R Gui (32-bit)

File Edit View Misc Packages Windows Help

STOP

R Console

```
Family: gaussian
Link function: identity

Formula:
heart ~ cr + s(long, lat)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 21.2572    3.9363   5.400 1.64e-07 ***
cr          0.6103    0.1651   3.696 0.000273 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
        edf Ref.df      F p-value
s(long,lat) 17.94  22.79 15.58 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.593  Deviance explained = 62.4%
GCV = 48.203  Scale est. = 44.404    n = 253
>
> moran.test(gmod$residuals, listw, zero.policy=T)

        Moran's I test under randomisation

data: gmod$residuals
weights: listw

Moran I statistic standard deviate = -1.3187, p-value = 0.9064
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
-0.058574969     -0.004081633     0.001707730
```

# R-INLA를 이용한 공간 회귀분석

- <http://www.r-inla.org/download>
- Bayesian CAR(BYM)모형을 추정하기 위한 방법으로 원래는 WinBUGS를 이용하여 MCMC추정.
- Bayesian 회귀분석을 위해 기존 MCMC방법보다 빠른 INLA 방법을 사용.

```
> source("http://www.math.ntnu.no/inla/givemeINLA.R")  
> inla.upgrade()
```

# R-INLA를 이용한 CAR모형

- 심장질환 발생률과 흡연률 간 관계분석
- 데이터 및 인접지역간 연결 관계를 보여주는 weight matrix (.graph) 필요.

```
temp <- poly2nb(map)
nb2INLA("map.graph", temp)

g="C:\\\\Users\\\\cyberdoc\\\\desktop\\\\2014대한예방의학회\\\\지역사회조사\\\\map.graph"

model = heart ~ cr + f(PHC_NUM, model="besag", graph=g)

result = inla(model, family="gaussian",
              data=map@data, control.predictor=list(compute=TRUE),
              control.compute=list(dic=TRUE, cpo=TRUE))

summary(result)
```



## R Console

```
>
> summary(result)

Call:
c("inla(formula = model, family = \"gaussian\")", data = map@data, control.compute = list(dic = TRUE, ",    cpo = TRUE$"

Time used:
Pre-processing      Running inla Post-processing          Total
2.4201              4.1582            0.3430           6.9214

Fixed effects:
  mean     sd 0.025quant 0.5quant 0.975quant   mode kld
(Intercept) 25.6949 4.9536    15.9394  25.7016   35.4029 25.7155  0
cr          0.4226 0.2078    0.0150   0.4223    0.8313  0.4218  0

Random effects:
Name      Model
PHC_NUM  Besags ICAR model

Model hyperparameters:
                                         mean     sd 0.025quant 0.5quant 0.975quant mode
Precision for the Gaussian observations 0.0280 0.0071  0.0167      0.0271  0.0444  0.0253
Precision for PHC_NUM                  0.0057 0.0014  0.0034      0.0055  0.0089  0.0053

Expected number of effective parameters(std dev): 135.08(24.59)
Number of equivalent replicates : 1.873

Deviance Information Criterion: 1770.49
Effective number of parameters: 133.99

Marginal Likelihood: -1078.99
CPO and PIT are computed

Posterior marginals for linear predictor and fitted values computed
```

&gt;

7/5/15

# 앞의 회귀 추정 결과를 그림으로

```
mapping=map@data

mapping$car  <- lm1$fit$fitted.values
mapping$gwr  <- col.gauss$lm$fitted.values
mapping$gam  <- gmod$fitted.values
mapping$INLA <- result$summary.fitted.values[,4] #중앙값, 만약 평균값 추정지로 하고자 하면 1

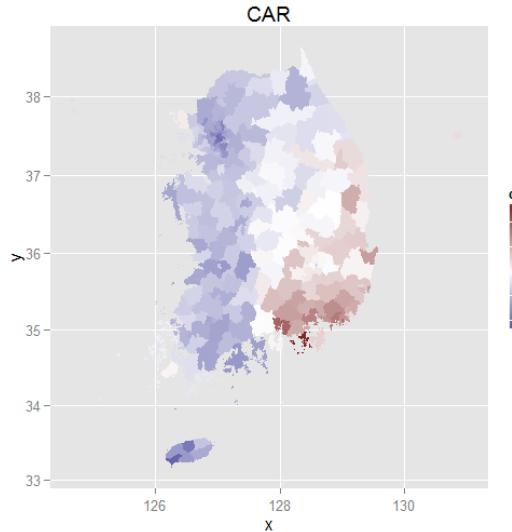
map_ft <- fortify(map, region="PHC_NUM")

# car 추정값의 최소값=22.68352, 최대값=52.45239, midpoint는 37로 설정
ggplot() + geom_map(data = mapping, aes(map_id = PHC_NUM, fill = car),
  map = map_ft) + expand_limits(x=map_ft$long, y=map_ft$lat) + scale_fill_gradient2(low = muted("blue"),
  mid = "white", midpoint = 37, high = muted("red"), limits = c(22, 53)) + coord_map(projection="mercator")

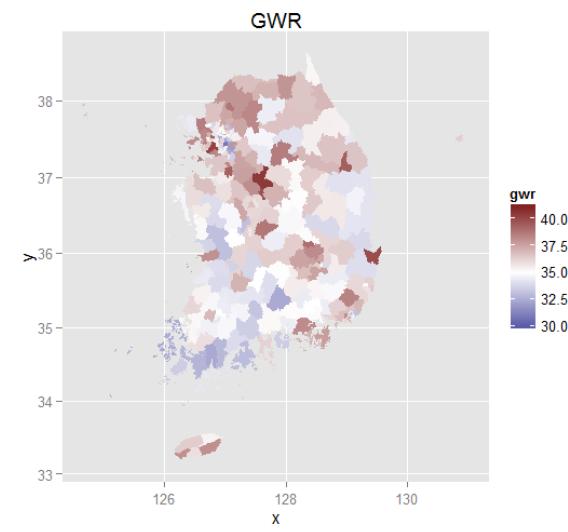
# gwr 추정값의 최소값=30.8627, 최대값=40.42974, midpoint는 35로 설정
ggplot() + geom_map(data = mapping, aes(map_id = PHC_NUM, fill = gwr),
  map = map_ft) + expand_limits(x=map_ft$long, y=map_ft$lat) + scale_fill_gradient2(low = muted("blue"),
  mid = "white", midpoint = 35, high = muted("red"), limits = c(30, 41)) + coord_map(projection="mercator")

# gam 추정값의 최소값=23.1289, 최대값=55.96624, midpoint는 40으로 설정
ggplot() + geom_map(data = mapping, aes(map_id = PHC_NUM, fill = gam),
  map = map_ft) + expand_limits(x=map_ft$long, y=map_ft$lat) + scale_fill_gradient2(low = muted("blue"),
  mid = "white", midpoint = 40, high = muted("red"), limits = c(23, 56)) + coord_map(projection="mercator")

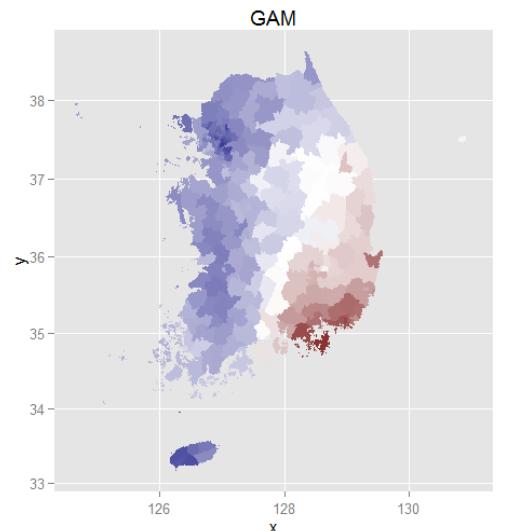
# INLA 추정값의 최소값=16.1302, 최대값=60.72727, midpoint는 39로 설정
ggplot() + geom_map(data = mapping, aes(map_id = PHC_NUM, fill = INLA),
  map = map_ft) + expand_limits(x=map_ft$long, y=map_ft$lat) + scale_fill_gradient2(low = muted("blue"),
  mid = "white", midpoint = 39, high = muted("red"), limits = c(16, 61)) + coord_map(projection="mercator")
```



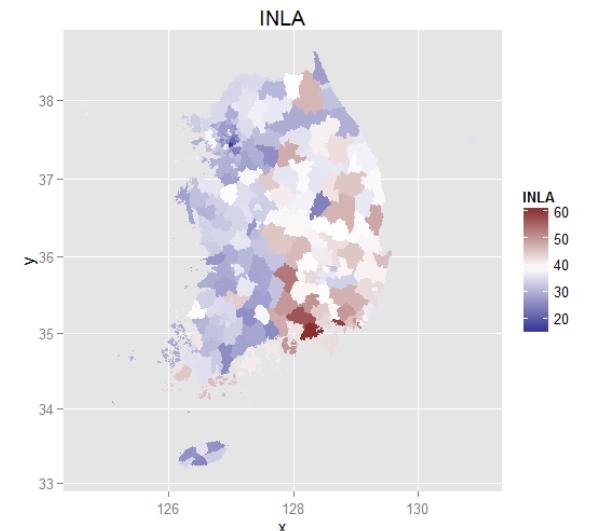
CAR model using spautolm()



GWR model using gwr()



GAM model using gam()



CAR-Besag model using INLA

# 참고 문헌

- Bivand RS, Pebesma EJ, Gomez-Rubio V. (2014) Applied Spatial Data Analysis with R, 2<sup>nd</sup> ed. Springer.
- Blangiardo M, Cameletti M. (2015) Spatial and Spatio-temporal Bayesian Models with R-INLA, Wiley.
- Lamigueiro OP. (2014) Displaying Time Series, Spatial, and Space-Time Data with R, CRC Press.
- Pfeiffer D, Robinson T, Stevenson M, et al. (2008) Spatial Analysis in Epidemiology, Oxford University Press.

