이학석사 학위논문

The Statistical Computation of Heat Disorder Risk with HGLM

HGLM을 활용한 온열질환 리스크의 통계적 계산

2017년 2월

서울대학교 대학원

협동과정 계산과학

이 진 희

The Statistical Computation of Heat Disorder Risk with HGLM

지도교수 이 영 조

이 논문을 이학석사 학위논문으로 제출함

2016년 12월

서울대학교 대학원

협동과정 계산과학

이 진 희

이진희의 이학석사 학위논문을 인준함

2017년 2월

위 원 장 이 재 용 (인)

부위원장 이 영 조 (인)

위 원 신 동 우 (인)

The Statistical Computation of Heat Disorder Risk with HGLM

by

Jinhee Lee

A Dissertation
submitted in fulfillment of the requirement
for the degree of
Master of Science
in
Interdisciplinary Program of
Computational Science and Technology

Interdisciplinary Program of
Computational Science and Technology
College of Natural Sciences
Seoul National University
February, 2017

Abstract

Jinhee Lee
Interdisciplinary Program of Computational Science and Technology
College of Natural Sciences
Seoul National University

This study provides the prediction result of heat disorder incidence risk using Hierarchical Generalized Linear Model (HGLM) on the basis of the relationship between climate variables (temperature and relative humidity) and heat disorder incidence. Basic descriptive statistics were calculated to track down to any change in climate variables over the past 43 years. The empirical (probability) density functions were simulated by four different times of 1970s, 1980s, 1990s and the recentest. Furthermore, we compared the statistics, regional ranges and regional standard deviations, of weather variables in 1973 and in 2015(t-test was applied).

Understanding the variables with these statistics, two types of response variable (the monthly sum of heat disorder and the monthly sum of heat disorder per 1 million people) were modeled to predict the risk with explanatary climate variables. Especially, a spatial correation structure was included in the models as a random effect. This spatial correlation sturcture had the location information of each region in terms of adjacency. We found that this could decrease the significance of nominal region variable. With the estimates obtained from the chosen model, we compared the simulated heat disorder incidence risk during the unobserved period to the observed current data.

Keyword :*climate, heat disorder incidence, hierarchical generalized linear model, random effect, spatial correlation*

Student Number : 2013-23011

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Climate change has been discussed as a top priority issue since Rio Earth Summit in 1992. Conference of Parties ("COP") of UN Framework on Climate Change ("UNFCCC") has taken place to review the current status of climate and implement any collective and political actions to prevent climate change and adapt to the changing environmental conditions. COP21 in December 2015 is referred as the recentest success of these kinds of international cooperations.

Meanwhile, Intergovernmental Panel on Climate Change ("IPCC") has contributed to the actions against climate change in the different position. IPCC has provided various kinds of scientific information on climate change. In 2012, it published a special report on the risks of extreme events, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*("SREX"). This special report provides us an idea of how extreme weather events could occur using statistical concepts. Figure 1.1 (IPCC, 2012) shows us three statistical causations of extreme events. Three causes

are depicted under the assumption that weather related variables follow the Normal Probability Distribution.

According to Figure 1.1, firstly, extreme events can be provoked by the shift of mean of temperature distribution. Secondly, they can be caused by the increase of variability of temperature distribution. And the third, the change of symmetry of temperature distribution can result in the extreme events.

This growing concern for climate change has motivated and forced the goverments to take any action to slow down the change and adapt to the changing environment. The South Korean Government also has moved against this environmental trend. For example, Korea Meteorological Administration ("KMA") has provided scientific information about climate change including relevant data and meteorological and climatological materials to the public.

Another example is the reporting system managed by Centers for Disease Control & Prevention of South Korea ("KCDC"). KCDC has operated this special reporting system to manage weather related disease like heat disorder and the disease caused by cold surge since 2011. KCDC counts a daily occurrence of patients who were transported to an emergency room of hospital due to these diseases nationally during a specific period. And, this data is opened to the public through its website.

Inspired by Figure 1.1, our study starts at this point. Our study aims at solving two questions. Firstly, we want to figure out if we can find any visible proof of climate change in South Korea. And secondly, we also want to find out how we can calculate the risk of climate as an effect to health.

Briefly, the occurrence of patient can be a measure to capture the weather

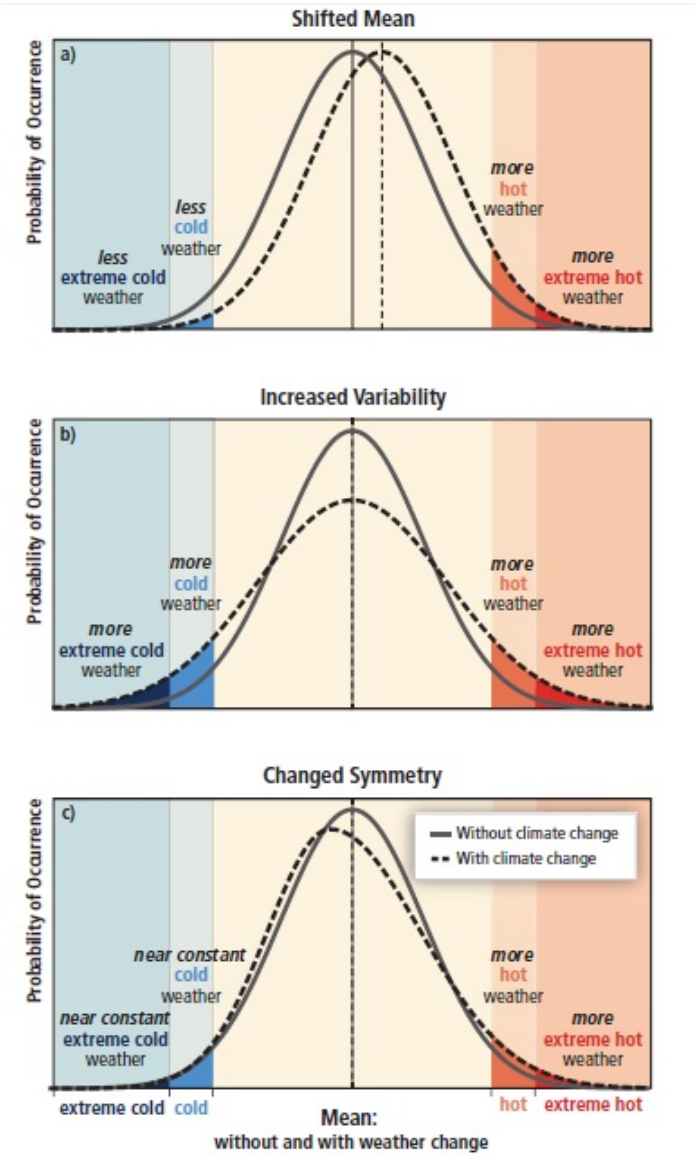


Figure 1.1: The effects of changes in temperature distribution on extremes. (IPCC, 2012)

or climate relevant risk especially for summer season. In this study, we try to explore one way to measure the risk by applying Generalized Linear Model ("GLM"), Hierarchical Generalized Linear Model ("HGLM") and the data from KMA and that from KCDC.

Chapter 2 covers the description of data used in our study. Chapter 3 contains the basic analysis which would lead to the modeling analysis. And, we generated the (probability) density functions of temperature and humidity to capture any change in them over the past 43 years. Chapter 4 consists of the results of GLM and HGLM modelling with data and their implications. This chapter also provides the reconstructed (simulated) heat disorder incidence during 10 years which starts 1973 and ends 1982. Finally, Chapter 5 is composed of the findings we have obtained by this study. Additionally, Appendix conveys the additional information and graphs that supported our study.

Chapter 2

Data Description

For this study, we used the data from two sources described at the end of Chapter 1, KMA and KCDC. We collected three temperature variables and one relative humidity variable from KMA website. We got one heat disorder incidence variable from KCDC website. We give the details of our dataset below.

2.1. Details on KMA Data

Considering the first question mentioned in Chapter 1, we tried to get data from as many as possible stations which have been measured for a long time. Thinking these two conditions, the number of stations and the available data measuring period, we chose the data from 60 KMA stations accumulated for 43 years. The stations are depicted in Figure 2.1.

There are various kinds of available data on the KMA website (http://data.kma.go.kr). We selected daily average temperature, daily highest temperature, daily low-

est temperature and daily average relative humidity which were expected to show a proof of climate change and be relevant with heat disorder.

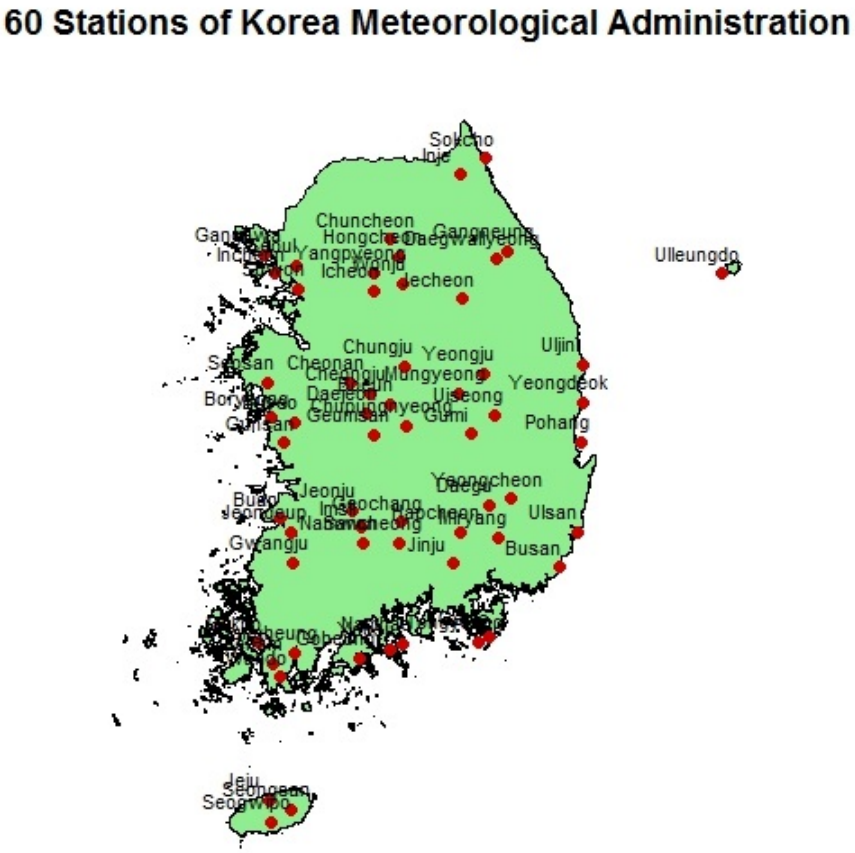


Figure 2.1: 60 KMA Stations Chosen Dispersed Over South Korea. The details of each station are available in Appendix.

These 4 variables (3 temperature variables, 1 relative humidity variable) are the measurements in 60 stations all over South Korea between 1st June and 31st August, 3 months (92 days) corresponding to the specific period that KCDC operates its special reporting system, for 43 years from 1973 to 2015.

2.2. Details on KCDC Data

KCDC started its special reporting system to manage the occurrence of patients transported to emergency room in hospital due to *the hot weather* who were diagnosed as *heat disorder*. This data is also available on the KCDC website (www.cdc.go.kr). We made a special request to receive this data to KCDC. It provided daily data between 1st June and 31st August (92 days) for 3 years ($2012 \sim 2014$) excluding the measurement during the test period in 2011. And, we updated this data with the newly measurements between 1st June and 31st August in 2015 using the daily reports released on the its website. So, we could use 4-year data to study.

Looking through the report published by KCDC (KCDC, 2016), we can see more dimensions of this data such as sex, age, etc.. But, these features have been shared only as a total count without any detailed time and space information.

The final data used for our analysis contains daily occurrence of patients due to hot weather of 16 provinces of South Korea accumulated for 4 years. Actually, KCDC collects data from 17 provinces of South Korea, but we excluded the data from Sejong city due to its lack of KMA data.

2.3. Preprocess for KMA and KCDC Data

In Section 2.1 and Section 2.2, we provided the description of raw data we collected from two different governmental bodies. There are some differences between the datasets, due to the difference in their original sources. So, we had to take a step to modify these differences before our analysis. Table 2.1 gives us a feature summary of our raw data.

	Raw Data from KMA	Raw Data from KCDC
Time Unit	"Day" (Daily Measurement)	
Period	1st June $\sim$ 31st August (92 days)	
	1973 $\sim$ 2015 (43 years)	2012 $\sim$ 2015 (4 years)
Region	60 Stations	16 Provinces
Variable	Average Temperature Highest Temperature Lowest Temperature Average Relative Humidity	Heat Disorder Incidence

Table 2.1: The Feature Summary of Raw Dataset (KMA & KCDC). KMA provides the exact loacation information of each station, while KCDC data is only gathered on the regional level.
"Province" has 16 values such as Busan, Chungbuk, Chungnam, Daegu, Daejeon, Gangwon, Gwangju, Gyeonggi, Gyeongbuk, Gyeongnam, Incheon, Jeju, Jeonbuk, Jeonnam, Seoul and Ulsan.

A couple of adjustments were implemented to match the spatial condition

of KMA data to that of KCDC data. At first, we re-classified 60 KMA stations into 16 provinces.

Province	Number of KMA Stations
Busan	1
Chungbuk	4
Chungnam	5
Daegu	1
Daejeon	1
Gangwon	7
Gwangju	1
Gyeonggi	3
Gyeongbuk	10
Gyeongnam	8
Incheon	2
Jeju	3
Jeonbuk	6
Jeonnam	6
Seoul	1
Ulsan	1

Table 2.2: The Number of KMA Stations Re-classified by Province. This table shows each Metropolitan City has only one station on the contrary to the multiple stations in each Province. So, we should keep in mind that the statistics of each Province might show bigger variabilities.

And, we calculated regional descriptive statistics as like minimum, mean, maximum and standard deviation of the temperature variables and the humidity variable. This calculation was proceeded twice to get daily statistics and its monthly version. These statistics from KMA data and CDC data were merged into one dataset to analyze.

2.4. Details on Population Data

It was necessary to include the population data into our study. Because we deal with a heat disorder (disease incidence) data. It may be easily guessed that more patients could occur where the population is bigger. So, we downloaded the population data from the website of KOrean Statistical Information Service (KOSIS). There are two kinds of available population data. One is the result of census and the other is the data from residence registration system. We used the latter because it is more updated one than the census.

Chapter 3

Basic Analysis

It is necessary to draw a big picture of our data to understand them easily. Therefore, we deliver the basic features of each data at first, and then move to the analysis on their relationships. Our two main questions are noted here again to remind the objectives of this study.

(1) *First, might there be any visible proof of climate change in terms of probability in South Korea?*

(2) *Second, how can we calculate the risk of climate as an effect to health?*

It may be possible to find a solution for the first question with the results of basic analysis. A proof of climate change might appear as a change in weather variables in the long term. We calculated temperature empirical density functions and a humidity empirical density function as Figure 1.1 which has motivated us to study. We used re-classified daily temperature variables (KMA), the re-classified daily relative humidity variable (KMA),

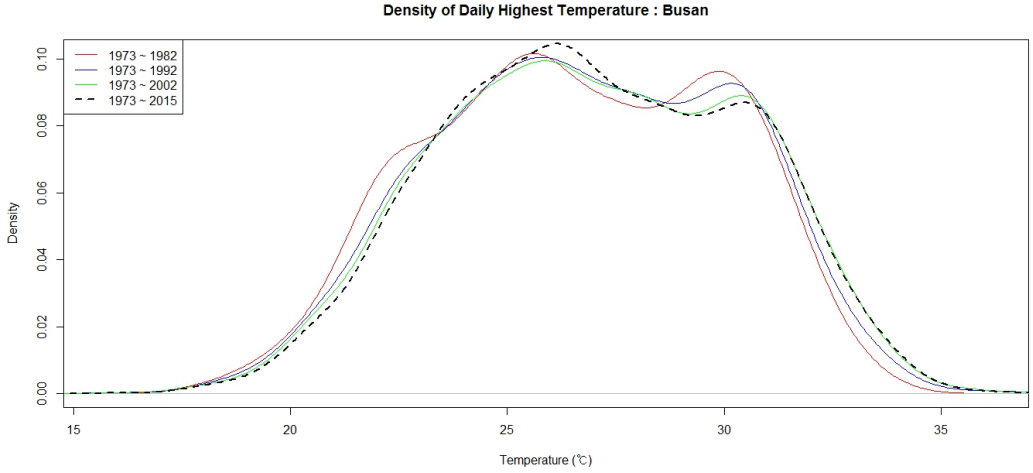
and the heat disorder incidence data (KCDC). We used density() function which is offered by the R (version 3.2.5) stats package.

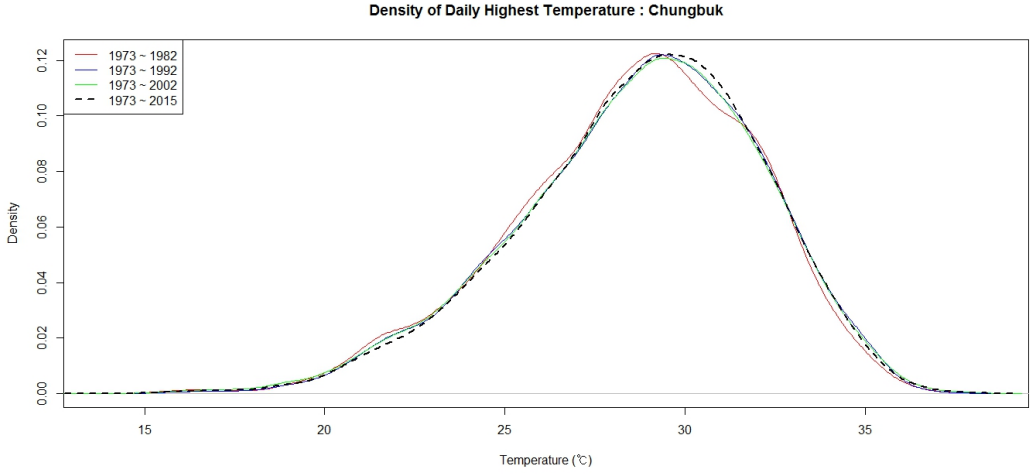
3.1. Density Functions of Daily Temperature

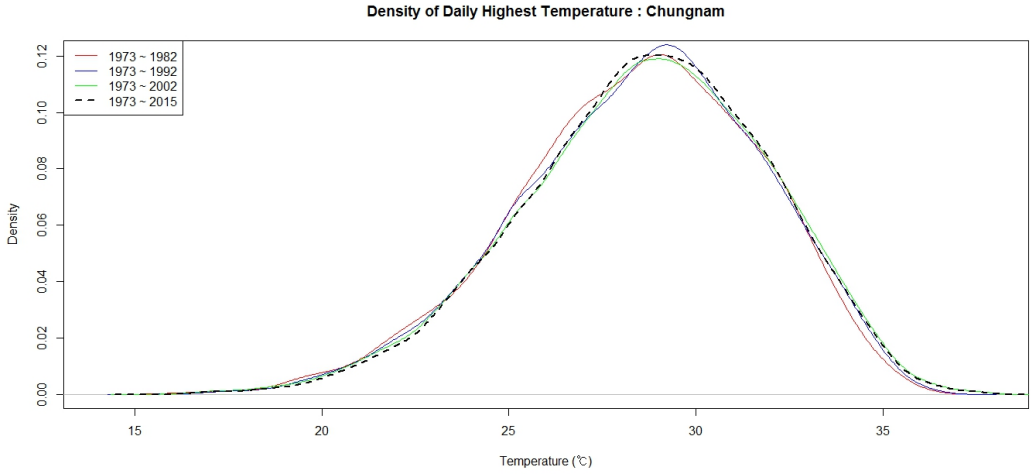
Here, we present re-classified daily highest temperature and daily average relative humidity. The density plots for daily average temperature variables are available in Appendix.

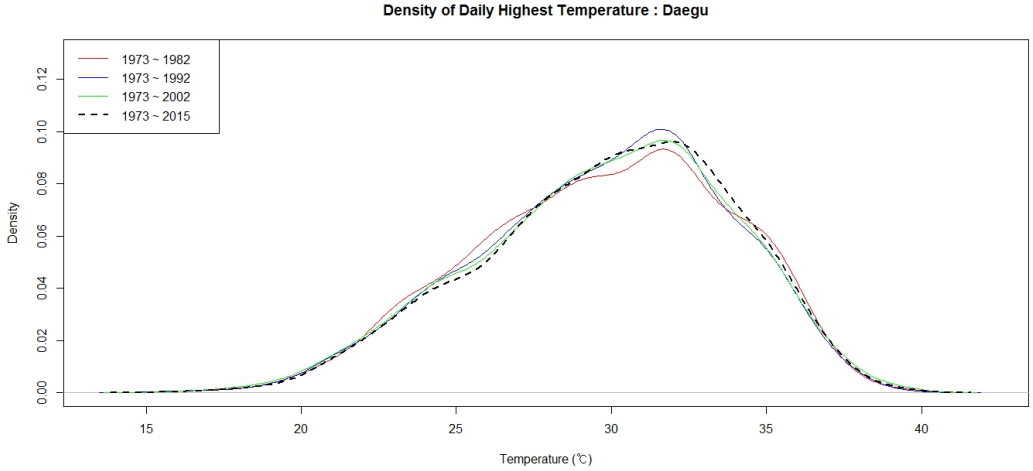
Our intent is to capture any change in empirical density function which can be considered as a proof of climate change as Figure 1.1. For this, we depicted 4 density functions by province to capture any change in density function in the long term by updating our variables adding newly measured ones.

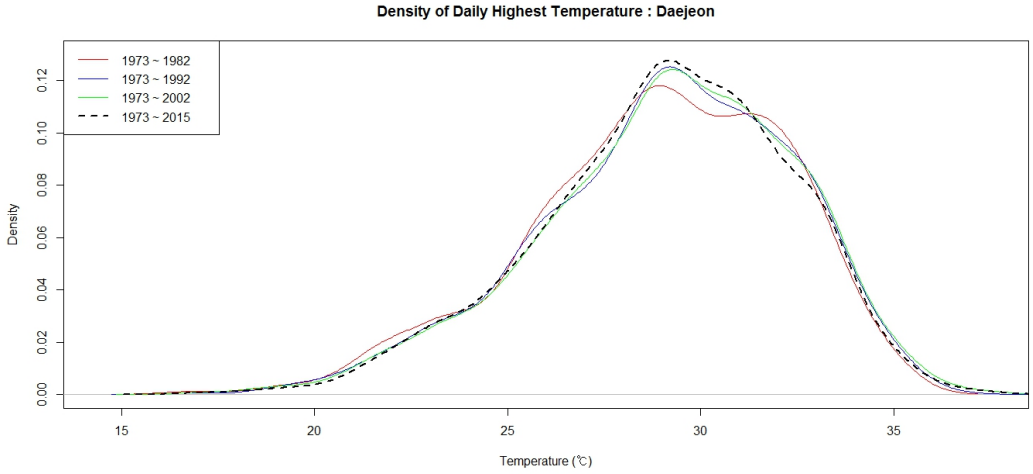
Thanks to density() function of R, we can review the details of density function derived. Figure 3.1. gives us the regional empirical density functions of daily highest temperature. If we compare the density lines in each province province, the most updated density function (Dashed line) seems to move a little bit to the right than the least updated one (Red line). And, the density functions of Busan, Jeju, Jeonnam, and Ulsan show twin-peak. We interpreted this twin-peak as a bigger temperature change between the hottest month (August) and the other months (June, July).

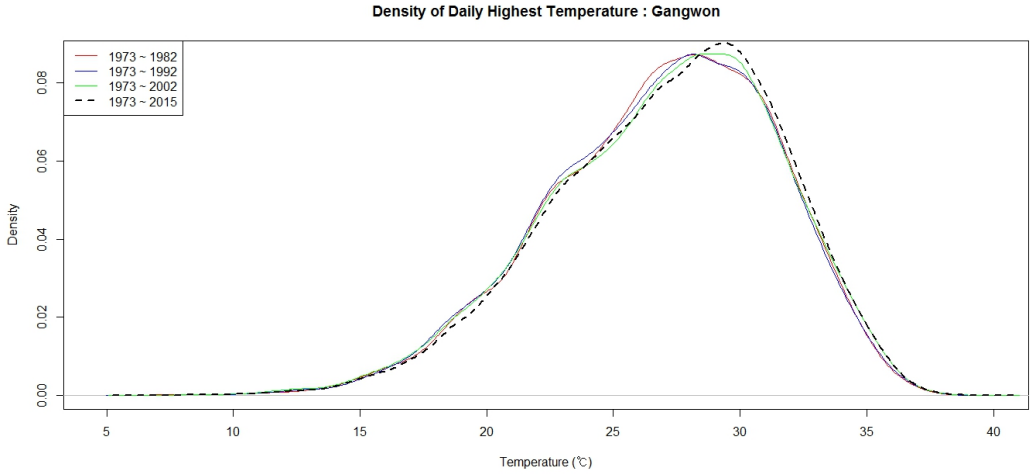

Density of Daily Highest Temperature : Busan
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
15
20
25
30
35
Temperature (℃)

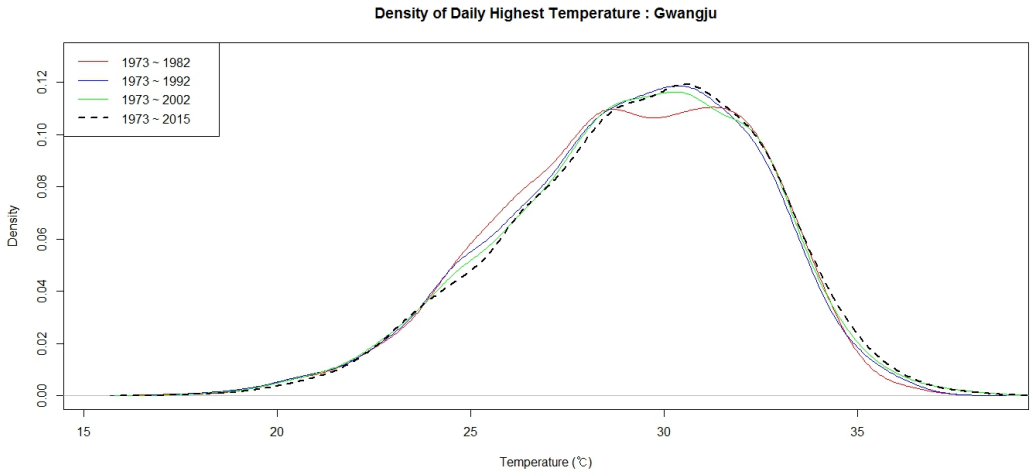

Density of Daily Highest Temperature : Chungbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)

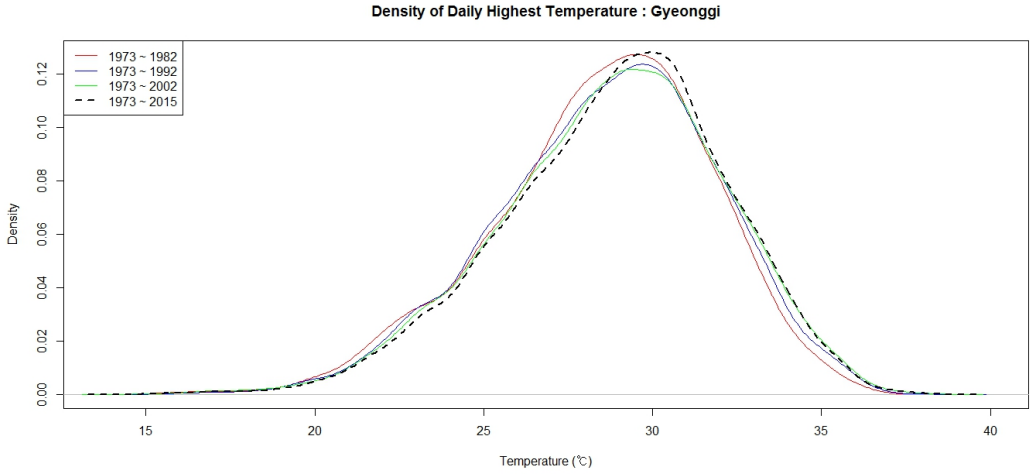

Density of Daily Highest Temperature : Chungnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)

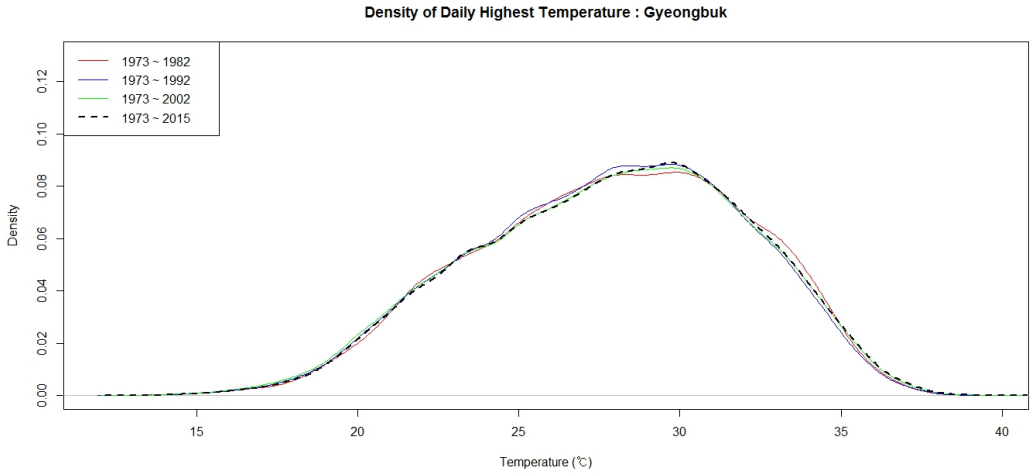

Density of Daily Highest Temperature : Daegu
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
40
Temperature (℃)

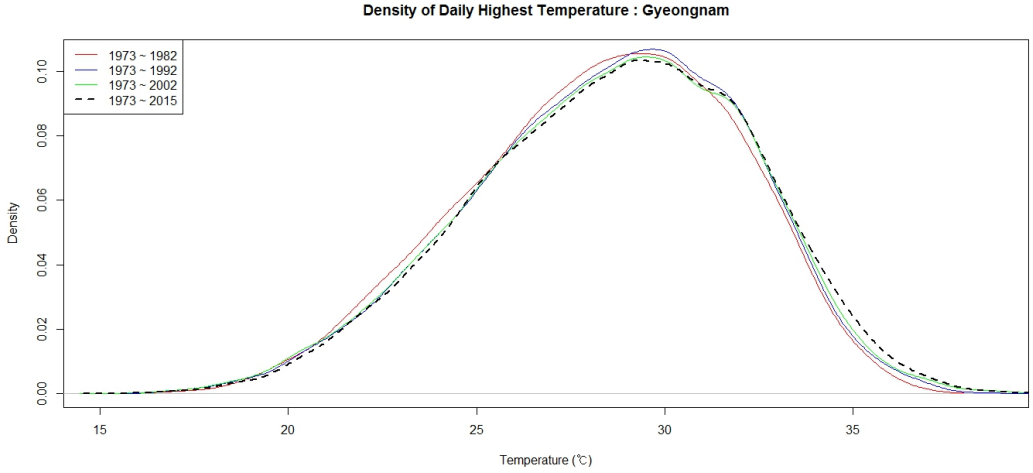

Density of Daily Highest Temperature : Daejeon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)

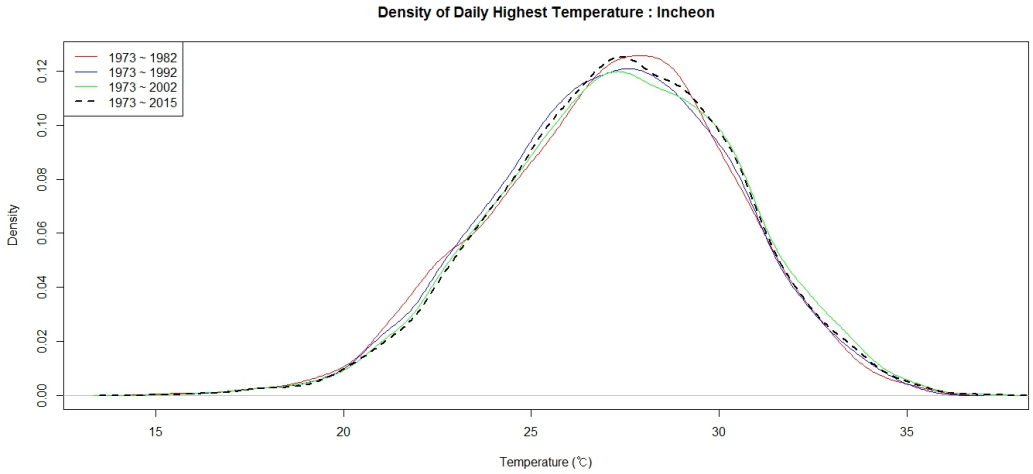

Density of Daily Highest Temperature : Gangwon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
5
10
15
20
25
30
35
40
Temperature (℃)

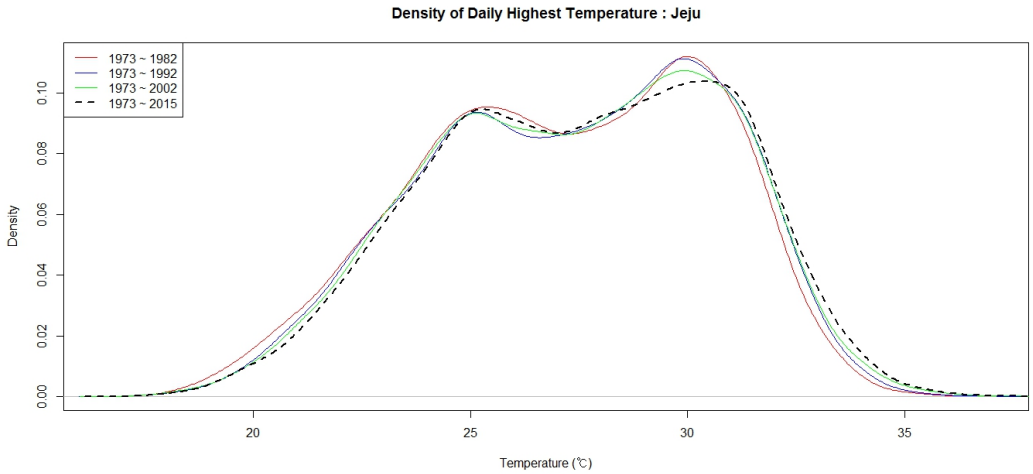

Density of Daily Highest Temperature : Gwangju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)

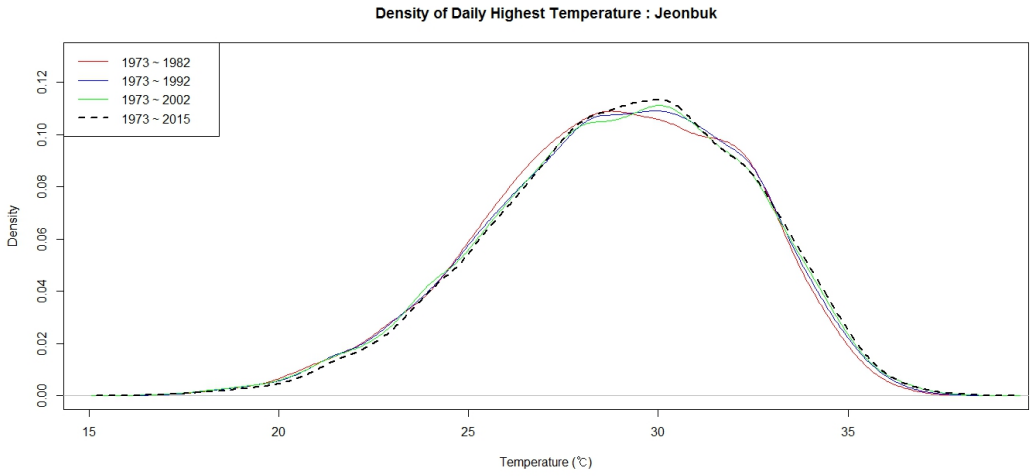

Density of Daily Highest Temperature : Gyeonggi
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
40
Temperature (℃)

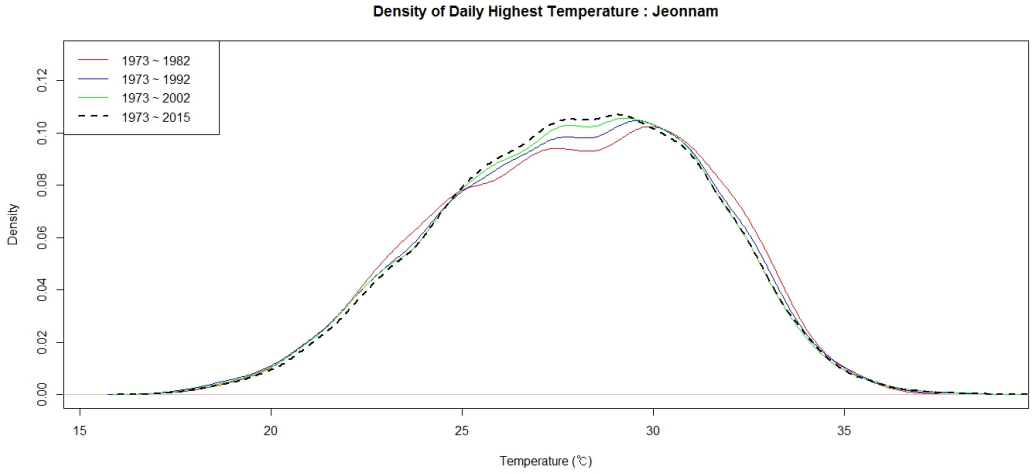

Density of Daily Highest Temperature : Gyeongbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
40
Temperature (℃)

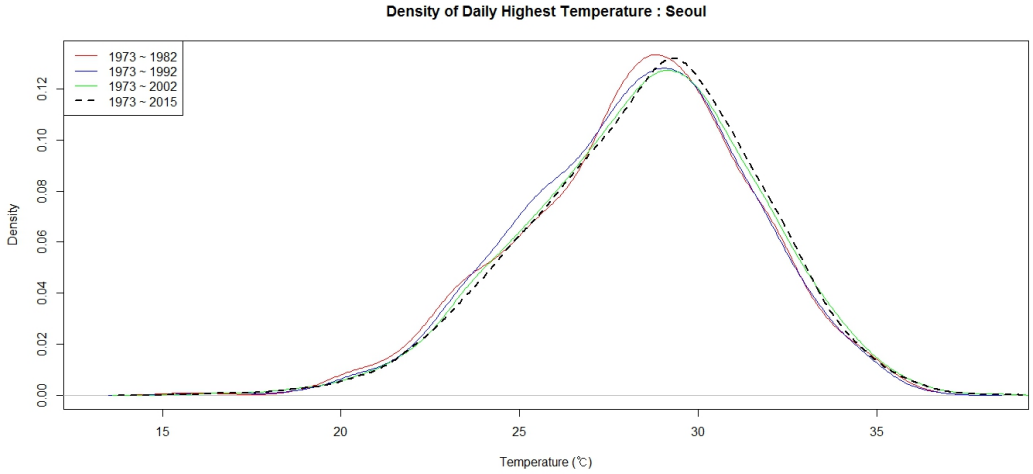

Density of Daily Highest Temperature : Gyeongnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
15
20
25
30
35
Temperature (℃)


Density of Daily Highest Temperature : Incheon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)


Density of Daily Highest Temperature : Jeju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
20
25
30
35
Temperature (℃)


Density of Daily Highest Temperature : Jeonbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)


Density of Daily Highest Temperature : Jeonnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)


Density of Daily Highest Temperature : Seoul
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
35
Temperature (℃)

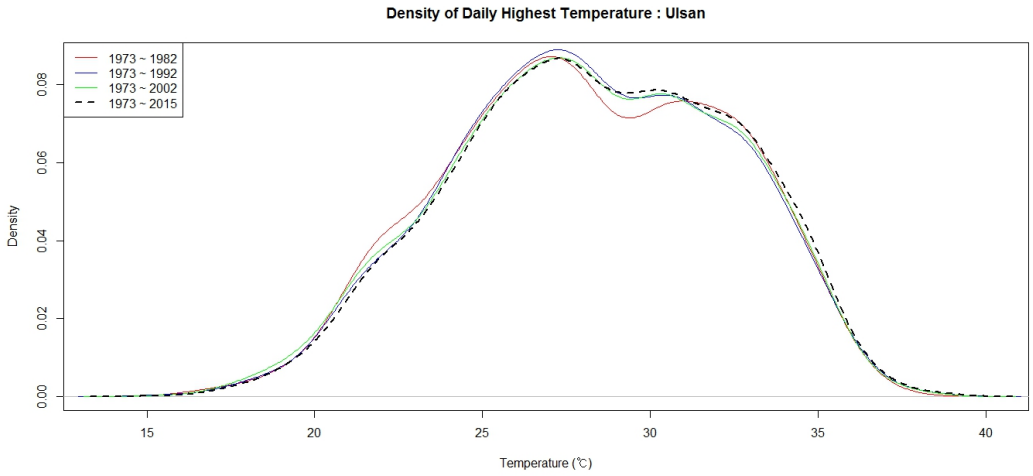


Figure 3.1: Density of Daily Highest Temperature by Province.

Table 3.1 shows the estimates from the generated density function. The slight inceases in estimated statistics support our observation and rough conclusion on Figure 3.1 by eyes. We can think the current empirical probability density function has changed to the hotter direction over the past 43 years.

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Busan	Min	15.60	Min	14.35
	Mean	25.90	Mean	26.40
	Max	36.20	Max	38.45
Chungbuk	Min	13.75	Min	12.09
	Mean	26.15	Mean	25.65
	Max	38.55	Max	39.21

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Chungnam	Min	14.06	Min	14.48
	Mean	26.00	Mean	26.75
	Max	37.94	Max	39.02
Daegu	Min	13.5	Min	13.83
	Mean	27.9	Mean	27.70
	Max	42.3	Max	41.57
Daejeon	Min	14.43	Min	15.05
	Mean	26.00	Mean	27.20
	Max	37.57	Max	39.35
Gangwon	Min	4.72	Min	5.23
	Mean	22.35	Mean	23.05
	Max	39.96	Max	40.87
Gwangju	Min	15.40	Min	15.92
	Mean	26.95	Mean	28.05
	Max	38.50	Max	40.18
Gyeonggi	Min	13.62	Min	13.27
	Mean	26.85	Mean	26.45
	Max	40.08	Max	39.63
Gyeongbuk	Min	11.98	Min	12.13
	Mean	25.85	Mean	26.45
	Max	39.72	Max	40.77
Gyeongnam	Min	14.99	Min	14.55
	Mean	26.85	Mean	27.60
	Max	38.71	Max	40.65

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Incheon	Min	13.51	Min	13.49
	Mean	25.40	Mean	26.05
	Max	37.29	Max	38.61
Jeju	Min	16.46	Min	16.11
	Mean	26.75	Mean	27.45
	Max	37.04	Max	38.79
Jeonbuk	Min	15.28	Min	15.19
	Mean	27.05	Mean	27.35
	Max	38.82	Max	39.51
Jeonnam	Min	15.51	Min	15.97
	Mean	27.20	Mean	27.95
	Max	38.89	Max	39.93
Seoul	Min	13.24	Min	13.75
	Mean	25.75	Mean	26.90
	Max	38.26	Max	40.05
Ulsan	Min	13.59	Min	13.30
	Mean	26.90	Mean	27.10
	Max	40.21	Max	40.90

Table 3.1: Minimum, Mean and Maximum Estimates of Density of Daily Highest Temperature (°C) by Province. density() function computes density estimates with kernel density method. So, this table shows the estimates not the real values from the uesd data.

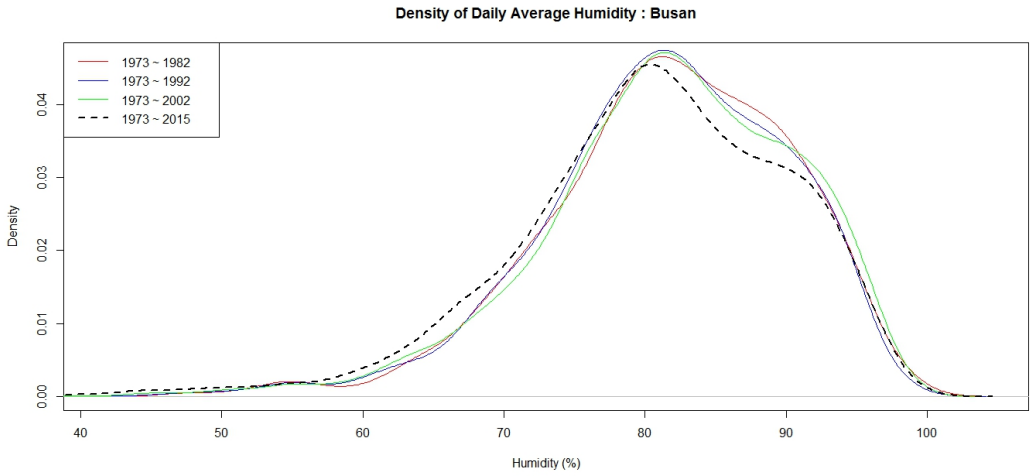
Generally, an abrupt change in temperature is cited as a proof of climate change. According to KMA report (KMA 2014), there is a steady increase in the annual average temperature of Korean Peninsula. (The increasing rate of temperature was 0.23 °C/10years from 1954 to 1999, 0.41 °C/10years from 1981 to 2010 and 0.5 °C/10years from 2001 2010). This fast increasing trend can be also explained by the changes of probability density of temperature.

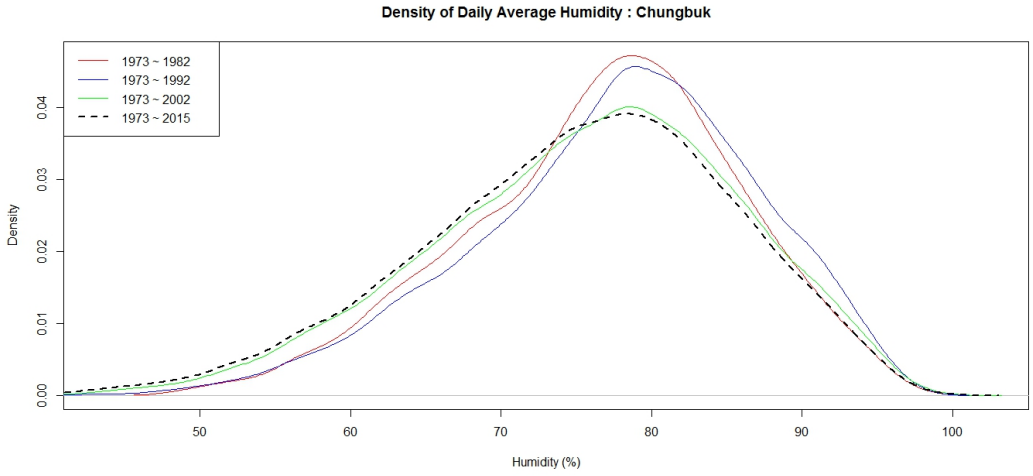
3.2. Density Functions of Daily Humidity

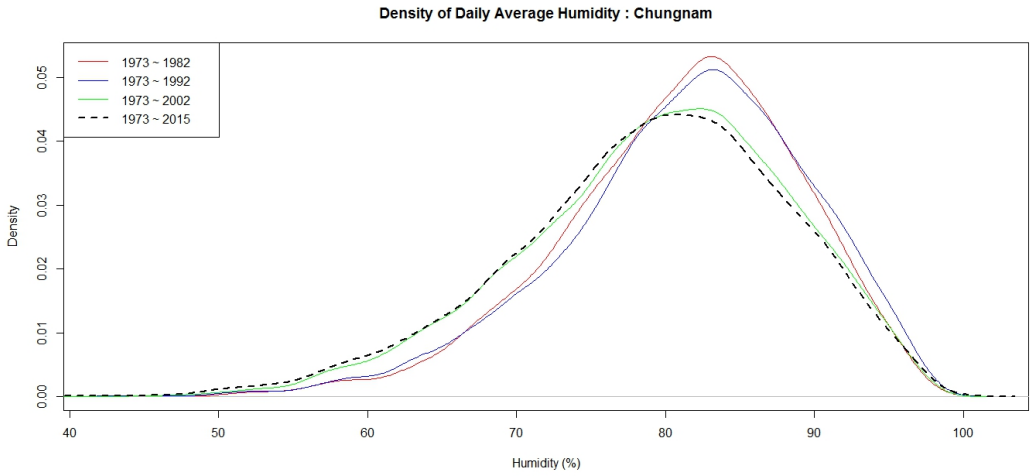
In this section, we discuss another variable from KMA, the empirical density of daily average relative humidity variable. The same method, density() function of R (version 3.2.5) stats package. was applied to this variable. Figure 3.2 below shows 4 kinds of density plots by province and Table 3.2 gives us the estimates as the results of density() function.

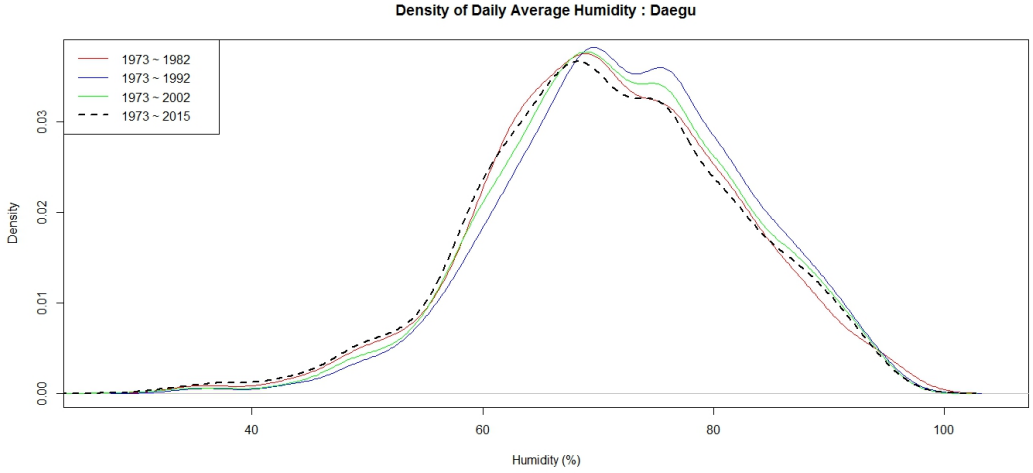
From them, we can observe a trend that the most updated density of daily average relative humidity is the flattest among them.

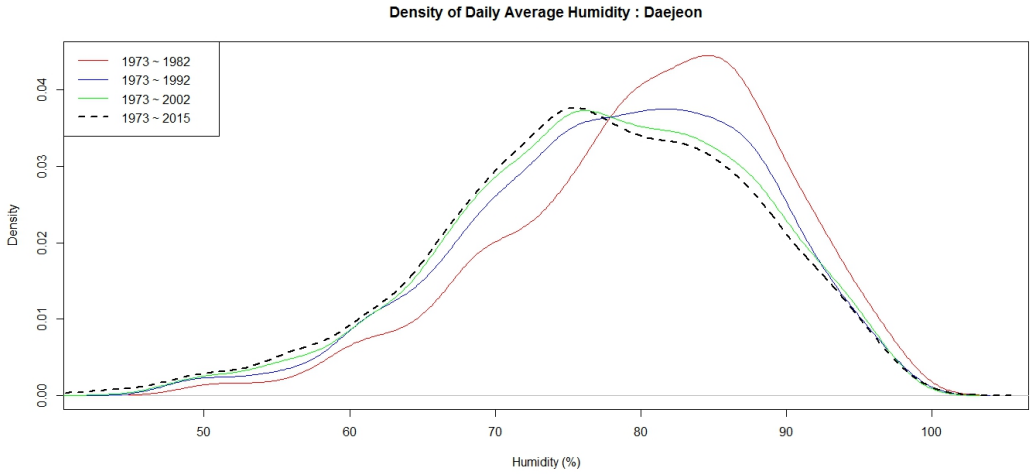
The densities becoming flatter seem to be as a decrease in the estimates from the least updated density to the most updated. And, it is also noted that due to the bigger decrease in the minimum estimates that those of maximum estimates, the ranges between maximums and minimums seem to increase.

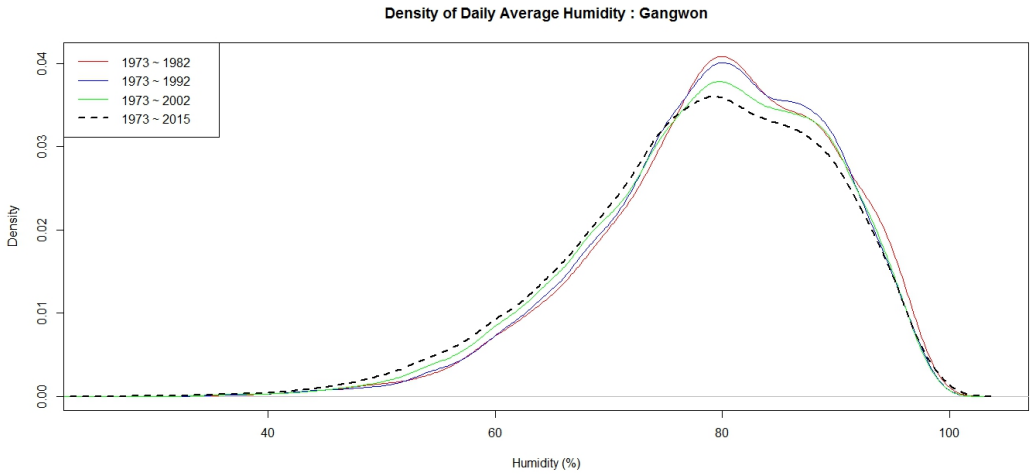

Density of Daily Average Humidity : Busan
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
40
50
60
70
80
90
100
Humidity (%)

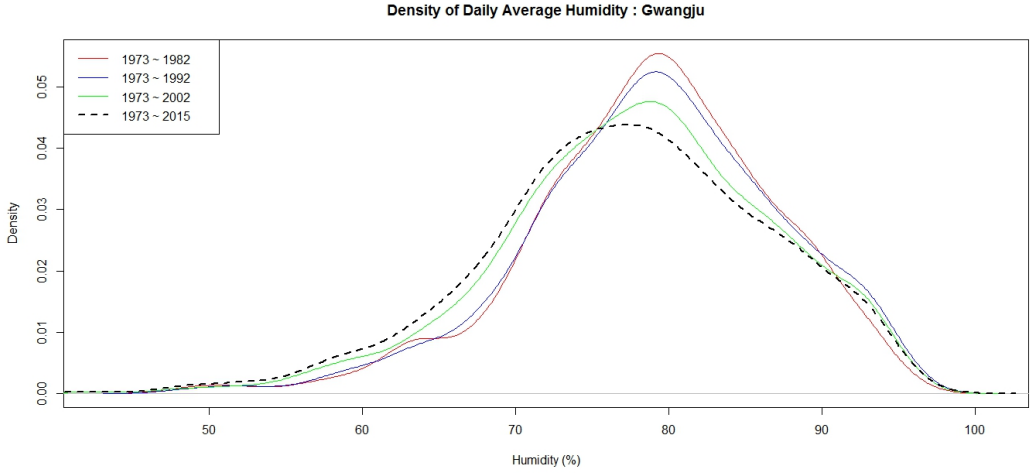

Density of Daily Average Humidity : Chungbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
50
60
70
80
90
100
Humidity (%)

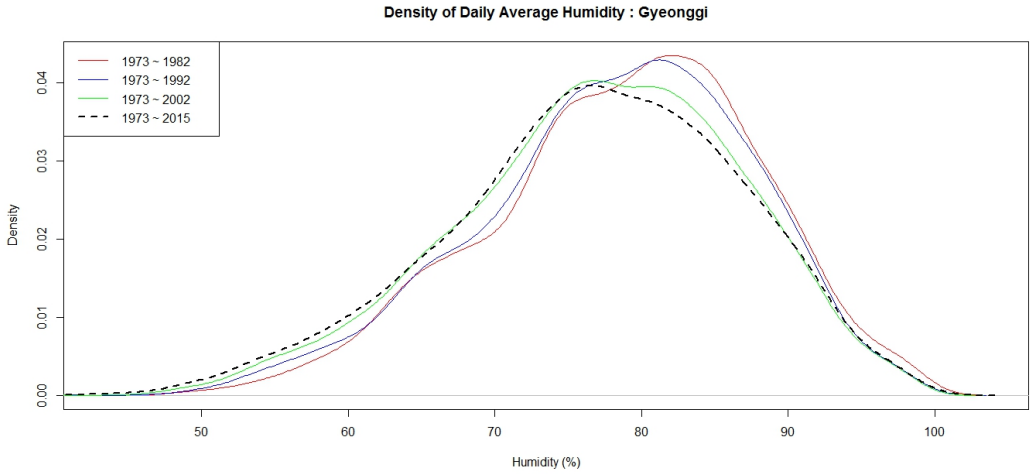

Density of Daily Average Humidity : Chungnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0.05
40
50
60
70
80
90
100
Humidity (%)

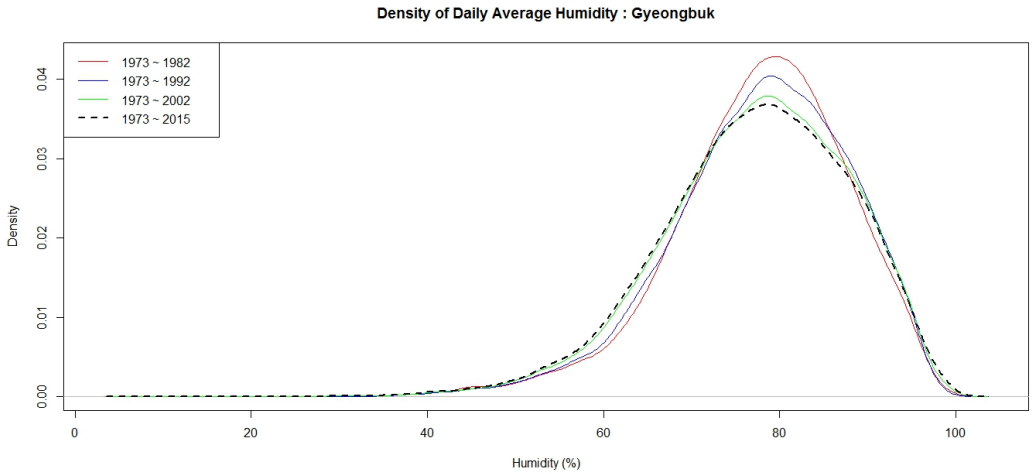

Density of Daily Average Humidity : Daegu
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
40
60
80
100
Humidity (%)

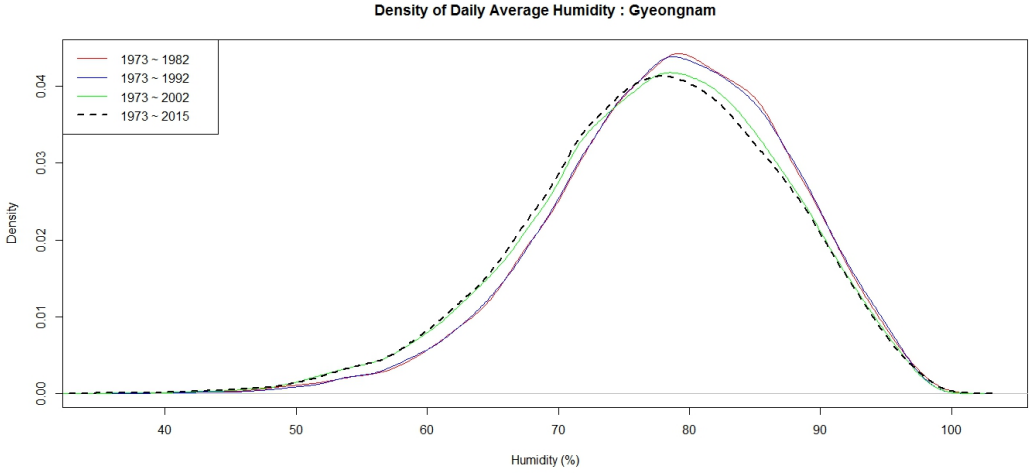

Density of Daily Average Humidity : Daejeon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
50
60
70
80
90
100
Humidity (%)

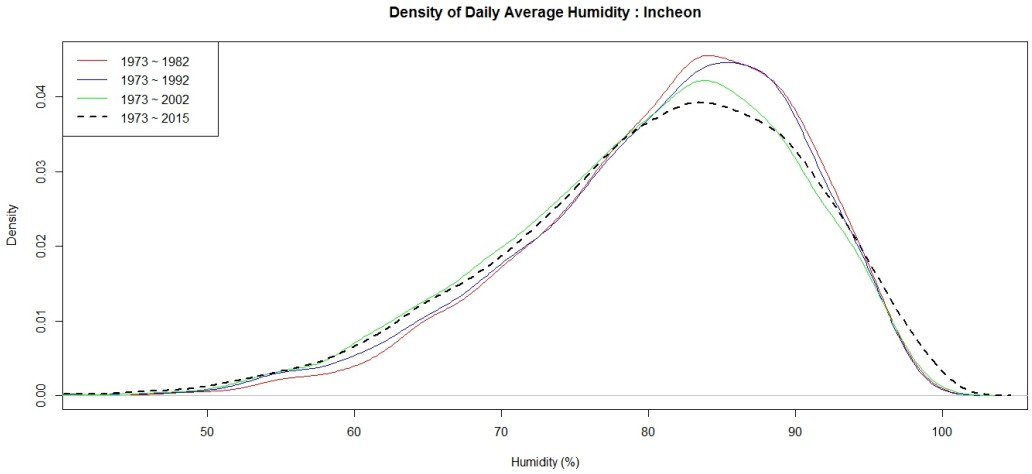

Density of Daily Average Humidity : Gangwon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
40
60
80
100
Humidity (%)

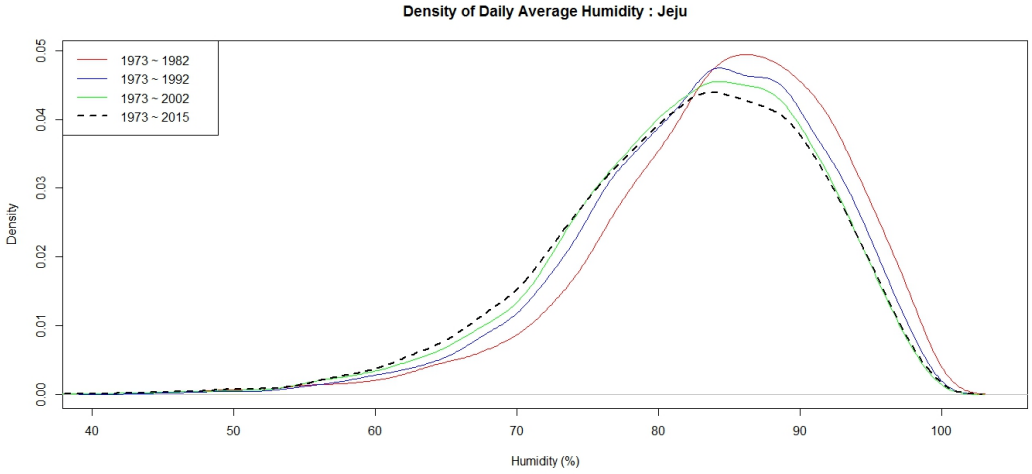

Density of Daily Average Humidity : Gwangju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0.05
50
60
70
80
90
100
Humidity (%)

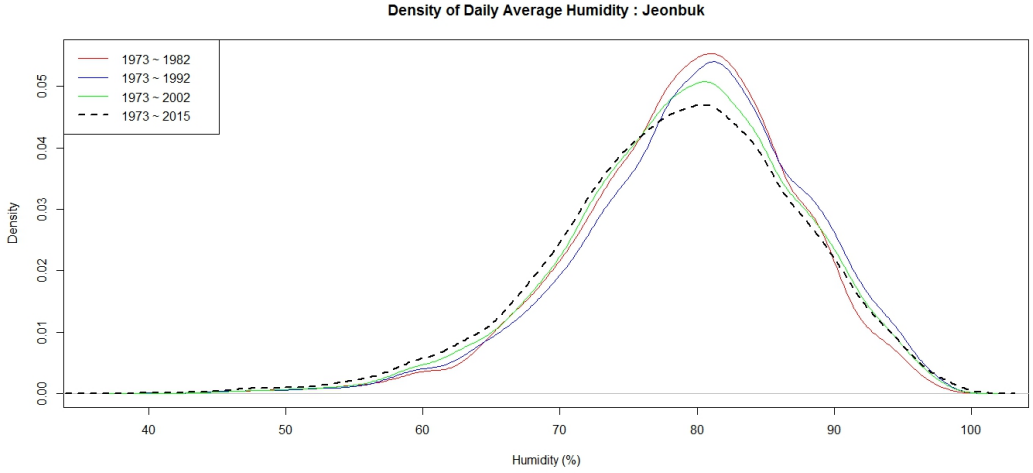

Density of Daily Average Humidity : Gyeonggi
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
50
60
70
80
90
100
Humidity (%)

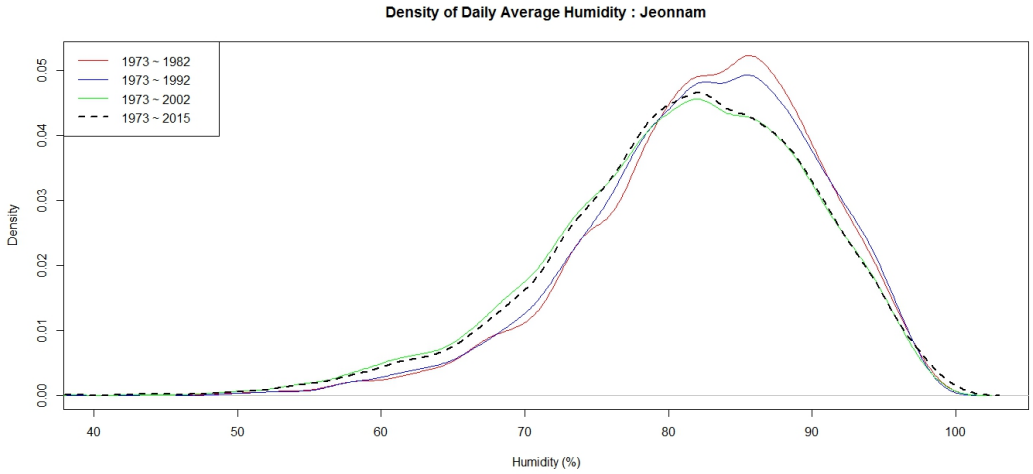

Density of Daily Average Humidity : Gyeongbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0
20
40
60
80
100
Humidity (%)

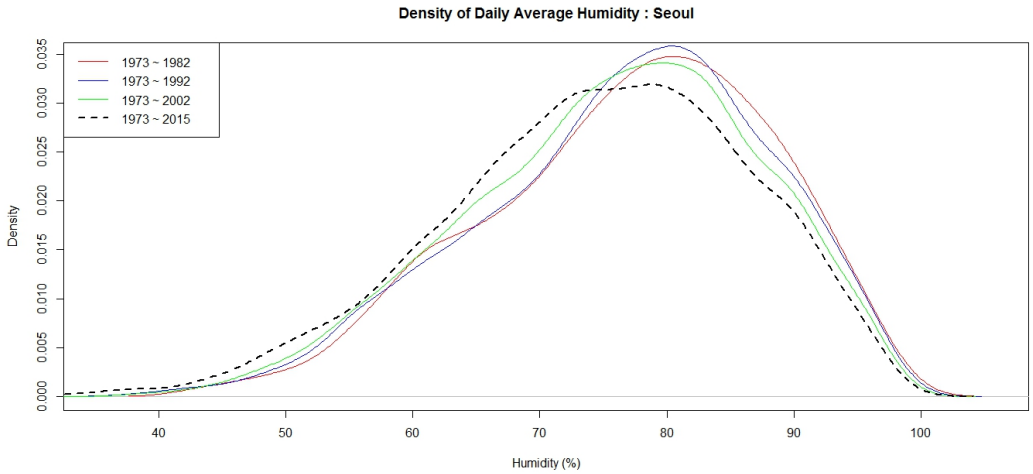

Density of Daily Average Humidity : Gyeongnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
40
50
60
70
80
90
100
Humidity (%)


Density of Daily Average Humidity : Incheon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
50
60
70
80
90
100
Humidity (%)


Density of Daily Average Humidity : Jeju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0.05
40
50
60
70
80
90
100
Humidity (%)


Density of Daily Average Humidity : Jeonbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0.05
40
50
60
70
80
90
100
Humidity (%)


Density of Daily Average Humidity : Jeonnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.01
0.02
0.03
0.04
0.05
40
50
60
70
80
90
100
Humidity (%)


Density of Daily Average Humidity : Seoul
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
40
50
60
70
80
90
100
Humidity (%)

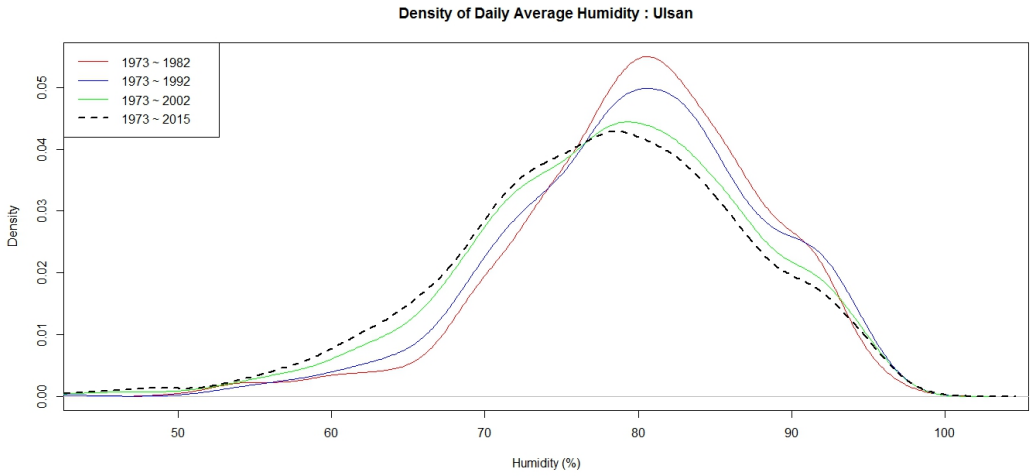


Figure 3.2: Density of Daily Average Relative Humidity by Province.

Here, we present the minimum, mean and maximum from the computed empirical density functions. Especially, the minimums and means show a decreasing trend.

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Busan	Min	41.34	Min	29.39
	Mean	73.00	Mean	67.00
	Max	104.66	Max	104.61
Chungbuk	Min	43.32	Min	22.96
	Mean	73.00	Mean	63.00
	Max	102.68	Max	103.04

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Chungnam	Min	42.01	Min	23.55
	Mean	72.00	Mean	63.50
	Max	101.99	Max	103.45
Daegu	Min	26.80	Min	22.26
	Mean	65.50	Mean	62.50
	Max	104.20	Max	102.74
Daejeon	Min	42.82	Min	24.59
	Mean	73.50	Mean	65.00
	Max	104.18	Max	105.41
Gangwon	Min	25.21	Min	17.13
	Mean	64.50	Mean	60.50
	Max	103.79	Max	103.87
Gwangju	Min	42.85	Min	29.79
	Mean	72.00	Mean	66.20
	Max	101.15	Max	102.61
Gyeonggi	Min	43.04	Min	27.85
	Mean	73.50	Mean	66.00
	Max	103.96	Max	104.15
Gyeongbuk	Min	2.78	Min	3.541
	Mean	53.50	Mean	53.50
	Max	104.22	Max	103.46
Gyeongnam	Min	34.93	Min	30.75
	Mean	69.00	Mean	67.00
	Max	103.07	Max	103.25

Province	Density (1973 $\sim$ 1982)		Density (1973 $\sim$ 2015)	
Incheon	Min	42.62	Min	21.37
	Mean	73.00	Mean	63.00
	Max	103.38	Max	104.63
Jeju	Min	40.46	Min	19.16
	Mean	72.00	Mean	61.00
	Max	103.54	Max	102.84
Jeonbuk	Min	36.4	Min	27.82
	Mean	69.0	Mean	65.50
	Max	101.6	Max	103.18
Jeonnam	Min	40.40	Min	29.96
	Mean	71.50	Mean	66.50
	Max	102.60	Max	103.04
Seoul	Min	35.31	Min	17.87
	Mean	70.50	Mean	61.00
	Max	105.69	Max	104.13
Ulsan	Min	44.85	Min	29.39
	Mean	74.00	Mean	67.00
	Max	103.15	Max	104.61

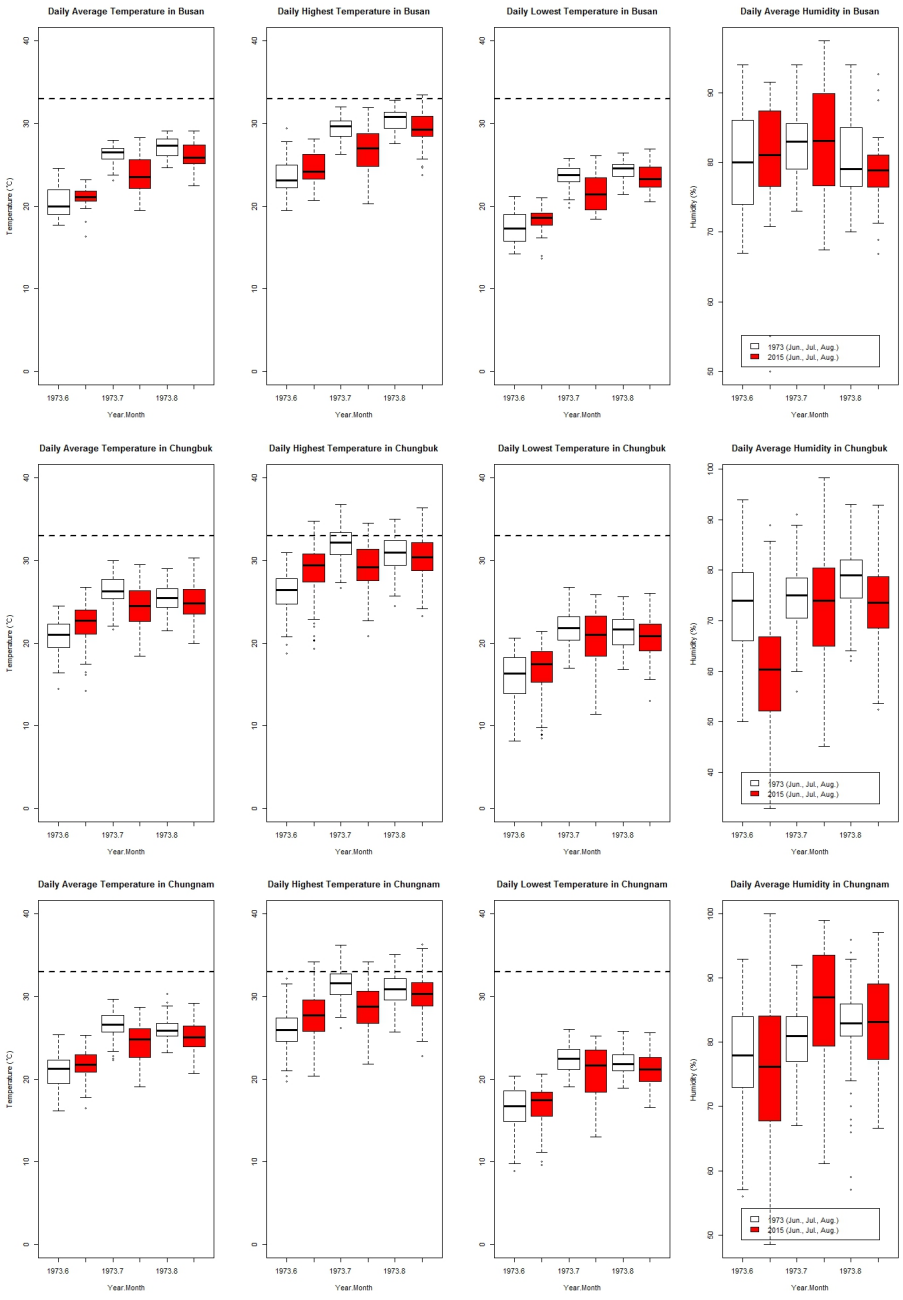
Table 3.2: Minimum, Mean and Maximum Estimates of Density of Daily Average Relative Humidity (%) by Province. density() function computes density estimates with kernel density method. So, this table shows the estimates not the real values from the uesd data.

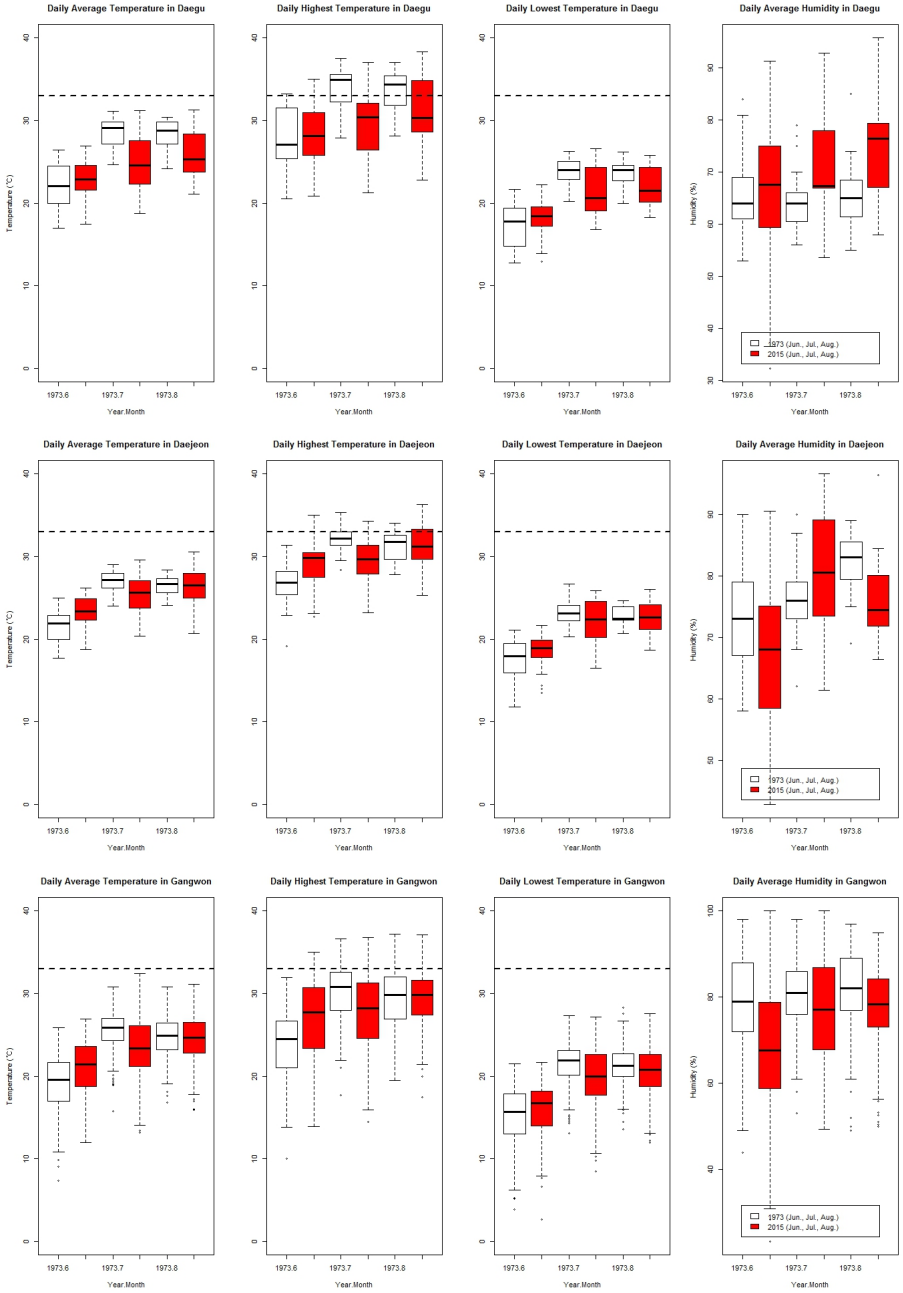
3.3. 1973 vs. 2015

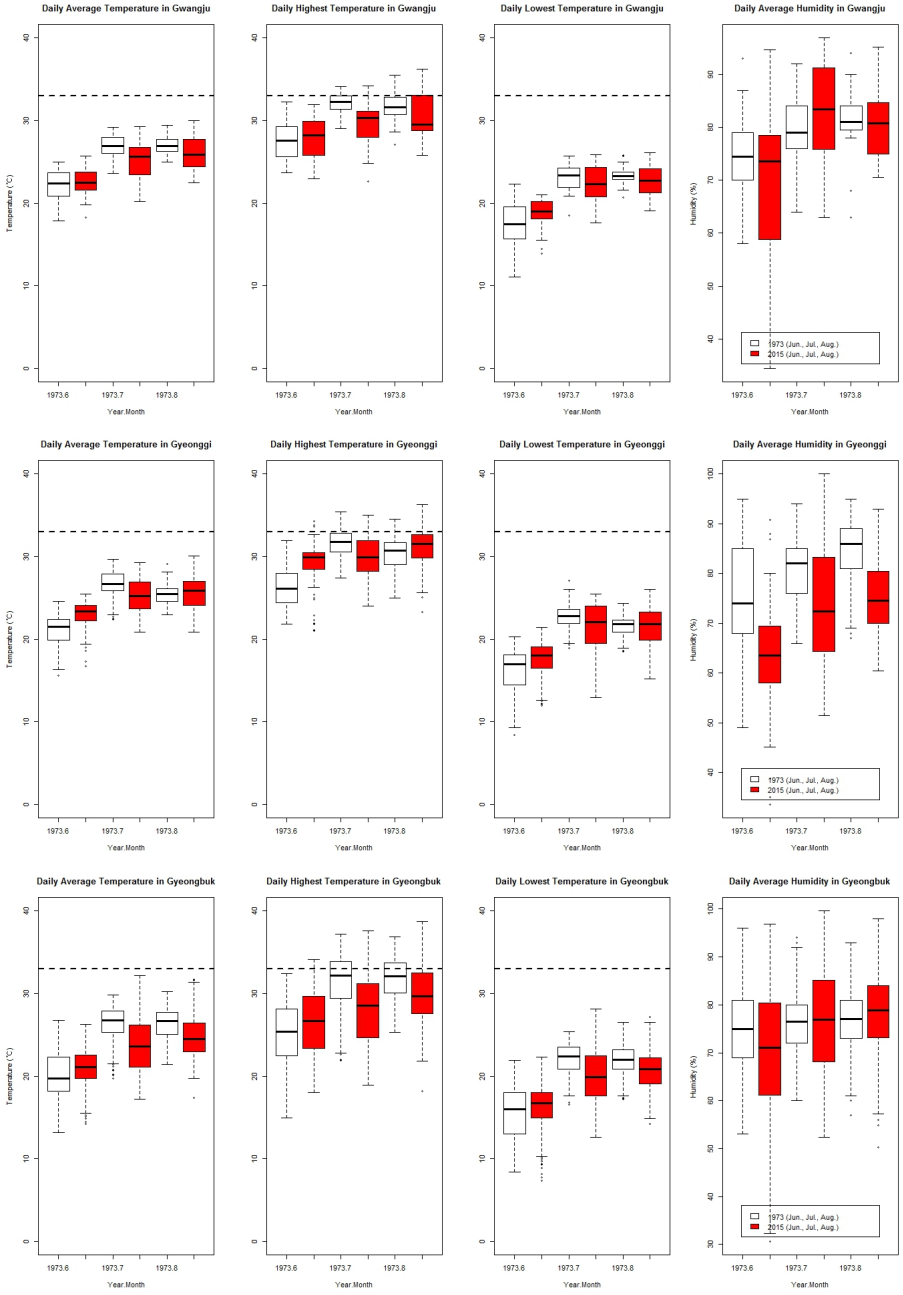
In general, the stationarity assumption is adopted for time-series variable. This means the statistical features, like mean, standard deviation, etc., do not change by time. But, as we have already seen, it is not true. We detected slight changes in empirical (probability) densities in daily highest temperature and daily average relative humidity, and provided them through Section 3.1 and Section 3.2.

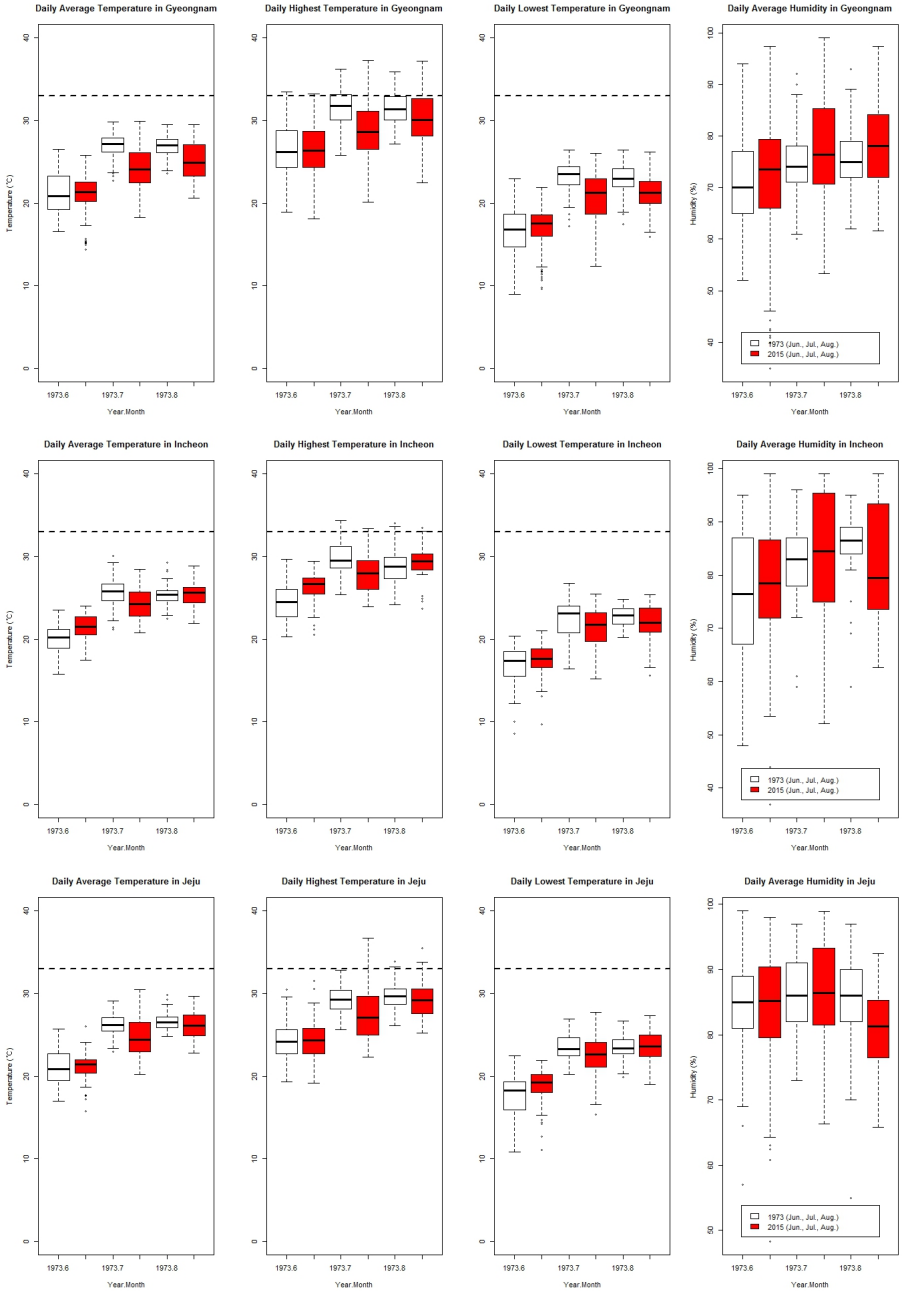
In this section, as we deal with an idea that changes in empirical (probability) densities of weather variable could be the proofs of climate change, we add another basic analysis comparing the first year to the last year of our study period. Our intent is to compare the variables to capture any changes appeared in empirical (probability) densities passing 42 years. Here, we present the boxplots of temperatures and humidity by province. We used R (version 3.2.5) and its boxplot() function. Dashed lines in the temperature boxplots indicate 33 °C which is one of the criteria that KMA declare an alert of heat wave.

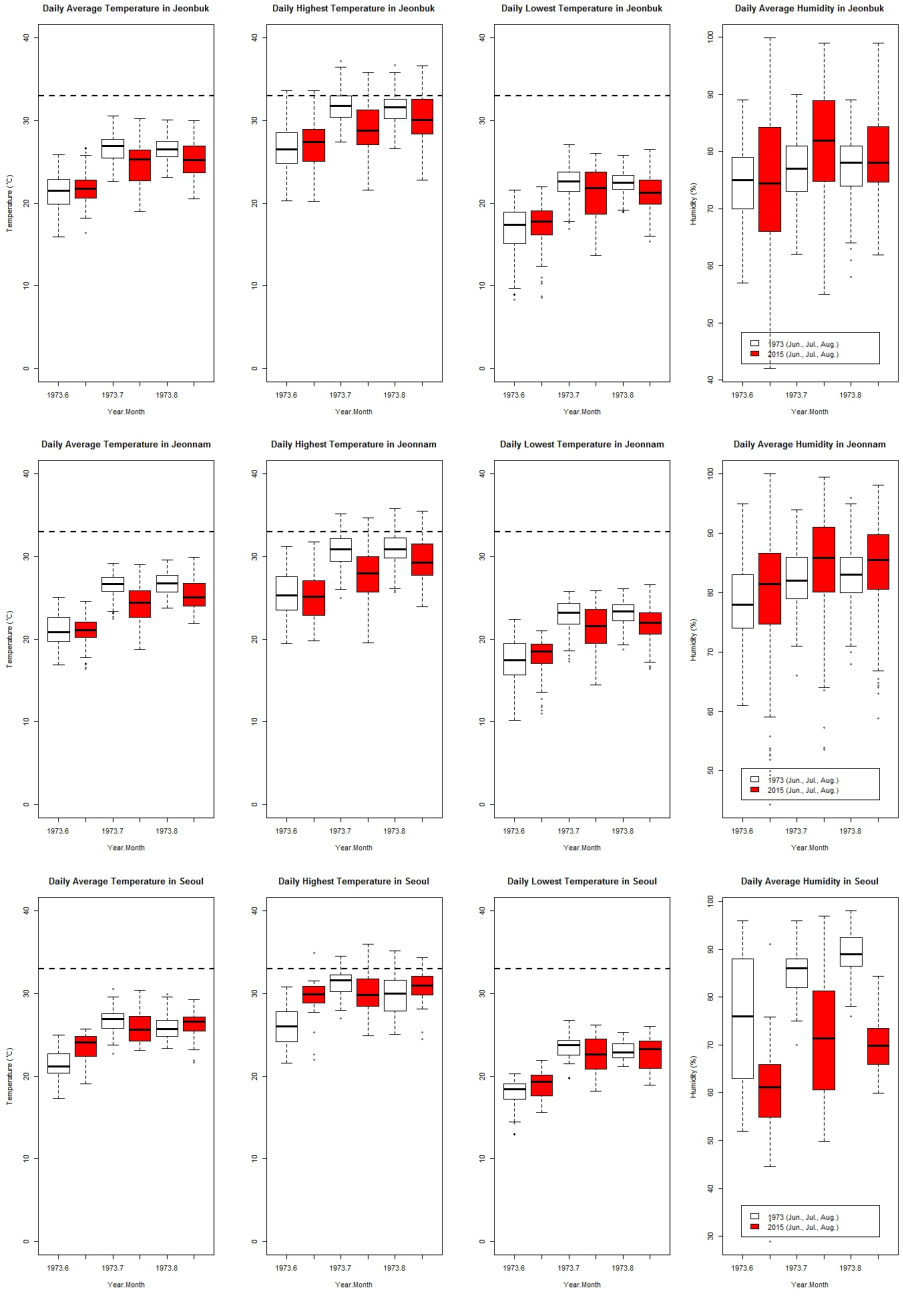
We depicted boxplots by month because of a going up trend of temperature during the summer season (from June to August). Reviewing the boxplots, we can find all of the distributions in 2015 are *different* from those of 1973. Most of boxplots of temperatures seem to move up after 42 years. Furthermore, each range of variable in 2015 is bigger than that in 1973 which implies a bigger variability. Heat wave and heat disorder patients provoked by heat wave have always existed but their risks have changed. If climate change occurs, the probability distributions of weather variables will change. Finally, this will have led to the changes in heat wave and heat disorder risk.


Daily Average Temperature in Busan
Daily Highest Temperature in Busan
Daily Lowest Temperature in Busan
Daily Average Humidity in Busan
Daily Average Temperature in Chungbuk
Daily Highest Temperature in Chungbuk
Daily Lowest Temperature in Chungbuk
Daily Average Humidity in Chungbuk
Daily Average Temperature in Chungnam
Daily Highest Temperature in Chungnam
Daily Lowest Temperature in Chungnam
Daily Average Humidity in Chungnam
Temperature (°C)
Humidity (%)
Year.Month
1973.6
1973.7
1973.8
1973 (Jun., Jul., Aug.)
2015 (Jun., Jul., Aug.)


Daily Average Temperature in Daegu
Daily Highest Temperature in Daegu
Daily Lowest Temperature in Daegu
Daily Average Humidity in Daegu
Daily Average Temperature in Daejeon
Daily Highest Temperature in Daejeon
Daily Lowest Temperature in Daejeon
Daily Average Humidity in Daejeon
Daily Average Temperature in Gangwon
Daily Highest Temperature in Gangwon
Daily Lowest Temperature in Gangwon
Daily Average Humidity in Gangwon
Temperature (°C)
Humidity (%)
Year.Month
1973.6
1973.7
1973.8
1973 (Jun., Jul., Aug.)
2015 (Jun., Jul., Aug.)


Daily Average Temperature in Gwangju
Daily Highest Temperature in Gwangju
Daily Lowest Temperature in Gwangju
Daily Average Humidity in Gwangju
Daily Average Temperature in Gyeonggi
Daily Highest Temperature in Gyeonggi
Daily Lowest Temperature in Gyeonggi
Daily Average Humidity in Gyeonggi
Daily Average Temperature in Gyeongbuk
Daily Highest Temperature in Gyeongbuk
Daily Lowest Temperature in Gyeongbuk
Daily Average Humidity in Gyeongbuk
Temperature (°C)
Humidity (%)
Year.Month
1973.6
1973.7
1973.8
1973 (Jun., Jul., Aug.)
2015 (Jun., Jul., Aug.)


Daily Average Temperature in Gyeongnam
Daily Highest Temperature in Gyeongnam
Daily Lowest Temperature in Gyeongnam
Daily Average Humidity in Gyeongnam
Daily Average Temperature in Incheon
Daily Highest Temperature in Incheon
Daily Lowest Temperature in Incheon
Daily Average Humidity in Incheon
Daily Average Temperature in Jeju
Daily Highest Temperature in Jeju
Daily Lowest Temperature in Jeju
Daily Average Humidity in Jeju
Temperature (°C)
Humidity (%)
Year.Month
1973.6
1973.7
1973.8
1973 (Jun., Jul., Aug.)
2015 (Jun., Jul., Aug.)


Daily Average Temperature in Jeonbuk
Daily Highest Temperature in Jeonbuk
Daily Lowest Temperature in Jeonbuk
Daily Average Humidity in Jeonbuk
Daily Average Temperature in Jeonnam
Daily Highest Temperature in Jeonnam
Daily Lowest Temperature in Jeonnam
Daily Average Humidity in Jeonnam
Daily Average Temperature in Seoul
Daily Highest Temperature in Seoul
Daily Lowest Temperature in Seoul
Daily Average Humidity in Seoul
Temperature (°C)
Humidity (%)
Year.Month
1973.6
1973.7
1973.8
1973 (Jun., Jul., Aug.)
2015 (Jun., Jul., Aug.)

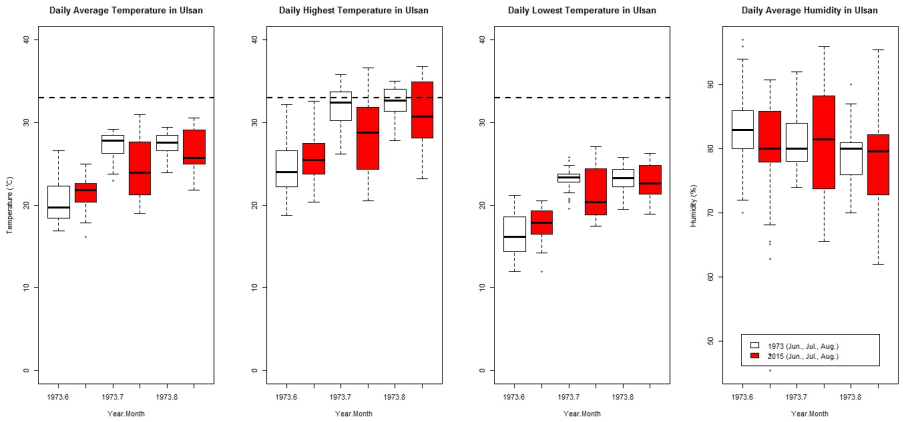


Figure 3.3: Boxplots of Daily Temperatures and Daily Average Relative Humidity by Province. From the left, Daily Average Temperature, Highest Temperature, Lowest Temperature and Average Relative Humidity. Each box displays the boxplots of June, July and August in 1973 (White) and the same months in 2015. The dashed line indicates 33 °C.

We add a simple hypothesis testing result with a help of t.test() function and R (version 3.2.5). We tested the null hypothesis that the range (maximum - minimum) of each variable in 2015 is the same to that in 1973 in each province. This null hypothesis means that the (probability) densities of each variable have not changed by time.

Variable	Month	t-value	df	p-value
Average Temperature	June	-1.1056	15	0.2863
	July	5.4664	15	0.0001
	August	8.5712	15	0.0000
Highest Temperature	June	2.0349	15	0.0599
	July	9.4873	15	0.0000
	August	5.7422	15	0.0000
Lowest Temperature	June	-4.6356	15	0.0003
	July	7.0173	15	0.0000
	August	8.3935	15	0.0000
Average Humidity	June	10.781	15	0.0000
	July	9.5295	15	0.0000
	August	0.7571	15	0.4607

Table 3.3: A Simple Hypothesis Testing Result with Range (Maximum - Minimum). This t.test() is conducted with daily data by month.

Temperature variables show different results from that of humidity variable. For daily average temperature, change in range by time exists except for

June. And, a similar tendency repeats in the results of daily highest temperature and daily lowest temperature. June shows minus t-value and the change in range of June shows the weakest p-value to reject the null hypothesis than July and August.

On the contrary, daily average relative humidity shows the reverse order. The most powerful p-value appears for June. Accepting results, we can assume the variabilities of densities have appeared even though it is not obviously relevant with climate change.

Here is another result of hypothesis testing with standard deviation of each variable. Null hypothesis means that the standard deviation of each variable in 2015 is same to that in 1973. We also gained the same result that we had already gotten with range. This implies that the variabilities of temperatures (range and standard deviation) and humidity are not same as time flows. With these results, we can get an impression that the variabilities of our variables are elevated regarding the ranges of the boxplots and the standard deviations. Especially, increase in temperatures seems to imply that heat disorder risk would arise.

Variable	Month	t-value	df	p-value
Average Temperature	June	-6.3987	15	0.0000
	July	8.1727	15	0.0000
	August	10.385	15	0.0000
Highest Temperature	June	1.0317	15	0.3186
	July	11.185	15	0.0000
	August	4.3067	15	0.0006
Lowest Temperature	June	-9.106	15	0.0000
	July	12.6110	15	0.0000
	August	9.2296	15	0.0000
Average Humidity	June	5.4813	15	0.0001
	July	13.2380	15	0.0000
	August	3.7986	15	0.0017

Table 3.4: A Simple Hypothesis Testing Result with Standard Deviation. This t.test() is conducted with daily data by month.

3.4. Histogram of Heat Disorder Incidence

In this section, we argue about heat disorder incidence (report counts) from KCDC data. Unfortunately, this data is shorter than KMA data. Because its reporting system has been just established. But this data contains the information how human beings would be affected by environment. But, this allowed us to calculate the risk of human beings due to the environment. We tried to think how to utilize this data for the prevention of disease.

Firstly, we start with total heat disorder incidence by year. These maps are drawn by R (version 3.2.5), ggplot2 package (Wickham and Cheong, 2016) and sp package (Edzer Pebesma *et al*, 2016).

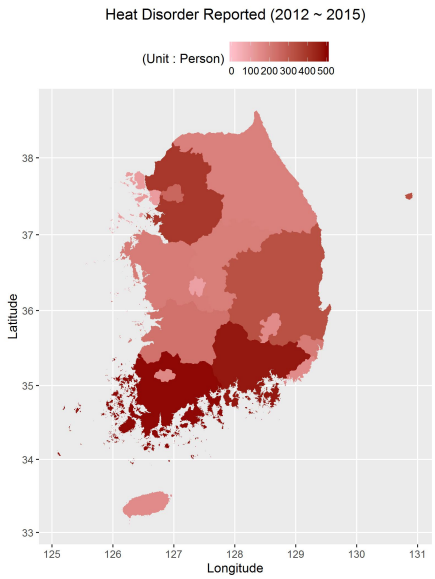


Figure 3.4: Total Heat Disorder Incidence by Province for 4 Years (2012 $\sim$ 2015).

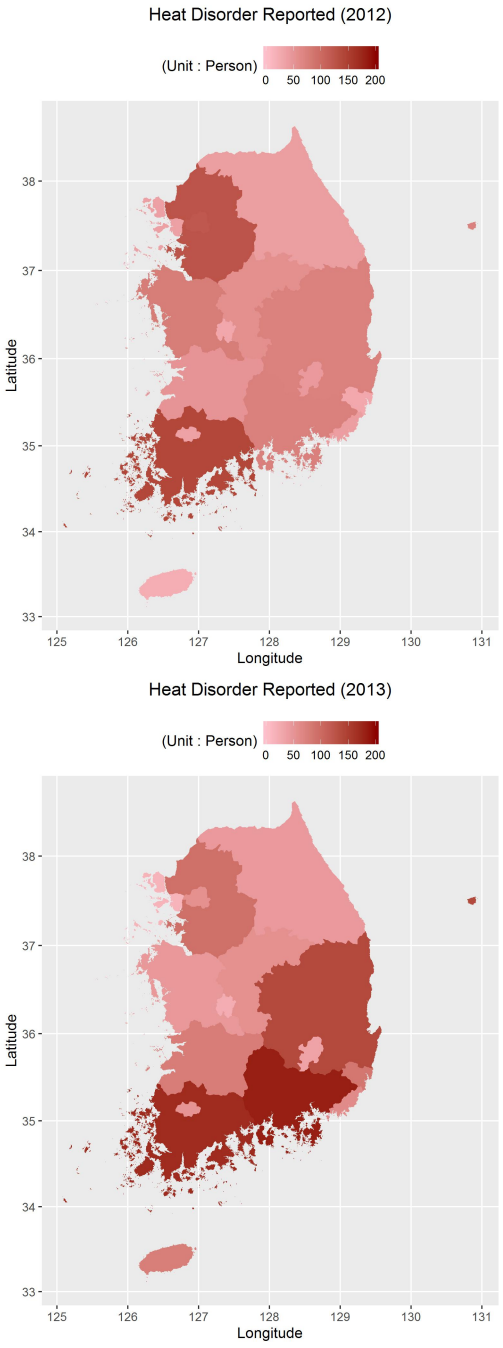


Figure 3.5: Total Heat Disorder Incidence in by Province 2012 & 2013.

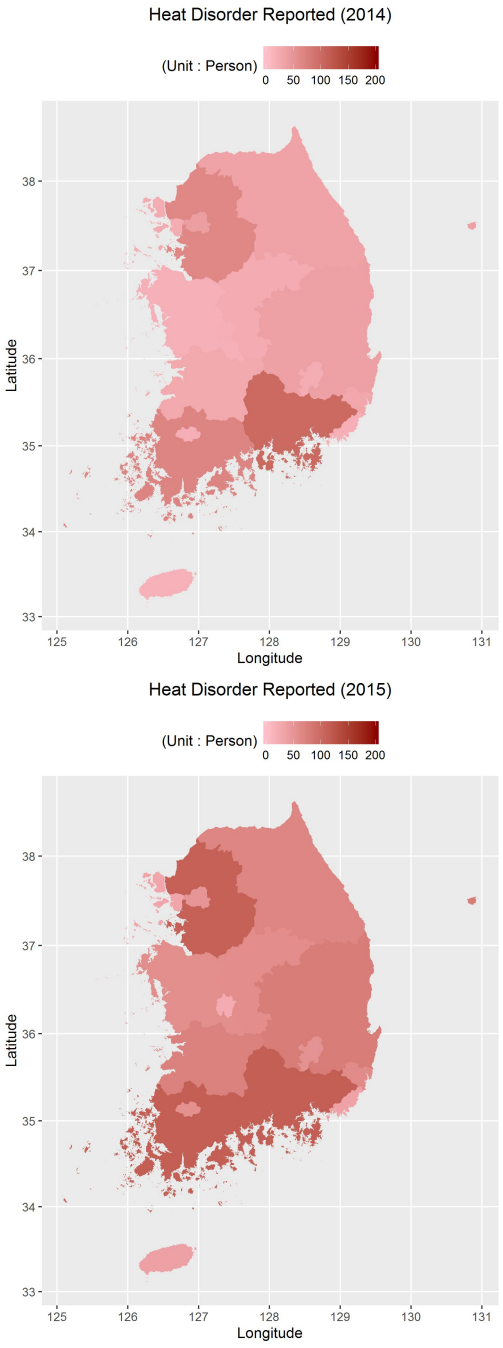
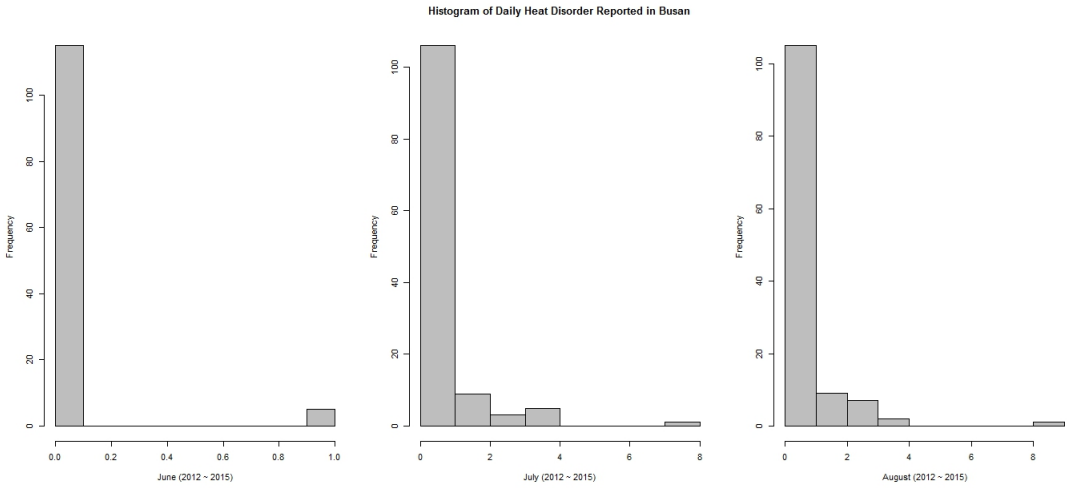
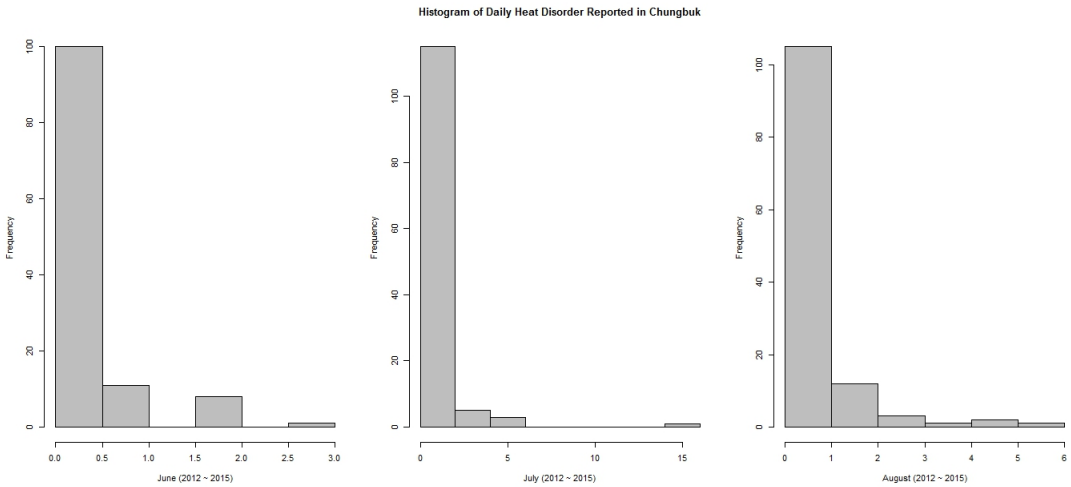


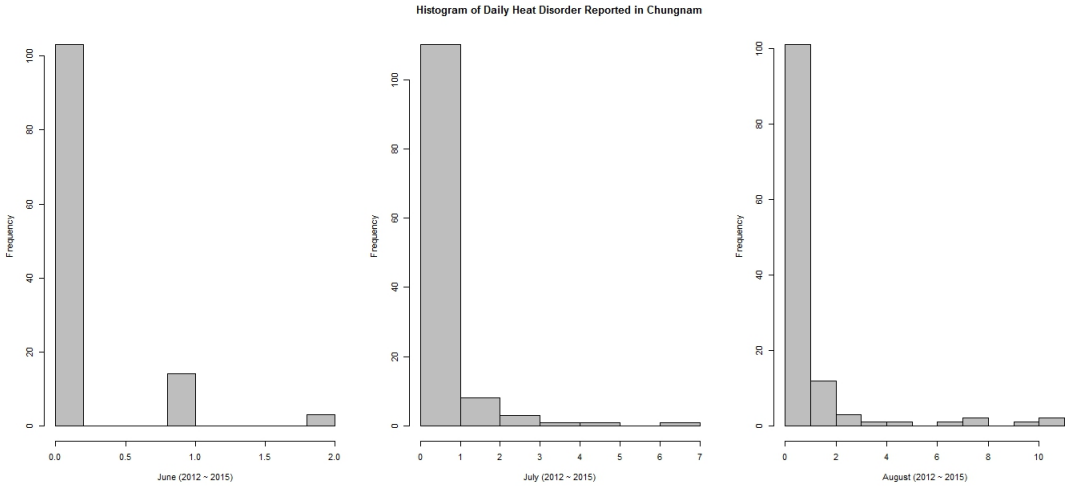
Figure 3.6: Total Heat Disorder Incidence by Province in 2014 & 2015.

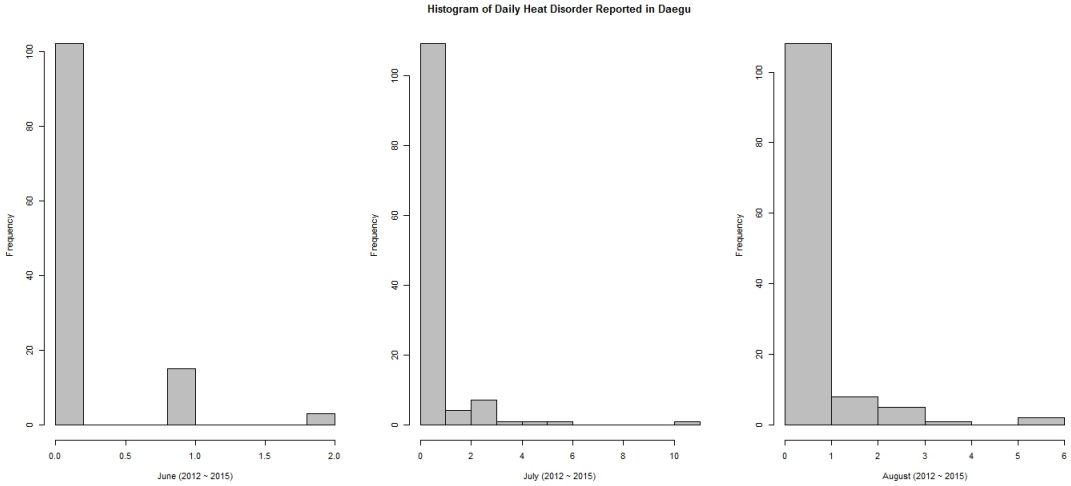
We can capture a couple of things from these yearly maps. There are yearly changes in the rank by regional heat disorder incidence. For example, Seoul shows the third largest incidence in 2012. But, it is not from 2013 to 2015. This means there is not the region where it is always dangerous for heat disorder. We can imagine it is hotter if it is located in more southern region. But, this location factor does not seem to be that deterministic if we observe Gyeonggi. The second thing is a spatial correlation might be assumed. Each province shows a similar incidence to its adjacent provinces.

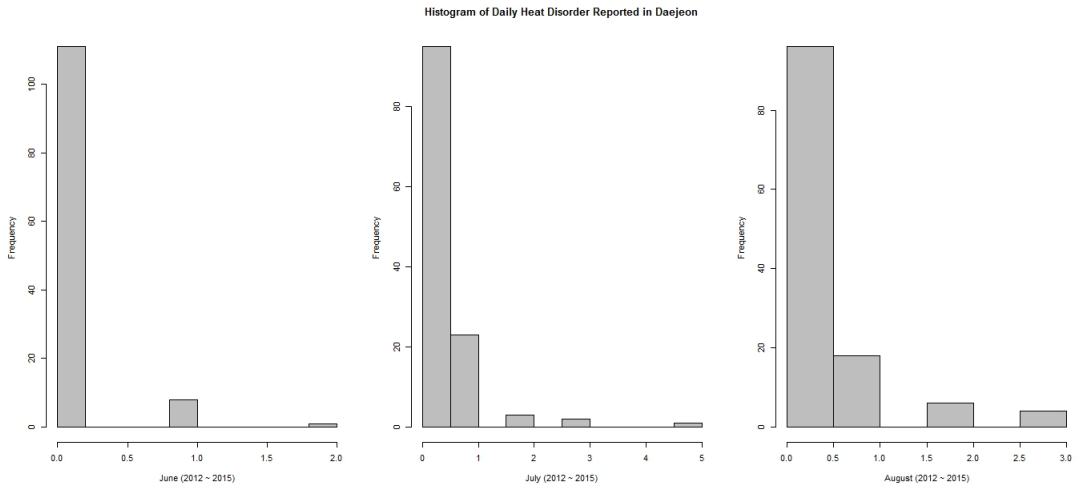
We can easily get histograms of daily heat disorder incidence by province with a help of hist() of R (version 3.2.5). We depicted histograms by month with 4-year data because of the increasing trend of temperature variables during summer season. Corresponding to this trend, each monthly histogram shows different maximum values on its x-axis. In addition to this, we can also observe a lot of zero frequencies at the left.

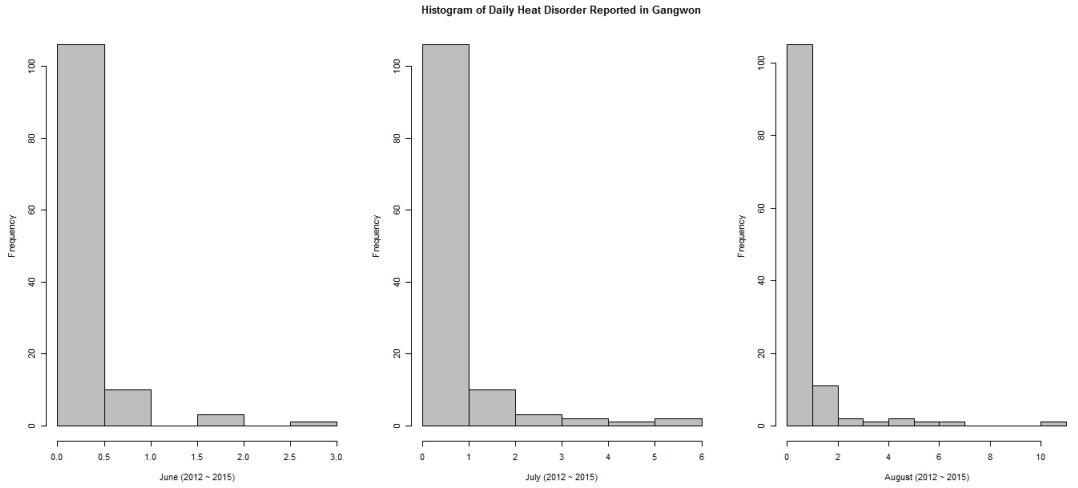


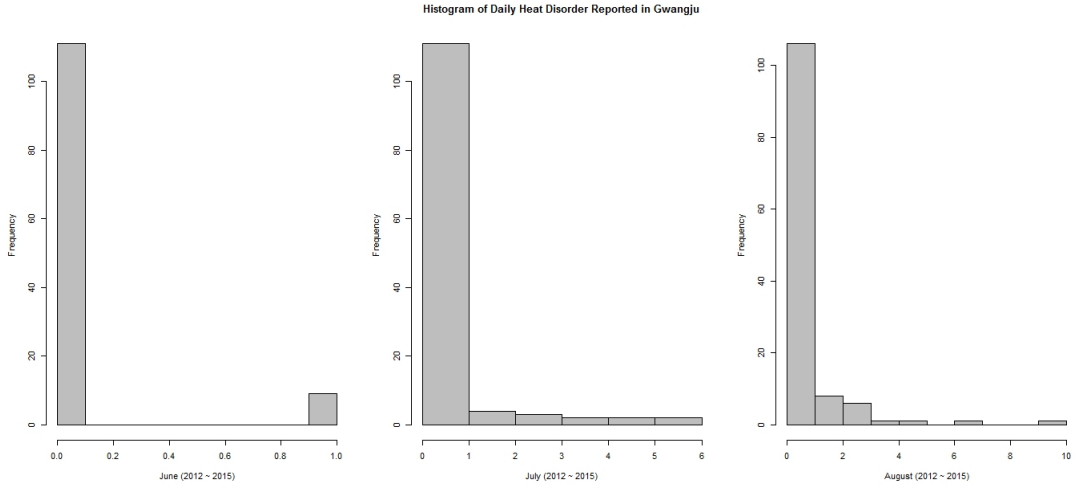

Histogram of Daily Heat Disorder Reported in Chungbuk
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

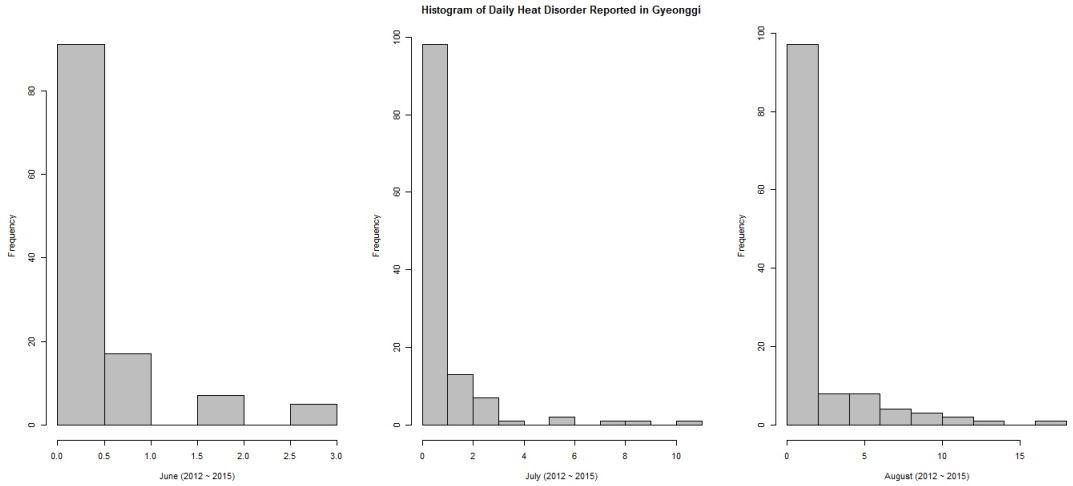

Histogram of Daily Heat Disorder Reported in Chungnam
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

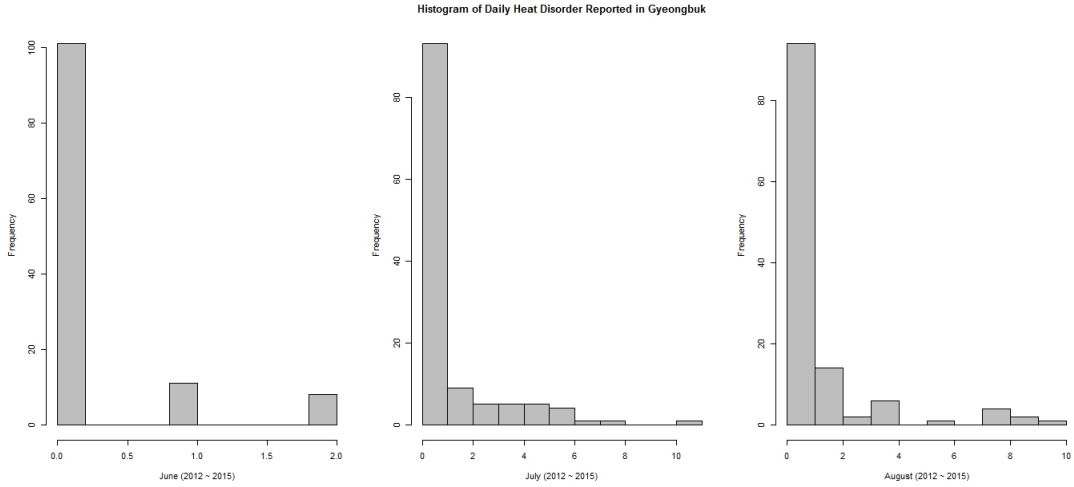

Histogram of Daily Heat Disorder Reported in Daegu
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

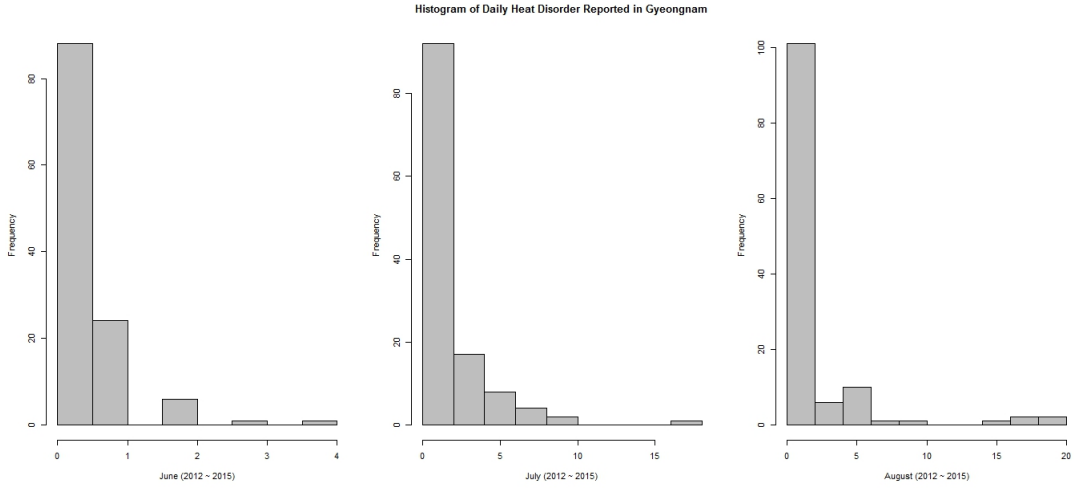

Histogram of Daily Heat Disorder Reported in Daejeon
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

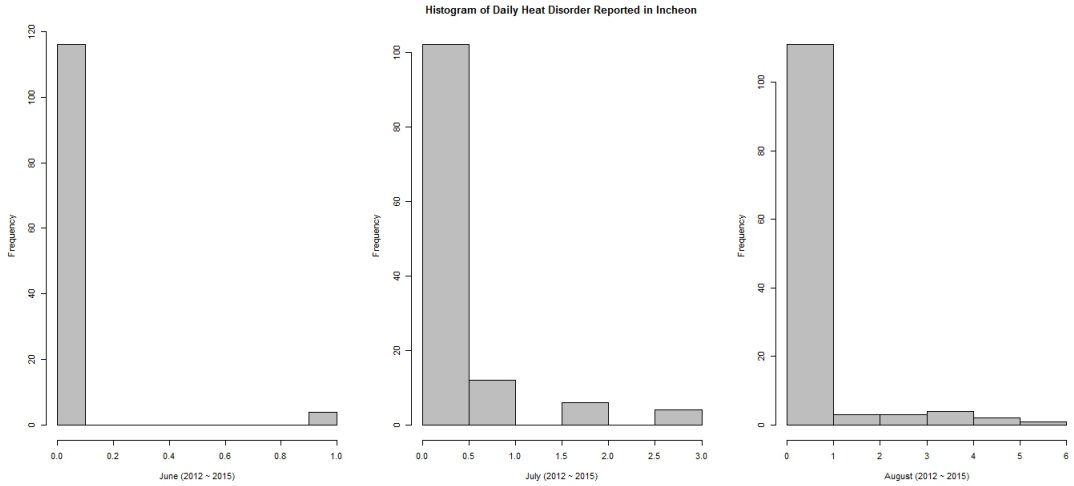

Histogram of Daily Heat Disorder Reported in Gangwon
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

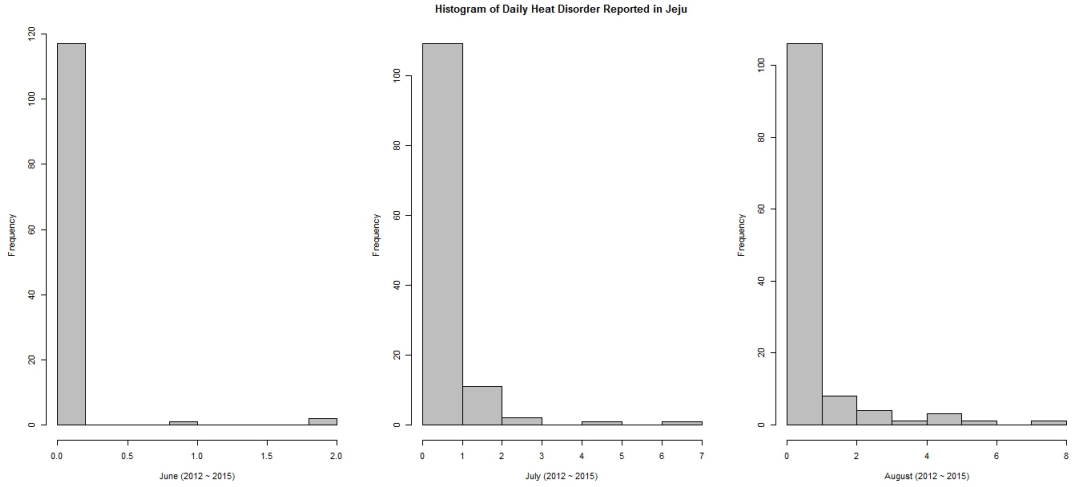

Histogram of Daily Heat Disorder Reported in Gwangju
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

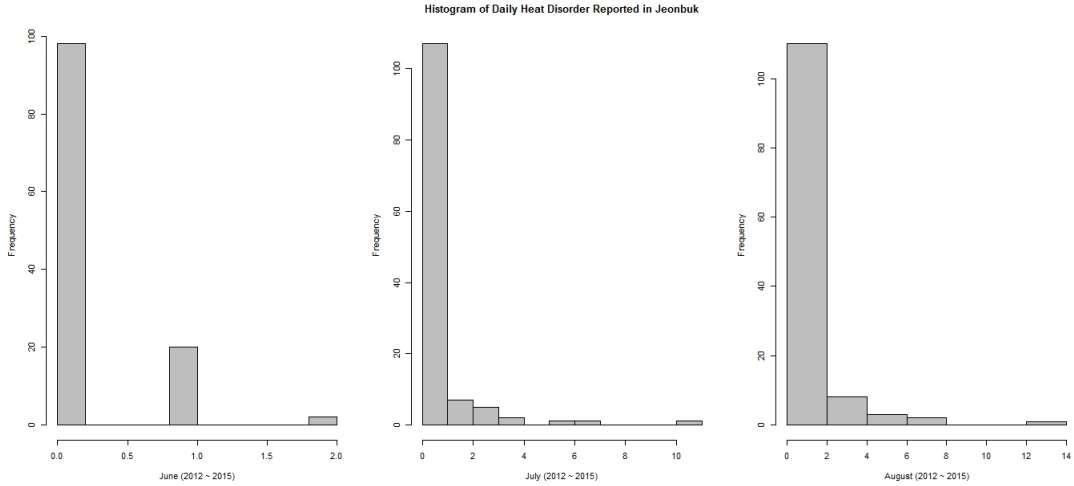

Histogram of Daily Heat Disorder Reported in Gyeonggi
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

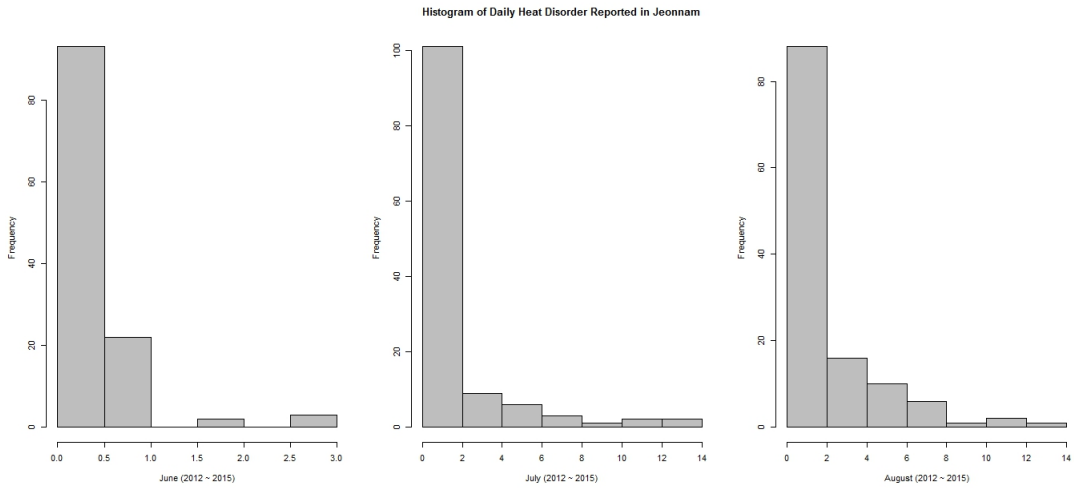

Histogram of Daily Heat Disorder Reported in Gyeongbuk
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

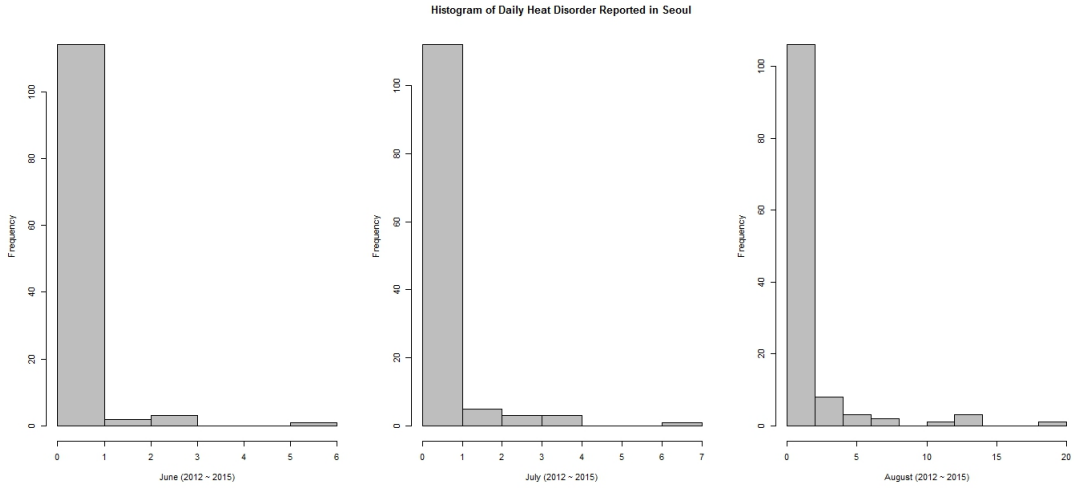

Histogram of Daily Heat Disorder Reported in Gyeongnam
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)


Histogram of Daily Heat Disorder Reported in Incheon
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)


Histogram of Daily Heat Disorder Reported in Jeju
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)


Histogram of Daily Heat Disorder Reported in Jeonbuk
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)


Histogram of Daily Heat Disorder Reported in Jeonnam
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)


Histogram of Daily Heat Disorder Reported in Seoul
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

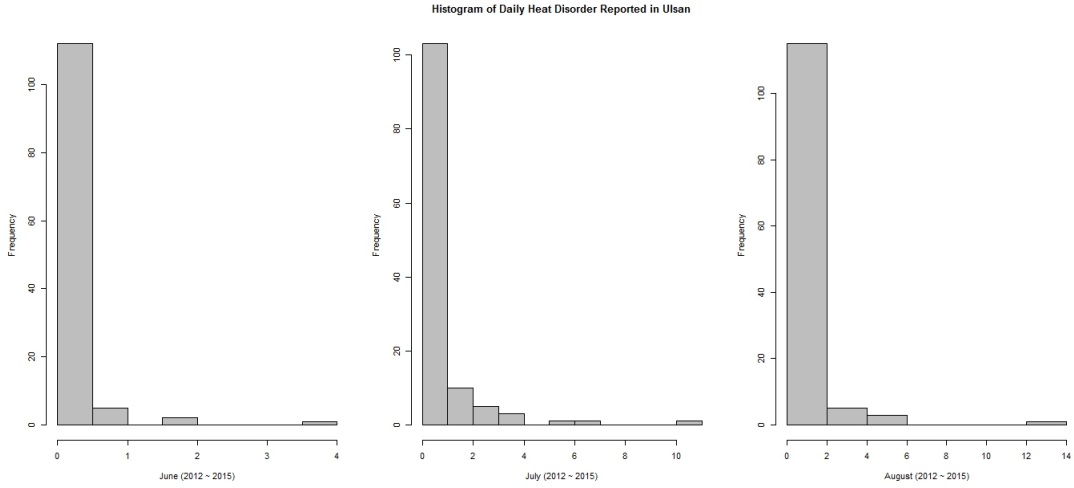

Histogram of Daily Heat Disorder Reported in Ulsan
Frequency
June (2012 ~ 2015)
July (2012 ~ 2015)
August (2012 ~ 2015)

Figure 3.7: Histogram of Daily Heat Disorder Incidence by Month & Province. This Histograms are depicted with 4-year data.

It is notable that the temperature variables and humidity variable fluctuate by day, there is an increasing trend of temperature during a summer season. And, it is hottest in August and the biggest incidence is usually recorded in August. These natures should be considered and used to analze the relationship between the explanatory variables (Temperatures, Humidity, etc.) and the response variable (Heat Disorder).

3.5. Population

It is required to note about the population variable though it is not a main factor of our study. This bar chart is depicted with R (version 3.2.5) and ggplot2 package (Wickham and Cheong, 2016).

This population data from KOSIS does not show any yearly big change in population by province. With no surprise, the population of Gyeonggi is the biggest during our study period. The second is that of Seoul. South Korea is highly concentrated onto the metropolitan area around City Seoul. Recalling the heat disorder incidence maps in Section 3.3, we can capture that the bigger population is not linked directly to the bigger incidence of heat disorder. If population is the main cause of heat disorder incidence, Gyeonggi and Seoul should have been the provinces where the biggest number of patients recorded. But, it is not like that. This fact is an important thing

when we analyze about the relationship between the cause and its effect.

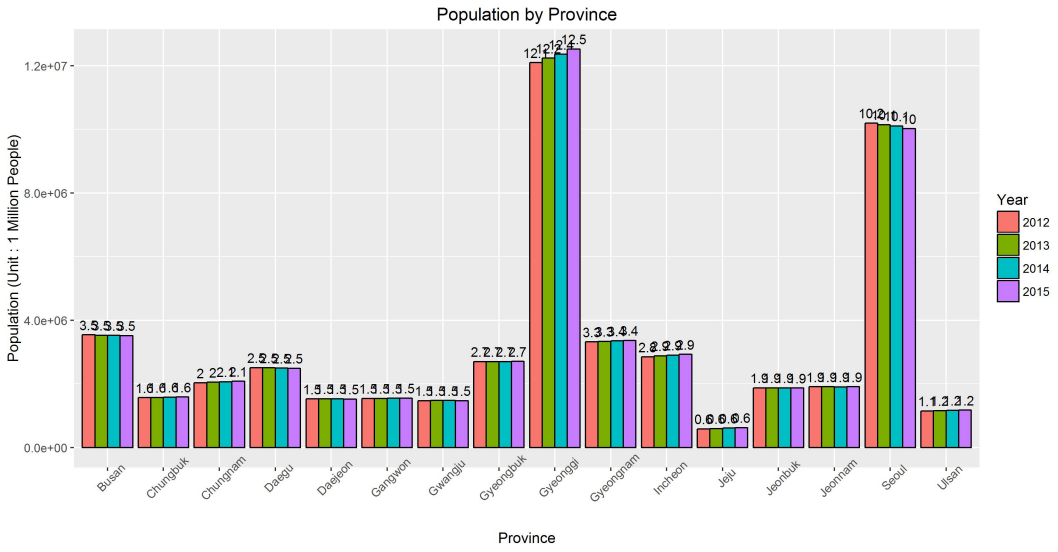


Figure 3.8: Yearly Population by Province from KOSIS.

3.6. The Relationship: Temperature, Humidity and Heat Disorder

In this section, we try to draw a relationship between the temperature variables, humidity variable and the heat disorder variable. One simple way to see a relationship among variables is a scatter plot. We used re-classified temperature variables, re-classified humidity variable and heat disorder incidence variable and plot() function R (version 3.2.5).

The scatter plots by province are also available in Appendix. In these plots, we can observe a linear relationship between temperatures and heat

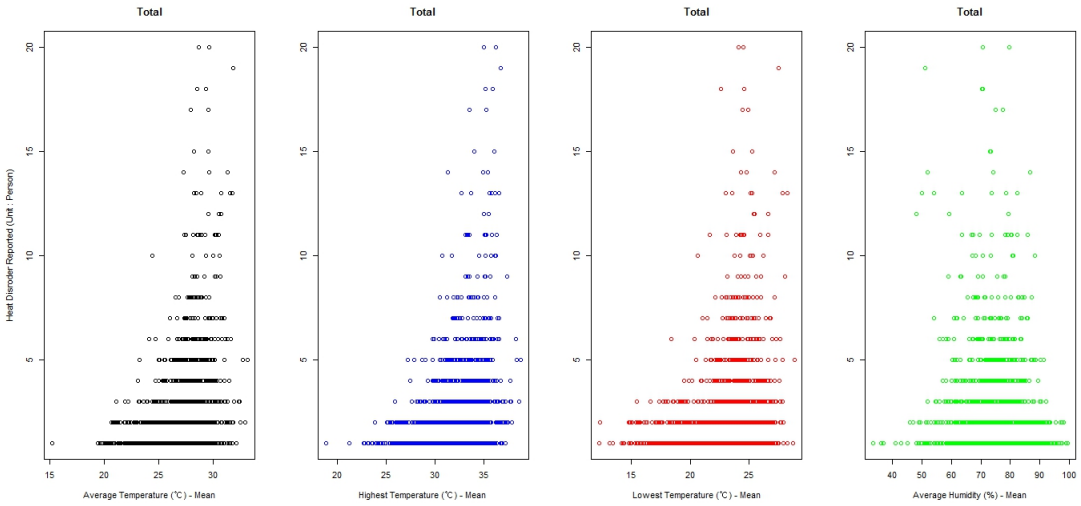


Figure 3.9: The Reltionship Between the Causes (Temperature, Humidity) and the Effect (Heat Disorder Incidence) at Total. We computed mean as a representative value by province during the re-classification process of temperature and humidity. And, we matched this data with heat disorder incidence of 4 years. Black is for daily average temperature, blue is daily highest temperature, the red is daily lowest temperature and the green is the daily average relative humidity.

disorder. The relationship between humidity and heat disorder is not that clear. Now, we approach to the solution to the second question of this study.

(2) *Second, how can we calculate the risk of climate as an effect to health?*

The results give us the information on changes in weather variables such as temperatures and humidity since 1973. We observed the steady increasing trend and variability of temperatures which lead to entail in higher risk of hot weather on our health in every province. Now, we can compute the risk of climate, especially *the hot weather*. We will discuss this in Chapter 4.

Chapter 4

Model Analysis

Regression analysis has been conducted to investigate a relationship of cause and effect under the assumption of normal distribution on response variable. General regression analysis is not applicable without this assumption.

In the field of disease mapping or epidemiology, the Generalized Linear Model (GLM) method has been broadly applied to regression analysis with non-normal data like logsitic regression with binomial respose variable. Our data of interest is the heat disorder incidence (counts) data by province. Generally, count data is known to follow Poisson distribution. Therefore, Poisson distribution has been widely adopted for disease mapping analysis or epidemiological study.

Meanwhile, our problem confronts an unobserved feature, the spatial dependence. Recalling the maps presented in Setion 3.4, we can expect *a similarity* in heat disorder incidence between adjacent provinces. This spatial dependence is much more difficult since spatial location is acting as a sur-

rogate for unobserved covariates; we need to choose an appropriate spatial model (Wakefield, 2007).

This spatial dependence or correlation can be dealt as a random effect in a generalized linear model. According to Lee and Nelder (2006), fixed effects can describe systematic mean patterns such as trend, while random effect may describe either correlation patterns between repeated measures within subjects or heterogeneities between subjects or both. The correlation can be represented by saturated random effects.

In general regression analysis, we can get the coefficients for fixed effect terms from the result so, random effects missed can appear as residuals. As Wakefield (2007) expressed as unobserved covariates, random effects might be neglected because of its difficulty to detect.

However, Lee and Nelder provide us an alternative way to analyze these kinds of problem; Hierarchical Generalized Linear Model (HGLM). In this study, we computed the relationship between temperature, humidity and heat disorder incidence by HGLM.

4.1. Method; Hierarchical GLM

Lee and Nelder (1996) have suggested a definition of HGLM as below:
(1) Conditional on random effects u, the responses y follow a GLM family, satisfying

$$E(y \mid u) = \mu \qquad \text{and} \qquad var(y \mid u) = \phi V(\mu)$$

The kernel of loglihood is given by

$$\sum\{y\theta - b(\theta)\}/\phi, \qquad \text{canonical parameter} : \theta = \theta(\mu)$$

The linear predicter takes the form below where $v = v(u)$, for some monotone function $v()$, are the random effects and β are the fixed effects.

$$\eta = g(\mu) = X\beta + Zv$$

(2) The random component u follows a distrbution conjugate to a GLM family of distrbutions with parameters λ.

This method uses h-loglihood for inferences where (ϕ, λ) are dispersion parameters as below.

$$h \equiv logf_{\beta,\phi}(y \mid v) + logf_{\lambda}(v)$$

4.2. Model Description

In general, GLM gives us a result only on fixed effects($X\beta$). Meanwhile, a result of HGLM consists of two parts such as fixed effects ($X\beta$) and random effects (Zv).

To evaluate a modeling result, we refer to R^2 in the case of normal regression analysis. For GLM, we cite Akaike Information Criterion (AIC) and deviance. AIC judges a model by how close its fitted values tend to be to the true mean values, in terms of a certain expected value. We can use the AIC to aid in variable selection with many potential predictors. Out of a set of candidate models, we identify the one with smallest AIC. However, the

models with similar AIC values are also of interest (Agresti, 2013).

And, the definition of deviance is provided as below:

$$D(y;\ \hat{\mu}) = -2[L(\hat{\mu};\ y) - L(y;\ y)]$$

Here is provided the definition of conditional (assuming a random effect) deviance as below :

$$D = D(y;\ \hat{\mu}) = -2\{l(\hat{\mu};\ y \mid v) - l(y;\ y \mid v)\}$$

Furthermore, Lee and Nelder (1996) propose to use the three deviances based upon $f_\theta(y, v)$, $f_\theta(y)$ and $f_\theta(y \mid \hat{\beta})$ for testing various components of HGLMs. For testing random effects they recommend using the deviance $-2h$, for fixed effects $-2l$ and for dispersion parameters $-2logf_\theta(y \mid \hat{\beta})$. When l is numerically hard to obtain, they used $p_v(h)$ and $p_{\beta,v}(h)$ as approximations to l and $logf_\theta(y \mid \hat{\beta})$.

HGLM has the same structure with GLM. Additionally, it assumes unobserved random effect as a variable. This assumption (condition) applies to its likelihood equation for inferences, deviances and conditional AICs.

Although we have many criteria to evaluate the goodnes of fit of model as previously mentioned, according to Agresti (2013), in selecting a model from a set of candidates, we are mistaken if we think that there is a *"correct"* one. Any model is a simplification of reality. So, when we choose a model, we should justify our choice considering our objectives to analyze.

4.3. Model Interpretation

We conducted Poisson GLM and Poisson HGLM with a log link on the bases of findings we discussed in Chapter 3. We chose regional monthly count sum (3 months for 4 years) as response variable because of an abundance of zero values in daily count data. We used re-classified temperature (mean of daily highest temperature) and re-classified humidity (mean of daily average relative humidity) each month by year. And, we included regional population by year into our models. The indexes: i refers to year, j does to month and k does to province. v_k represents unobserved area-specific random effects.

$$\eta_{ijk} = log\mu_{ijk} = tempH_{ijk} + humidityA_{ijk} + tempA_{ijk} + month_j + \\ + province_k + province_k : tempH_{ijk} \qquad (4.1)$$

$$\eta_{ijk} = log\mu_{ijk} = tempH_{ijk} + humidityA_{ijk} + population_{ik} + tempA_{ijk} + \\ + month_j + province_k + province_k : tempH_{ijk} \qquad (4.2)$$

$$\eta_{ijk} = log\mu_{ijk} = tempH_{ijk} + humidityA_{ijk} + tempA_{ijk} + month_j + \\ + province_k + province_k : tempH_{ijk} + v_k \qquad (4.3)$$

tempH means the mean of daily highest temperature of each month and year, humidityA does the mean of daily average relative humidity of each month and year, tempA does the standard deviation of daily average temperature of each month and year, month has three levels like June, July

and August and v_k is a term for an area(province)-specific random effect. We checked the correaltion between the mean of daily highest temperature and the standard deviation of daily average temperature in advance. We confirmed that the correlation between the two is close to 0 (No linear dependence).

Equation(Model) 4.1 is a Poisson GLM not assuming any random effect.

Equation(Model) 4.2 is a Poisson GLM including population variable but not assuming any random effect.

Equation(Model) 4.3 is a Poisson HGLM assuming an area-specific random effect following Markov Random Field (MRF) in which $[var(v)]^{-1} = (I - \rho M)/\lambda$ where M is the incidence matrix for neighbours of areas.

Before going to a comparison of estimates, it is necessary to review the deviances of each model. We obtained the calculation results with the help of R (version 3.3.2), dhglm package (Noh and Lee, 2015) and unpublished codes.

Table 4.1 shows the deviances to check our models. For the mean parameters, we can use $-2p_\beta(h)$ and $-2p_v(h)$, and for the dispersion parameters, $-2p_\beta(h)$ and $-2p_{\beta,v(h)}$ can be used. If we add the population variable, we can see an improvement in mean parameter estimation but no improvement in that of dispersion parameter. And, if we add the random effect term into our model, the deviances tend to increase. We can check further with the residual plots which are provided by dhglm package.

With residual plots, we can observe how the residuals are distributed

Deviance	Model 4.1	Model 4.2	Model 4.3
-2ML$(-2h)$	1227.63	1215.66	1236.02
-2RL$(-2p_\beta(h))$	1371.54	1387.91	n/a
-2RL$(-2p_v(h))$	n/a	n/a	1292.25
-2RL$(-2p_{\beta,v}(h))$	n/a	n/a	1371.55
cAIC	1299.63	1289.66	1299.63
Scaled Deviance	415.05	403.08	414.35
df	156.00	155.00	155.60

Table 4.1: Deviances of Models. For these models, heat disorder count sum of each month and year is used as the response variable.

along the fitted values. Figure 4.1 demonstrates that the explanatory variables are quite useful to predict the response variable because the trend lines appear almost flat without showing no tendencies.

The next residual plot is to review Model 4.2 (Figure 4.2). Recalling the deviance of mean parameter estimation from Table 4.1, the cAIC (Conditional AIC with a random effect) reduces from 1299.63 to 1289.66. So, we could conclude adding the population variable has the model better. However, we have found the trend lines of residual plot changed. They changed as to show a slight upward and downward trend. Adding the population variable provoked an unexpected fallout. The population variable affected a lot the estimation of other variable, especially the province. For example, the sign of estimate of ProvinceGangwon changed.

Table4.2, Table4.3 and Table4.4 lead us to the conclusion that the level

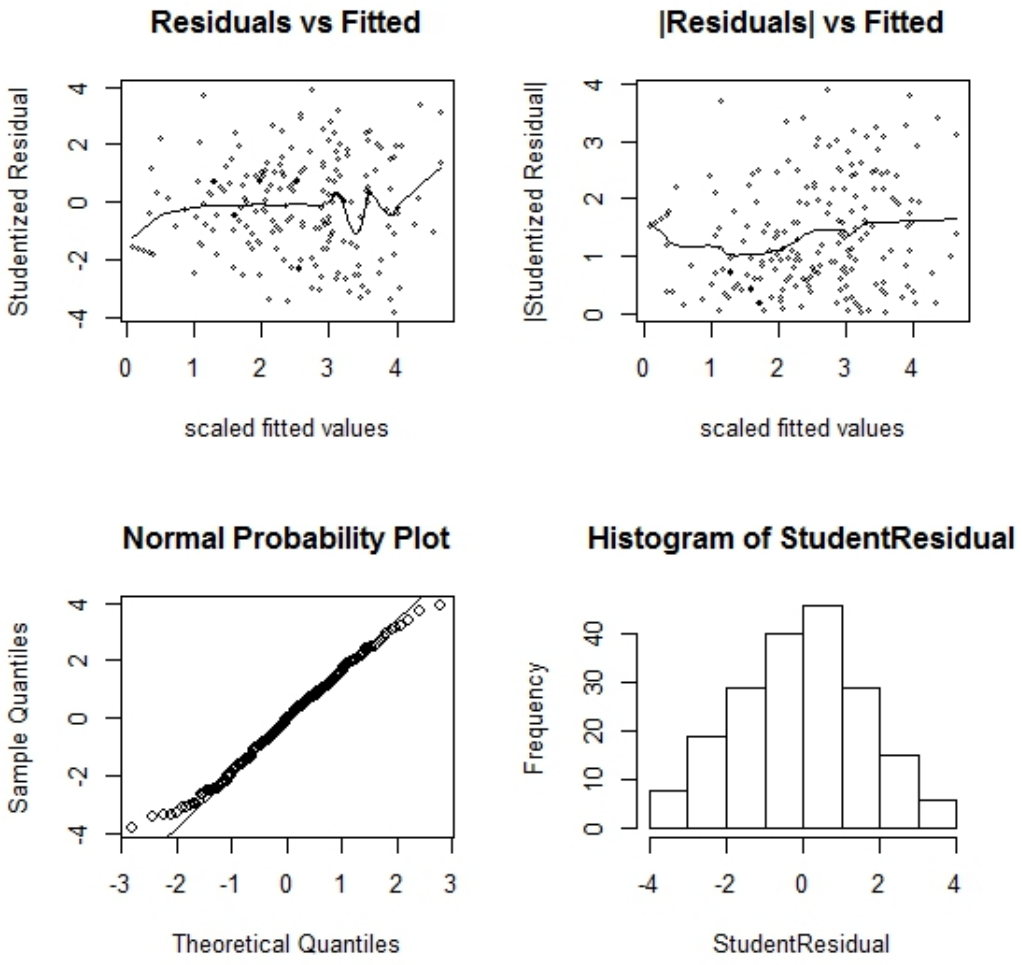


Figure 4.1: Residual Plot for Model 4.1. Heat disorder incidence count sum of each month and year was applied as the response variable. And, this plot is drawn by R (version 3.3.2) and dhglm package (Noh and Lee, 2015).

Variable	Model 4.1		
	Estimate	SE	t-value
Intercept	-12.6972	1.9182	-6.6193
ProvinceChungbuk	4.3600	2.5274	1.7251
ProvinceChungnam	0.8622	2.2326	0.3862
ProvinceDaegu	7.5190	1.9336	3.8887
ProvinceDaejeon	5.7785	3.0458	1.8972
ProvinceGangwon	-0.5701	2.0820	-0.2738
ProvinceGwangju	1.5506	2.0772	0.7465
ProvinceGyeongbuk	0.9956	1.6540	0.6020
ProvinceGyeonggi	-0.0810	2.2365	-0.0362
ProvinceGyeongnam	3.4583	1.5938	2.1698
ProvinceIncheon	-1.5102	2.8032	-0.5387
ProvinceJeju	0.2801	1.7949	0.1561
ProvinceJeonbuk	2.8009	1.9491	1.4370
ProvinceJeonnam	3.5631	1.5022	2.3719
ProvinceSeoul	-6.3928	2.6905	-2.3761
ProvinceUlsan	2.8947	1.5904	1.8200
as.factor(month)7	0.2625	0.1184	2.2166
as.factor(month)8	0.0706	0.1320	0.5350
tempH_mean	0.4336	0.0527	8.2266
humidityA_mean	0.0153	0.0081	1.8991
tempA_sd	0.8021	0.0487	16.4599

ProvinceChungbuk:tempH_mean	-0.1584	0.0847	-1.8697
ProvinceChungnam:tempH_mean	-0.0345	0.0753	-0.4583
ProvinceDaegu:tempH_mean	-0.2881	0.0635	-4.5394
ProvinceDaejeon:tempH_mean	-0.2319	0.1014	-2.2860
ProvinceGangwon:tempH_mean	-0.0027	0.0712	-0.0375
ProvinceGwangju:tempH_mean	-0.0778	0.0684	-1.1376
ProvinceGyeongbuk:tempH_mean	-0.0337	0.0558	-0.603
ProvinceGyeonggi:tempH_mean	0.0123	0.0744	0.1649
ProvinceGyeongnam:tempH_mean	-0.0978	0.0539	-1.8126
ProvinceIncheon:tempH_mean	0.0408	0.0973	0.4197
ProvinceJeju:tempH_mean	-0.0124	0.0608	-0.204
ProvinceJeonbuk:tempH_mean	-0.1047	0.0653	-1.6048
ProvinceJeonnam:tempH_mean	-0.0858	0.0515	-1.6667
ProvinceSeoul:tempH_mean	0.2127	0.0892	2.3836
ProvinceUlsan:tempH_mean	-0.1240	0.053	-2.3376

Table 4.2: Result of Model 4.1. In general, interaction terms help decrease the deviances of model. But, it is necessary to test the significance of interaction terms. After testing several interaction terms, an interaction term of province and mean of highest temperature was only selected to be included in a model.

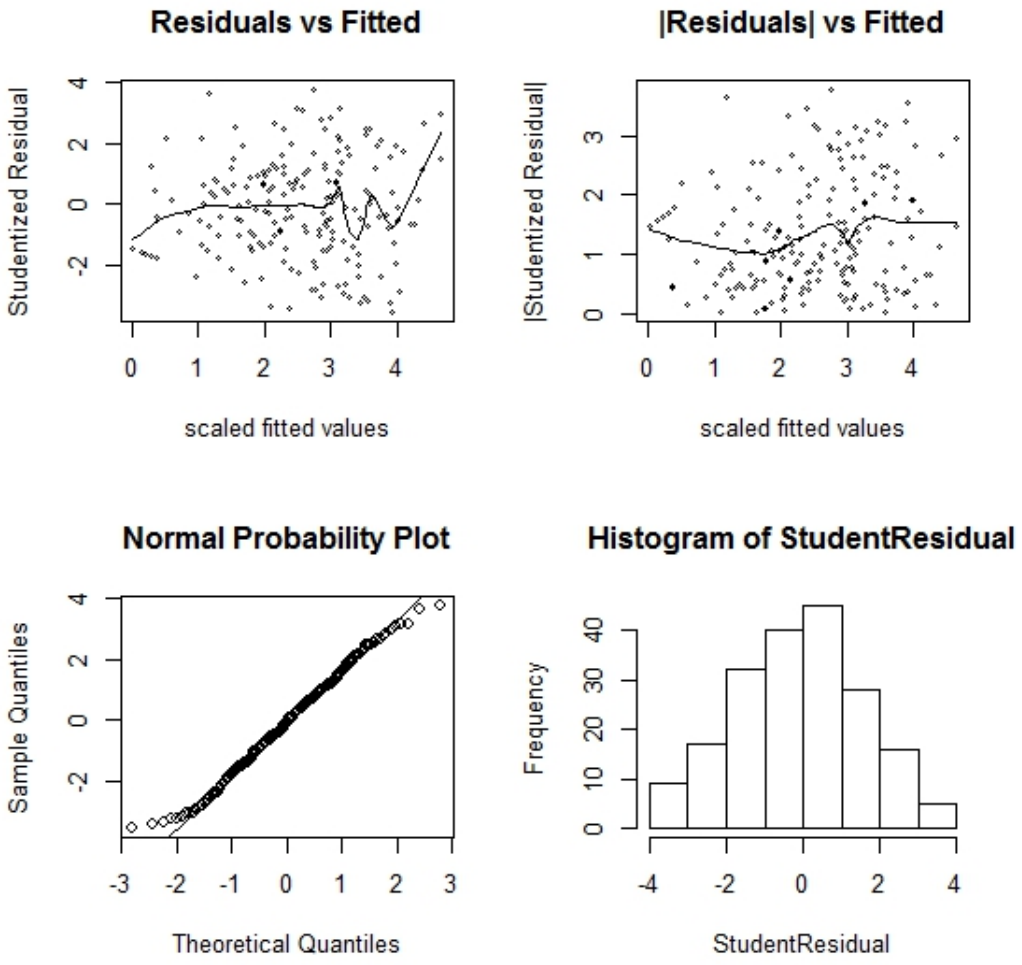


Figure 4.2: Residual Plot for Mode l4.2. Heat disorder incidence count sum of each month and year was applied as the response variable. And, this plot is drawn by R (version 3.3.2) and dhglm package (Noh and Lee, 2015).

Variable	Model 4.2		
	Estimate	SE	t-value
Intercept	-17.4800	2.3770	-7.3548
ProvinceChungbuk	6.8520	2.6290	2.6060
ProvinceChungnam	2.5390	2.2850	1.1111
ProvinceDaegu	8.9850	1.9860	4.5238
ProvinceDaejeon	8.1700	3.1220	2.6166
ProvinceGangwon	1.6430	2.1810	0.7532
ProvinceGwangju	3.8570	2.1900	1.7615
ProvinceGyeongbuk	2.1480	1.6930	1.2684
ProvinceGyeonggi	-8.8980	3.3810	-2.6320
ProvinceGyeongnam	4.0050	1.6090	2.4899
ProvinceIncheon	-0.7345	2.8130	-0.2611
ProvinceJeju	3.2600	1.9970	1.6322
ProvinceJeonbuk	4.8470	2.0420	2.3738
ProvinceJeonnam	5.4880	1.6090	3.4105
ProvinceSeoul	-12.6300	3.2210	-3.9199
ProvinceUlsan	5.3530	1.7470	3.0643
as.factor(month)7	0.1638	0.1217	1.3465
as.factor(month)8	-0.0387	0.1357	-0.2851
tempH_mean	0.4610	0.0536	8.5984
humidityA_mean	0.0226	0.0083	2.7058
tempA_sd	0.8269	0.0496	16.6867

population	0.0000	0.0000	3.4695
ProvinceChungbuk:tempH_mean	-0.1780	0.0849	-2.0958
ProvinceChungnam:tempH_mean	-0.0438	0.0754	-0.5813
ProvinceDaegu:tempH_mean	-0.3036	0.0638	-4.7560
ProvinceDaejeon:tempH_mean	-0.2479	0.1015	-2.4421
ProvinceGangwon:tempH_mean	-0.0125	0.0714	-0.1749
ProvinceGwangju:tempH_mean	-0.0884	0.0687	-1.2861
ProvinceGyeongbuk:tempH_mean	-0.0461	0.0562	-0.8215
ProvinceGyeonggi:tempH_mean	0.0167	0.0744	0.2241
ProvinceGyeongnam:tempH_mean	-0.1112	0.0543	-2.0482
ProvinceIncheon:tempH_mean	0.0340	0.0973	0.3495
ProvinceJeju:tempH_mean	-0.0174	0.0610	-0.2846
ProvinceJeonbuk:tempH_mean	-0.1203	0.0656	-1.8340
ProvinceJeonnam:tempH_mean	-0.0989	0.0518	-1.9080
ProvinceSeoul:tempH_mean	0.2036	0.0888	2.2939

Table 4.3: Result of Model 4.2. In general, interaction terms help decrease the deviances of model. But, it is necessary to test the significance of interaction terms. After testing several interaction terms, an interaction term of province and mean of highest temperature was only selected to be included in a model.

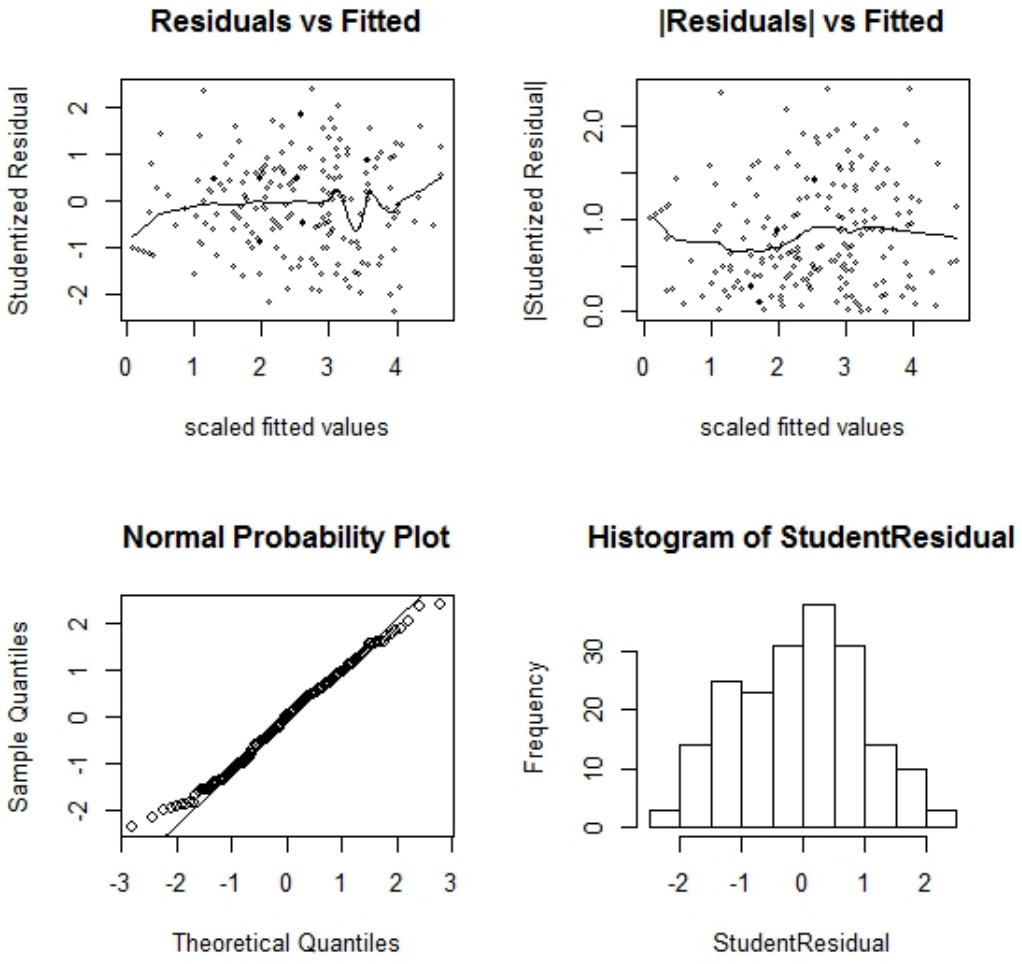


Figure 4.3: Residual Plot for Model 4.3. Heat disorder incidence count sum of each month and year was applied as the response variable. And, this plot is drawn by R (version 3.3.2), dhglm package (Noh and Lee, 2016) and unpublished R code.

Variable	Model 4.3		
	Estimate	SE	t-value
Intercept	-12.6972	1.9887	-6.3845
ProvinceChungbuk	4.3600	2.6448	1.6485
ProvinceChungnam	0.8622	2.3572	0.3658
ProvinceDaegu	7.5190	2.0623	3.6460
ProvinceDaejeon	5.7785	3.1338	1.8440
ProvinceGangwon	-0.5701	2.2108	-0.2579
ProvinceGwangju	1.5506	2.2005	0.7047
ProvinceGyeongbuk	0.9956	1.8222	0.5464
ProvinceGyeonggi	-0.0810	2.3588	-0.0343
ProvinceGyeongnam	3.4583	1.7356	1.9926
ProvinceIncheon	-1.5102	2.8988	-0.5210
ProvinceJeju	0.2801	1.9358	0.1447
ProvinceJeonbuk	2.8009	2.0887	1.3410
ProvinceJeonnam	3.5631	1.6678	2.1364
ProvinceSeoul	-6.3928	2.7899	-2.2914
ProvinceUlsan	2.8947	1.7215	1.6815
as.factor(month)7	0.2625	0.1184	2.2166
as.factor(month)8	0.0706	0.1320	0.5350
tempH_mean	0.4336	0.0527	8.2266
humidityA_mean	0.0153	0.0081	1.8991
tempA_sd	0.8021	0.0487	16.4599
ProvinceChungbuk:tempH_mean	-0.1584	0.0847	-1.8697
ProvinceChungnam:tempH_mean	-0.0345	0.0753	-0.4583

ProvinceDaegu:tempH_mean	-0.2881	0.0635	-4.5394
ProvinceDaejeon:tempH_mean	-0.2319	0.1014	-2.2860
ProvinceGangwon:tempH_mean	-0.0027	0.0712	-0.0375
ProvinceGwangju:tempH_mean	-0.0778	0.0684	-1.1376
ProvinceGyeongbuk:tempH_mean	-0.0337	0.0558	-0.6030
ProvinceGyeonggi:tempH_mean	0.0123	0.0744	0.1649
ProvinceGyeongnam:tempH_mean	-0.0978	0.0539	-1.8126
ProvinceIncheon:tempH_mean	0.0408	0.0973	0.4197
ProvinceJeju:tempH_mean	-0.0124	0.0608	-0.2040
ProvinceJeonbuk:tempH_mean	-0.1047	0.0653	-1.6048
ProvinceJeonnam:tempH_mean	-0.0858	0.0515	-1.6667
ProvinceSeoul:tempH_mean	0.2127	0.0892	2.3836
ProvinceUlsan:tempH_mean	-0.1240	0.053	-2.3376
λ	0.2500	0.0000	
ρ	0.1740	1.2960	

Table 4.4: Result of Model 4.3. In general, interaction terms help decrease the deviances of model. But, it is necessary to test the significance of interaction terms. After testing several interactions terms, an interaction term of province and mean of highest temperature was only selected to be included in a model.

of highest temperature may be the main factor to result in the incidence of heat disorder. Even though we applied an interaction between province and highest temperature, highest temperature itself still shows a high significance (t-value = 8.2266, 8.5984, 8.2266) And, notably, the standard deviation of daily average temperature of each month and year shows a high significance (t-value = 16.4599, 16.6867, 16.4599).

From this result, we can suggest that the level of temperature and its variability can provoke the incidence of heat disorder. If temperature varies more, its probability of reaching too high may escalate.

Our discussion has started from the Figure 1.1 of IPCC which implies a possibility that climate change results in some change in probability distribution of weather phenomenon. If climate change moves us into the world of bigger variability, we may confront the more incidences of heat disorder patients. Our modeling results provide us an empirical proof on this mechanism of incidence of heat disorder.

Comparing Model 4.1, Model 4.2 and Model 4.3, we confronted a problem of model selection. In this case, how to select a proper model might depend on the purpose of analysis. If we start with the purpose of stable prediction of response variable, a method to predict giving unstable residuals as Model 4.2 would not be helpful. Reminding us of the finding of Figure3.8, heat disorder incidence does not appear according to the order of population. We had to reconsider the use of population variable.

So, we tested the same equations to another response variable, heat disorder incidence rate per 1,000,000 people. A disease incidence rate is often chosen as the response variable to verify its relationship with the explanatory

variables.

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidityA_{ijk} + tempA_{ijk} + \\ + month_j + province_k \quad (4.4)$$

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidity_{ijk} + tempSD_{ijk} + \\ + month_j + province_k + v_k \quad (4.5)$$

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidity_{ijk} + tempA_{ijk} + \\ + month_j + province_k + v_k \quad (4.6)$$

Equation(Model) 4.4 is a Poisson GLM not assuming any random effect.

Equation(Model) 4.5 is a Poisson HGLM assuming a random effect which follows $N(0, \lambda)$.

Equation(Model) 4.6 is a Poisson HGLM assuming an area-specific random effect following Markov Random Field (MRF) in which $[var(v)]^{-1} = (I - \rho M)/\lambda$ where M is the incidence matrix for neighbours of areas.

We used the mean of daily highest temperature of each month and year, the mean of daily relative humidity of each month and year, and the standard deviation of daily average temperature of each month and year, month as a factor variable and province as a nominal variable to predict the heat disorder incidence.

Then, we obtained the results of Model 4.4, Model 4.5 and Model 4.6 in which heat disorder incidence rate per 1,000,000 was used as the response variable. We can observe that we can reduce the deviances of our models when we test heat disorder incidence rate per 1,000,000 as the response variable. And, we could detect a similar trend lines in residual plot although we removed the interaction term we added in Model 4.1, Model 4.2, and Model 4.3. Any interaction term did not appear significant when we model.

Deviance	Model 4.4	Model 4.5	Model 4.6
-2ML$(-2h)$	917.46	917.46	925.85
-2RL$(-2p_\beta(h))$	983.95	n/a	n/a
-2RL$(-2p_v(h))$	n/a	n/a	969.08
-2RL$(-2p_{\beta,v}(h))$	n/a	983.95	983.95
cAIC	959.46	959.46	959.46
Scaled deviance	264.97	264.97	264.27
df	171.00	171.00	170.60

Table 4.5: Deviances of Models. For these models, heat disorder incidence rate per 1,000,000 of each month and year is used as the response variable.

Changing the response variable from heat disorder incidence count sum to incidence rate per 1,000,000, we can find that the estimates and their significances are changed. Moreover, if we add a random effect term to a model, the significances for the province estimates are reduced (From Model 4.4 to Model 4.6).

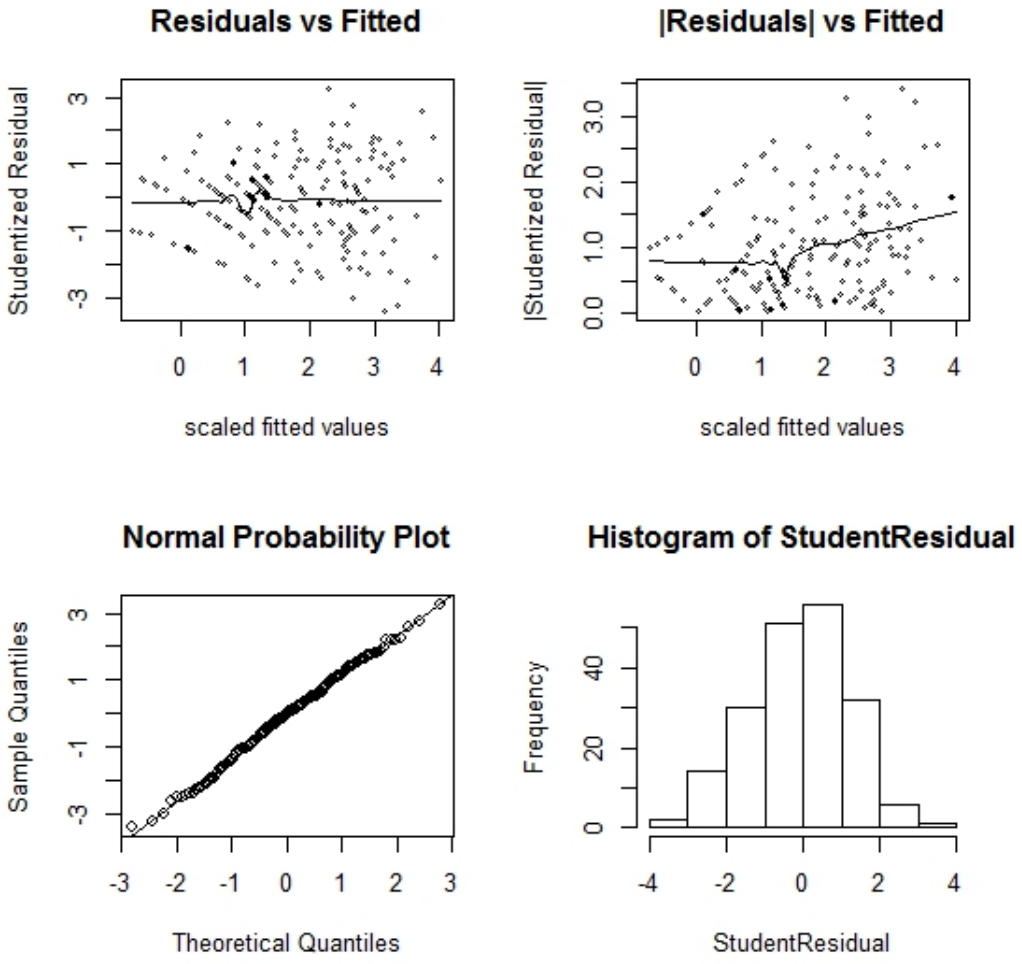


Figure 4.4: Residual Plot for Model 4.4. Heat disorder incidence rate per 1,000,000 of each month and year was applied as the response variable. And, his plot is drawn by R (version 3.3.2) and dhglm package (Noh and Lee, 2015).

Variable	Model 4.4		
	Estimate	SE	t-value
Intercept	-9.5906	1.3320	-7.2003
ProvinceChungbuk	0.4948	0.1791	2.7622
ProvinceChungnam	0.4792	0.1949	2.4586
ProvinceDaegu	-1.0222	0.2155	-4.7445
ProvinceDaejeon	-0.2163	0.2117	-1.0221
ProvinceGangwon	0.2523	0.1939	1.3011
ProvinceGwangju	0.1929	0.1867	1.0332
ProvinceGyeongbuk	0.3652	0.1846	1.9778
ProvinceGyeonggi	-0.8623	0.2379	-3.6249
ProvinceGyeongnam	0.6751	0.1796	3.7594
ProvinceIncheon	-0.1265	0.2462	-0.5138
ProvinceJeju	1.7709	0.1808	9.7975
ProvinceJeonbuk	0.4068	0.1883	2.1607
ProvinceJeonnam	1.7362	0.1827	9.5023
ProvinceSeoul	-1.0343	0.2551	-4.0549
ProvinceUlsan	0.3588	0.1841	1.9485
as.factor(month)7	0.4992	0.1578	3.1624
as.factor(month)8	0.3217	0.1751	1.8372
tempH_mean	0.3208	0.0270	11.8799
humidityA_mean	0.0011	0.0099	0.1149
tempA_sd	0.7067	0.0724	9.7570

Table 4.6: Result of Model 4.4. Mean of daily highest temperature of each month and year, mean of daily average relative humidity of each month and year and standard deviation of daily average temperature of each month and year are confirmed as to explain a lot about heat disorder incidence rate per 1,000,000.

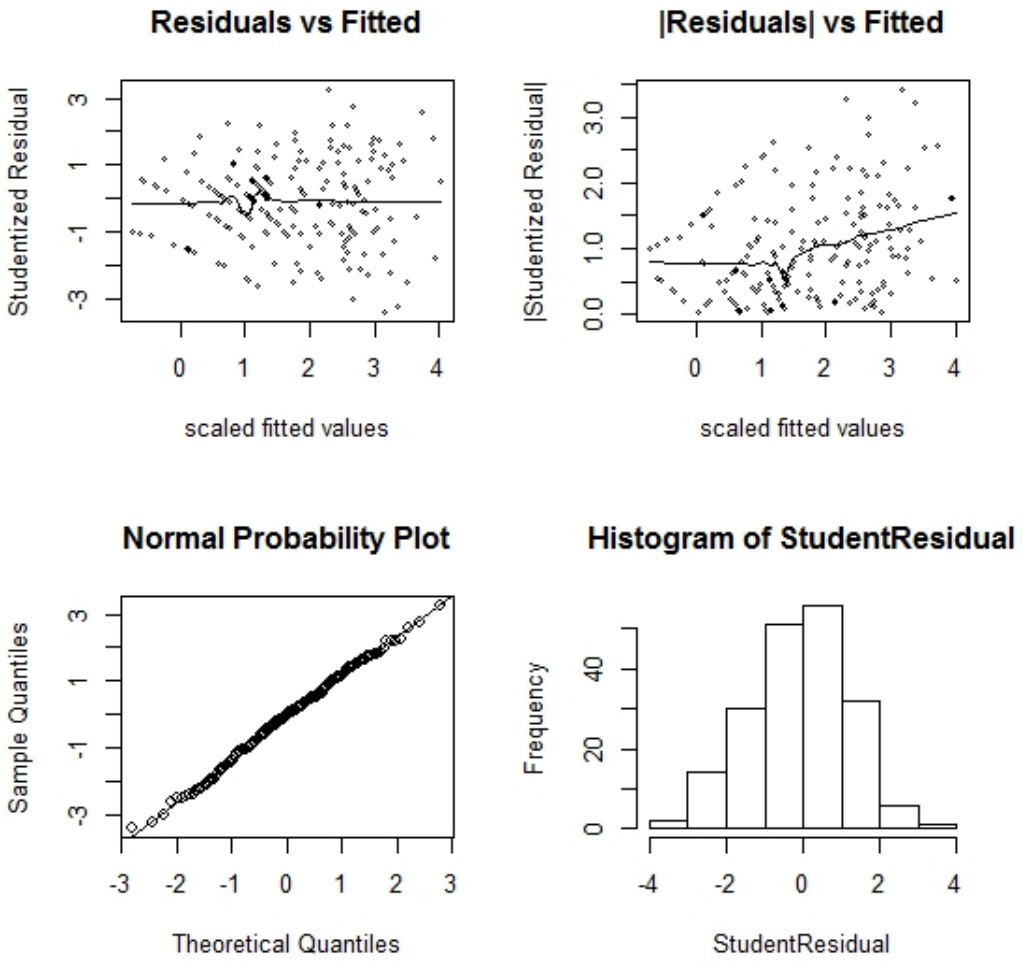


Figure 4.5: Residual Plot for Model 4.5. Heat disorder incidence rate per 1,000,000 of each month and year was applied as the response variable. And, this plot is drawn by R (version 3.3.2) dhglm package (Noh and Lee, 2015).

Variable	Model 4.5		
	Estimate	SE	t-value
Intercept	-9.5906	1.332	-7.2003
ProvinceChungbuk	0.4948	0.1791	2.7622
ProvinceChungnam	0.4792	0.1949	2.4586
ProvinceDaegu	-1.0222	0.2155	-4.7445
ProvinceDaejeon	-0.2163	0.2117	-1.0221
ProvinceGangwon	0.2523	0.1939	1.3011
ProvinceGwangju	0.1929	0.1867	1.0332
ProvinceGyeongbuk	0.3652	0.1846	1.9778
ProvinceGyeonggi	-0.8623	0.2379	-3.6249
ProvinceGyeongnam	0.6751	0.1796	3.7594
ProvinceIncheon	-0.1265	0.2462	-0.5138
ProvinceJeju	1.7709	0.1808	9.7975
ProvinceJeonbuk	0.4068	0.1883	2.1607
ProvinceJeonnam	1.7362	0.1827	9.5023
ProvinceSeoul	-1.0343	0.2551	-4.0549
ProvinceUlsan	0.3588	0.1841	1.9485
as.factor(month)7	0.4992	0.1578	3.1624
as.factor(month)8	0.3217	0.1751	1.8372
tempH_mean	0.3208	0.027	11.8799
humidityA_mean	0.0011	0.0099	0.1149
tempA_sd	0.7067	0.0724	9.7570
λ	-95.2400	4.2040	-22.6500

Table 4.7: Result of Model 4.5. Mean of daily highest temperature, mean of daily average relative humidity and year and standard deviation of daily average temperature of each month and year are confirmed as to explain well about heat disorder incidence rate per 1,000,000.

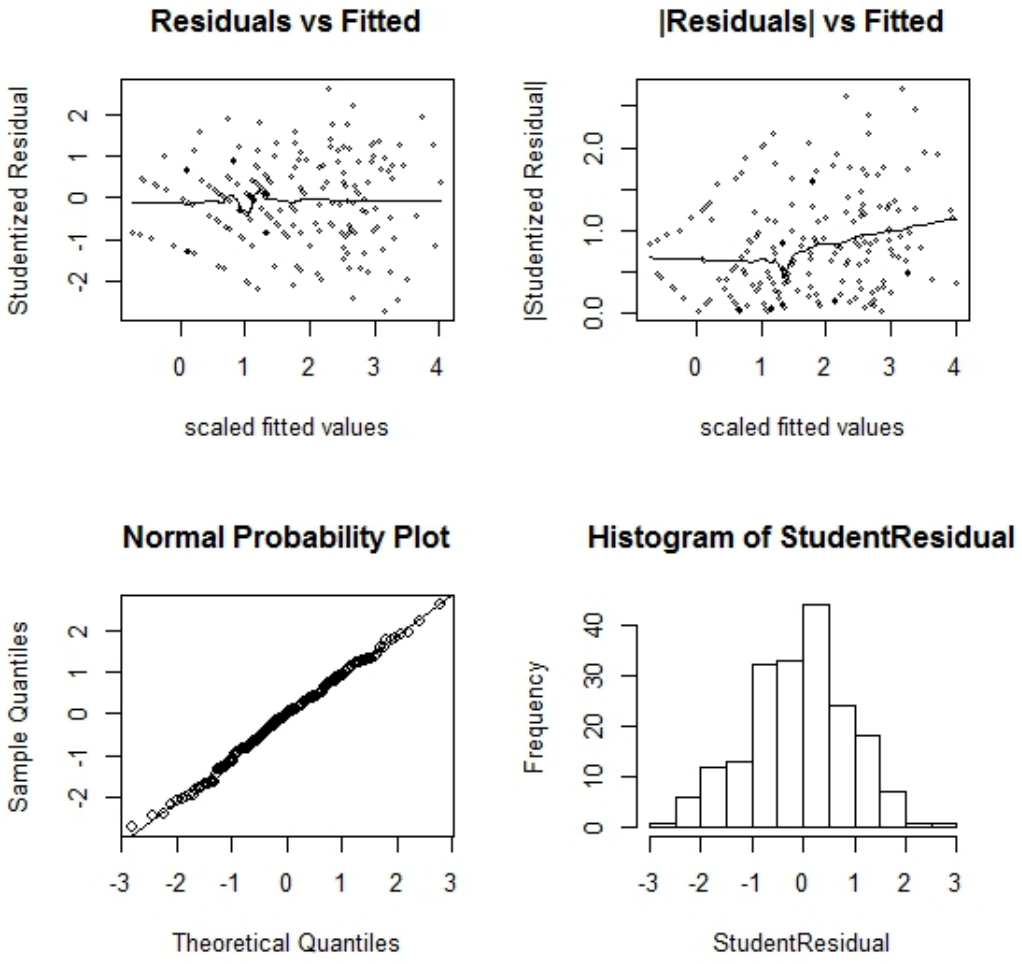


Figure 4.6: Residual Plot for Model 4.6. Heat disorder incidence rate per 1,000,000 of each month and year was applied as the response variable. And, this plot is drawn by R (version 3.3.2), dhglm package (Noh and Lee, 2015) and unpublished R code.

Variable	Model 4.6		
	Estimate	SE	t-value
Intercept	-9.5906	1.4317	-6.6988
ProvinceChungbuk	0.4948	0.7995	0.6189
ProvinceChungnam	0.4792	0.7810	0.6136
ProvinceDaegu	-1.0222	0.7488	-1.3651
ProvinceDaejeon	-0.2163	0.7672	-0.2820
ProvinceGangwon	0.2523	0.7684	0.3283
ProvinceGwangju	0.1929	0.7500	0.2572
ProvinceGyeongbuk	0.3652	0.7722	0.4729
ProvinceGyeonggi	-0.8623	0.8007	-1.0770
ProvinceGyeongnam	0.6751	0.7100	0.9508
ProvinceIncheon	-0.1265	0.7780	-0.1626
ProvinceJeju	1.7709	0.7472	2.3702
ProvinceJeonbuk	0.4068	0.7740	0.5256
ProvinceJeonnam	1.7362	0.7472	2.3236
ProvinceSeoul	-1.0343	0.7808	-1.3246
ProvinceUlsan	0.3588	0.6840	0.5246
as.factor(month)7	0.4992	0.1578	3.1624
as.factor(month)8	0.3217	0.1751	1.8372
tempH_mean	0.3208	0.0270	11.8799
humidityA_mean	0.0011	0.0099	0.1149
tempA_sd	0.7067	0.0724	9.7570
λ	0.2500	0.0000	
ρ	0.1740	1.2960	

Table 4.8: Result of Model 4.6. We got the same output of λ and ρ with that of Model4.3. However, notably, all of the significance of province are reduced.

Secondly, we excluded the interaction term in Model 4.4, Model 4.5 and Model 4.6. It is interesting to see that the significance level of mean of highest temperature arises a lot, meanwhile that of province estimates do not.

These modifications of modelling have an effect of elimination of explanatory power of province variable. Even though, the spatial correlation structure did not improve goodness of fits, but addition as a random effect into our model reduced the significance levels of province to the point that only two regions (Jeju and Jeonnam) showed significant estimates. It is remarkable because we can consider a way to rule out the province variable.

4.4. Conditioning Spatial Correlation

Inspired by the result of Model 4.6, we tested other models in which province variable was excluded. In fact, our study design did not cover the socioeconomic aspects of heat disorder incidence. So, we should bear in mind that the importances (usefulness to estimation) of our variables could be exaggerated due to the lack of information. Let us review these equations without considering the province variable.

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidityA_{ijk} + tempA_{ijk} + \\ + month_j \quad (4.7)$$

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidity_{ijk} + tempSD_{ijk} + \\ + month_j + v_k \quad (4.8)$$

$$\eta_{ijk} = log\frac{\mu_{ijk}}{population_{ik}} * 1000000 = tempH_{ijk} + humidity_{ijk} + tempA_{ijk} + \\ + month_j + v_k \quad (4.9)$$

Equation(Model) 4.7 is a Poisson GLM not assuming any random effect.

Equation(Model) 4.8 is a Poisson HGLM assuming a random effect which follows $N(0, \lambda)$.

Equation(Model) 4.9 is a Poisson HGLM assuming an area-specific random effect following Markov Random Field (MRF) in which $[var(v)]^{-1} = (I - \rho M)/\lambda$ where M is the incidence matrix for neighbours of areas.

We used the same variables which were taken into account for our models in Section4.3.

Deviance	Model 4.7	Model 4.8	Model 4.9
-2ML$(-2h)$	1612.17	997.25	954.49
-2RL$(-2p_\beta(h))$	1643.95	n/a	n/a
-2RL$(-2p_v(h))$	n/a	n/a	997.34
-2RL$(-2p_{\beta,v}(h))$	n/a	1022.84	1022.52
cAIC	1624.17	959.26	959.23
Scaled deviance	265.52	264.97	264.81
df	186.00	171.38	170.99

Table 4.9: Deviances of Models. For these models, heat disorder incidence rate per 1,000,000 of each month and year is used as the response variable.

In fact, province variable does not give any meaningful information to help understanding the risk of heat disorder incidence. The estimate of each province level means the relative effect of *being itself* comparing to the referred province level (it was Busan in our study). If we want to take proper action to prevent a disease, it would be much better to know the causes of regional differences not just the effects of being each province. Moreover, our explanatary variables (the temperature variables and humidity variable) are such regional features, too.

For a better and more practical interpretation of our linear models, we should select the independant variables which may be a part of the explaination of heat disorder incidence mechanism. In our situation, the addition of random effect condnitioning spatial correlation gives us one alternative way to build better models with lower cAICs without using this ambiguous province variable (cAICs from 1624.17 of Model 4.7 to 959.23 of Model 4.9). It is notable that we can replace the province variable of fixed effect with spatial correlation as random effect. Table4.10. shows that this uncertain province variable influences the goodness of fit and coefficients a lot.

Model	Variable	Estimate	SE	t-value
Model 4.7	Intercept	-15.8445	1.0755	-14.7316
	as.factor(month)7	0.0678	0.1337	0.5075
	as.factor(month)8	-0.0931	0.1493	-0.6238
	tempH_mean	0.3660	0.0229	15.9644
	humidityA_mean	0.0807	0.0066	12.2788
	tempA_sd	0.4675	0.0535	8.7308

Model	Variable	Estimate	SE	t-value
Model 4.8	Intercept	-9.7198	1.3710	-7.0893
	as.factor(month)7	0.4730	0.1570	3.0137
	as.factor(month)8	0.2963	0.1741	1.7014
	tempH_mean	0.3242	0.0269	12.0603
	humidityA_mean	0.0049	0.0097	0.5076
	tempA_sd	0.6962	0.0718	9.7016
	λ	-0.4689	0.4418	-1.0610
Model 4.9	Intercept	-9.6895	1.3785	-7.0290
	as.factor(month)7	0.4734	0.1569	3.0190
	as.factor(month)8	0.2971	0.1740	1.7070
	tempH_mean	0.3240	0.0269	12.0550
	humidityA_mean	0.0048	0.0097	0.4970
	tempA_sd	0.6972	0.0717	9.7200
	λ	0.6162	0.0685	
	ρ	0.1071	0.1604	

Table 4.10: Reuslts of Model 4.7, Model 4.8 and Model 4.9. For these models, heat disorder incidence rate per 1,000,000 of each month and year is used as the response variable. The result of Model 4.7 is different from those of Model 4.8 and Model 4.9. And, this plot is drawn by R (version 3.3.2), dhglm package (Noh and Lee, 2015) and unpublished R code.

We can observe the estimates of humidity variable and their significance depend on the existence of province variable. Because of the deficiency of

our information, we could not ignore the province variable for this study. But, it is expected to be more improved if we add socioeconomic features by adopting random effect of spatial correlation structure.

For the same reason, we chose the result of Model 4.6 in order to have the simulation of next section.

4.5. Simulation

Recalling the form of linear predictor introduced in Section4.1, we rewrite the form due to the addition of correlation structure.

$$\eta_i = g(\mu_i) = X\beta + Z_i^* r_i, \qquad Z_i^* = Z_i L_i, \qquad r_i \sim MVN(0, \lambda_i)$$

Lee and Nelder (2006) propose a method in which a random effect follows MRF where $[var(v)]^{-1} = (I - \rho M)/\lambda$ (M is an incidence matrix for neighbors of areas) to implement a random effect with a correlation structure.

$$[var(v)]^{-1} = P(\rho)/\lambda = (I - \rho M)/\lambda, \qquad P(\rho) = (L(\rho)^t)^{-1} L(\rho)^{-1}$$

We can simulate heat disorder incidence with these equations using the parameter coefficients already obtained. If we apply Cholesky decomposition to $(I - \rho M)$, we can get $L(\rho)$.

One benefit of this simulation is we can *reconstruct* the response variable for the period of no observation. Here, we present two types of simulation. The first type is the simulated maximum heat disorder incidence in June,

July and August for 10 years from 1973 to 1982. The second type is the simulated maximum heat disorder incidence of June, July and August for 4 years from 2012 to 2015. And, we plotted the maximum value of observations for 4 years from 2012 to 2015. These plots allow us to compare the simulated features of past and those of present. They are depicted by R (version 3.3.2) and ggplot2 (Wickham, 2009).

For Jeju, the current actual population does not exceed 1 million people. So, the simulated numbers for Jeju are a little bit inflated (Current actual population of Province Jeju is about 625,000 as of 2015.).

According to our simulations, it is not obvious that heat disorder incidence have increased significantly in all of the provinces in June, July and August. But, the possibility of bigger incidence seems to increase in the most of provinces especially in August, the hottest month of year in South Korea because of its under-predicted values. The simulated incidences for Gwangju, Gyeongnam, Jeju, Jeonbuk, Jeonnam and Ulsan show an under-prediction. We noticed that these cities and provinces are mostly located in souther region of South Korea.

On the contrary, Busan, Daegu, Daejeon, Gyeongbuk, Incheon and Seoul show relatively accurate simulation results except for a few over-predictions on the incidence in August.

For Daegu, the risk of heat disorder incidence looks stable even though it is regarded as the hottest region in South Korea. And, Jeju and Jeonnam appear as the most vulnerable regions to heat disorder.

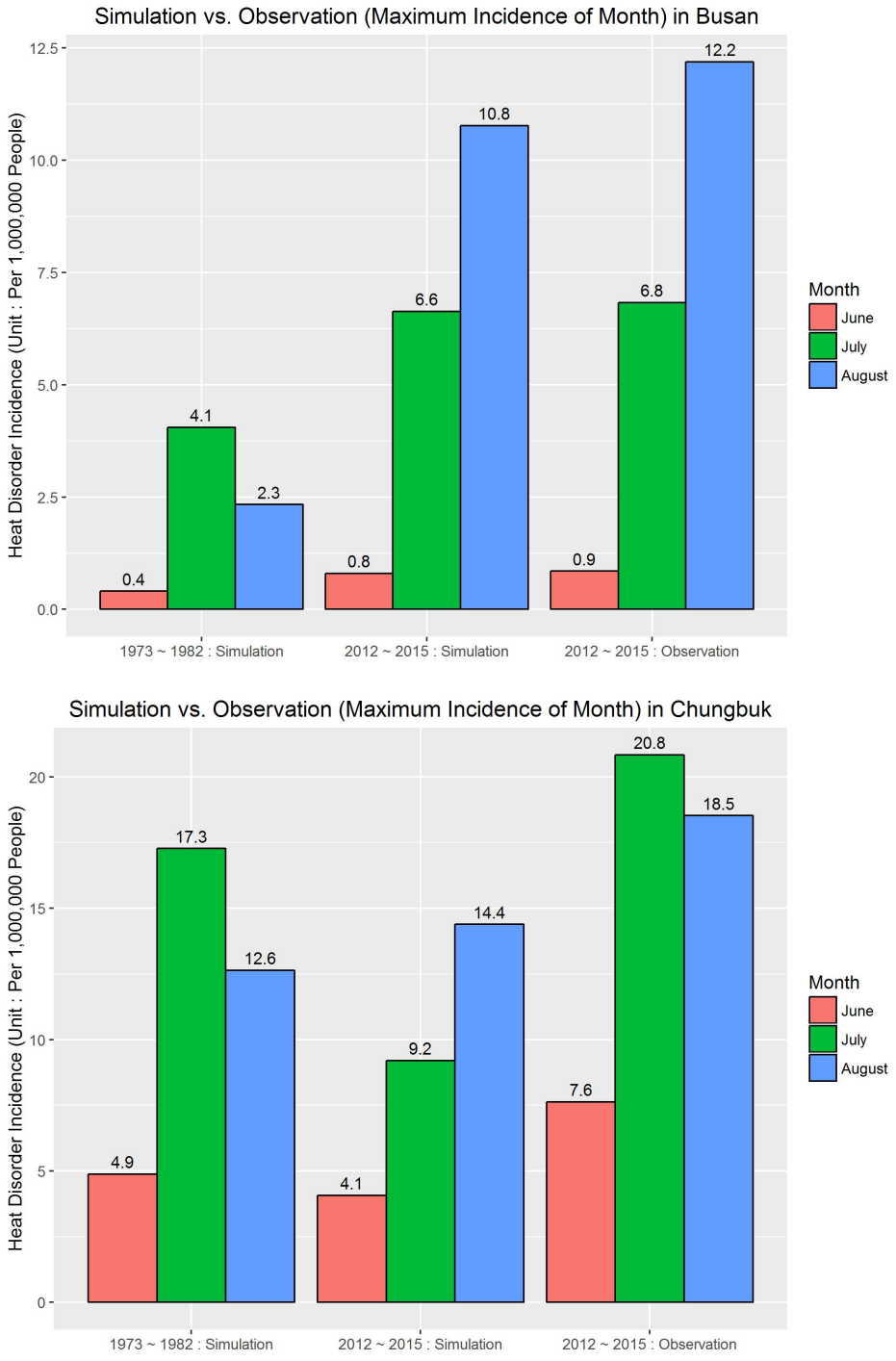
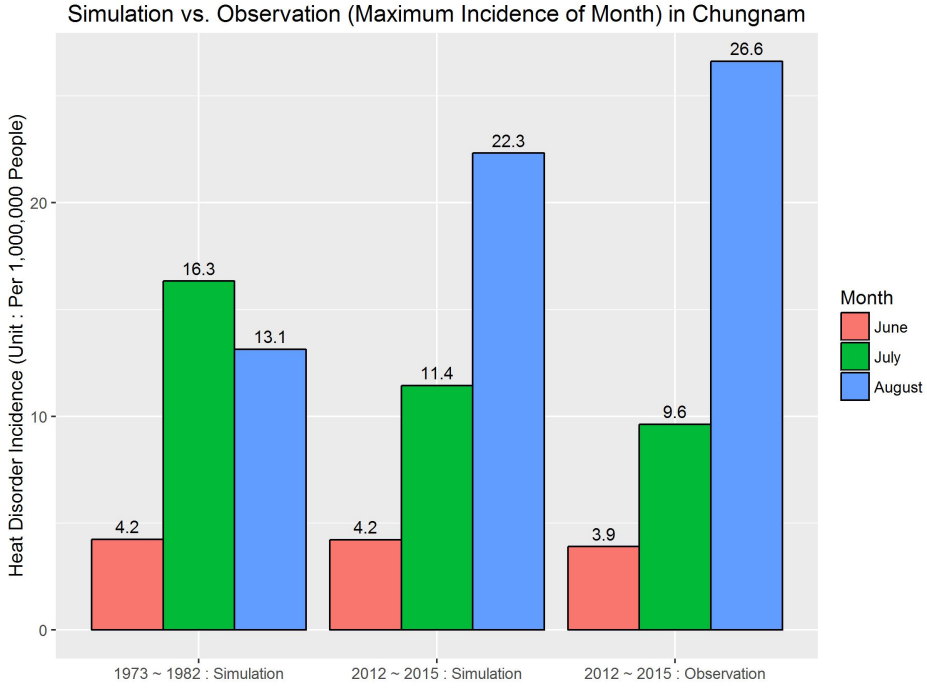
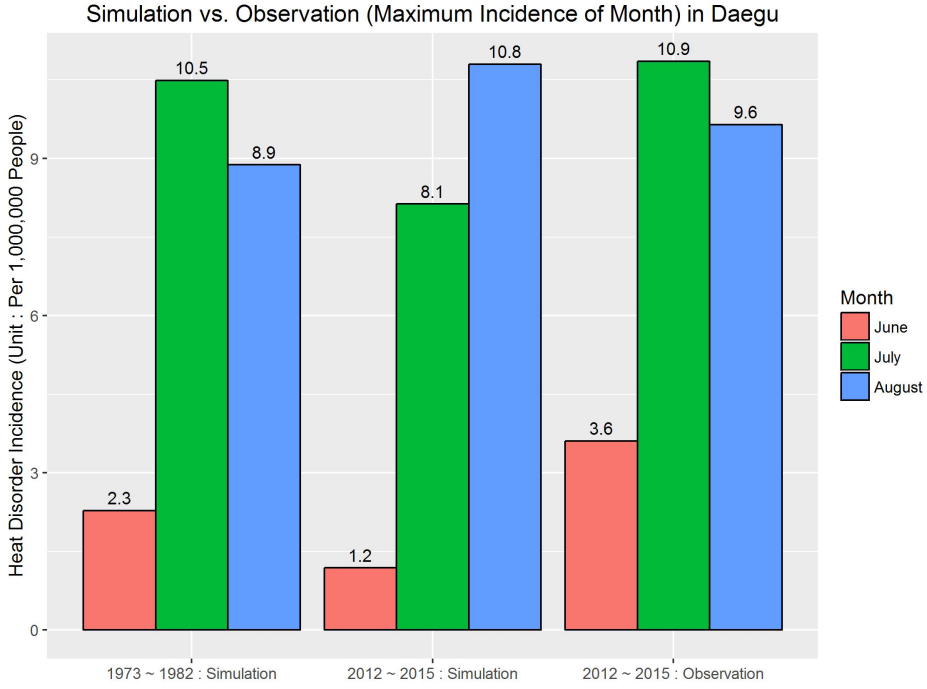
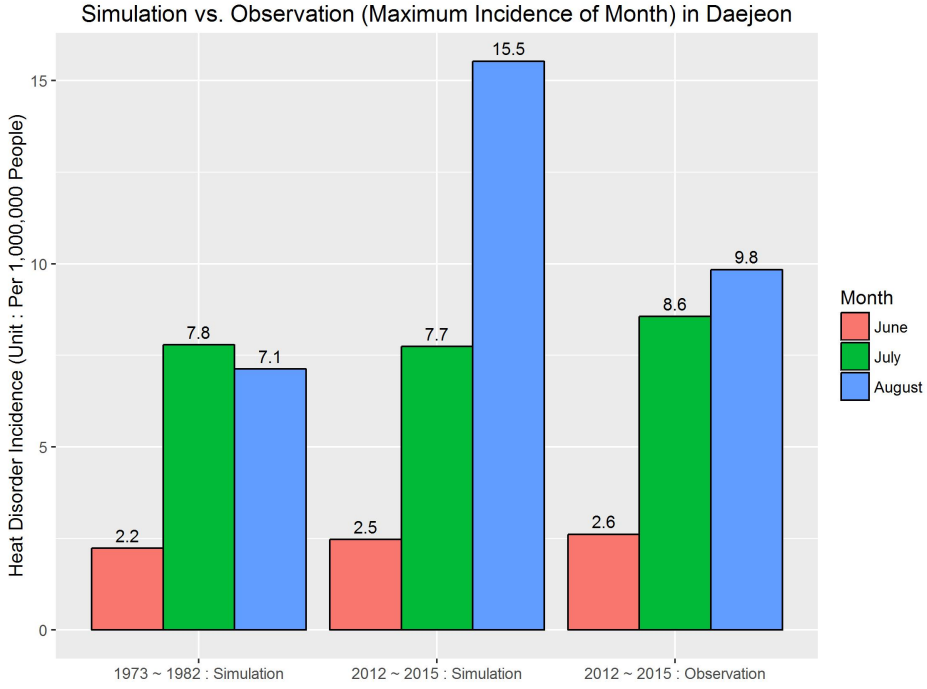
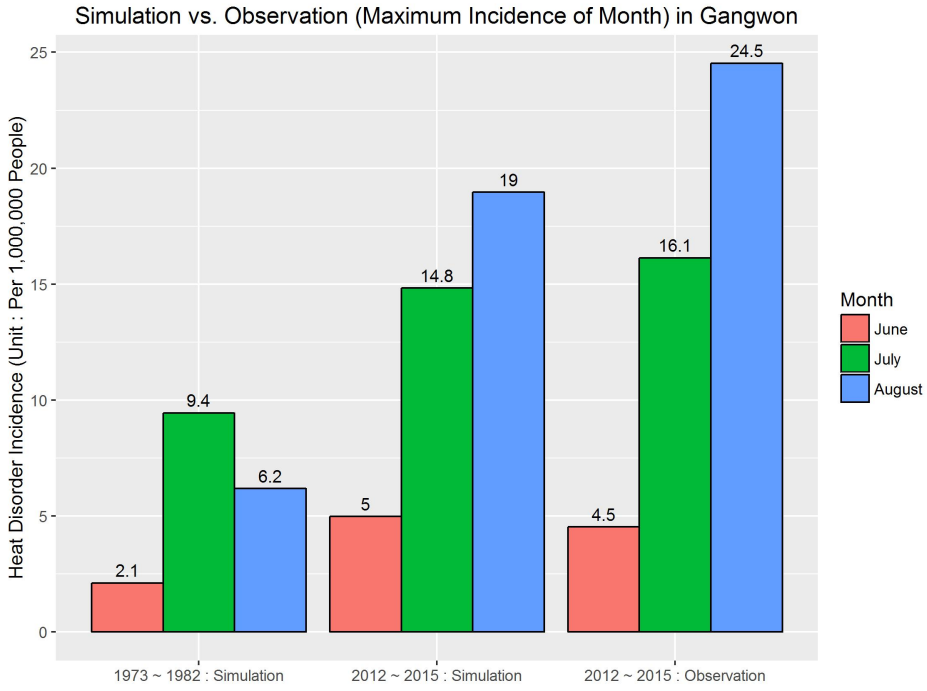


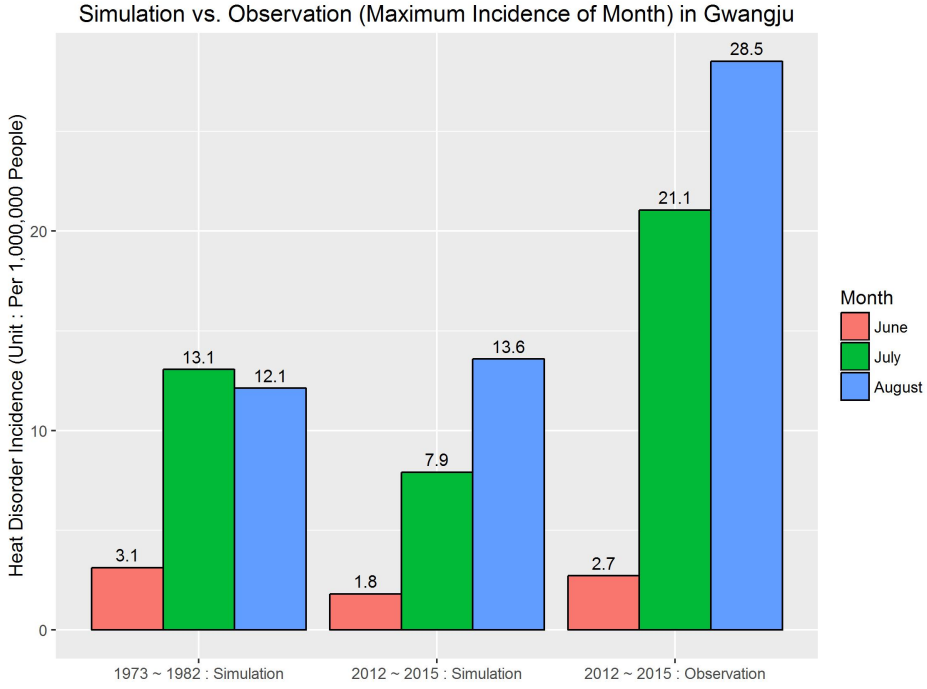
Figure 4.7: Simulation vs. Observation by Province. This plot is a summary of maximum incidence of heat disorder by month during the period.

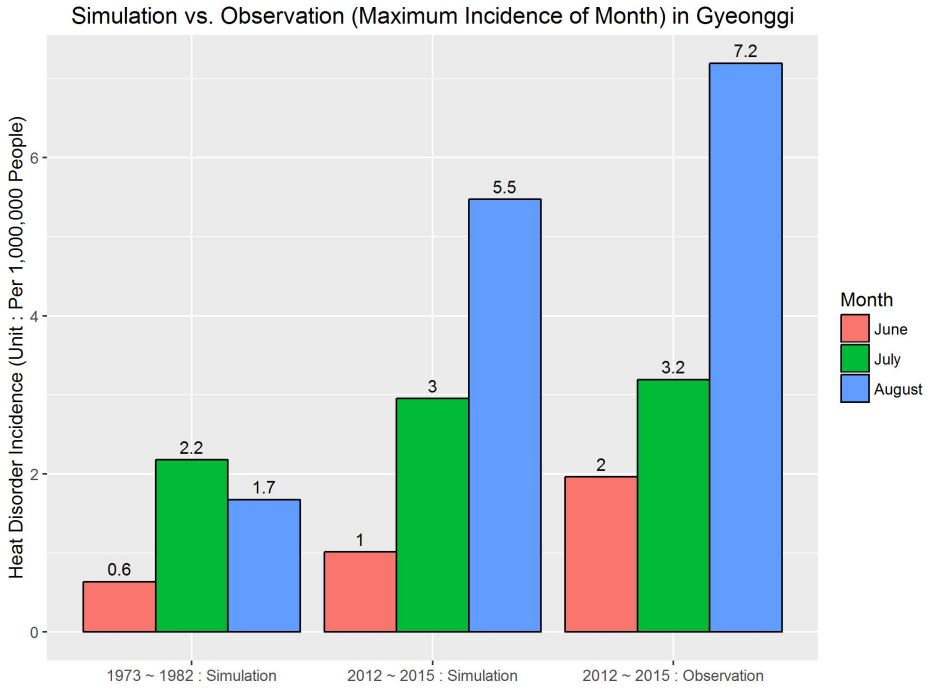

Simulation vs. Observation (Maximum Incidence of Month) in Chungnam
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
10
20
4.2
16.3
13.1
4.2
11.4
22.3
3.9
9.6
26.6
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

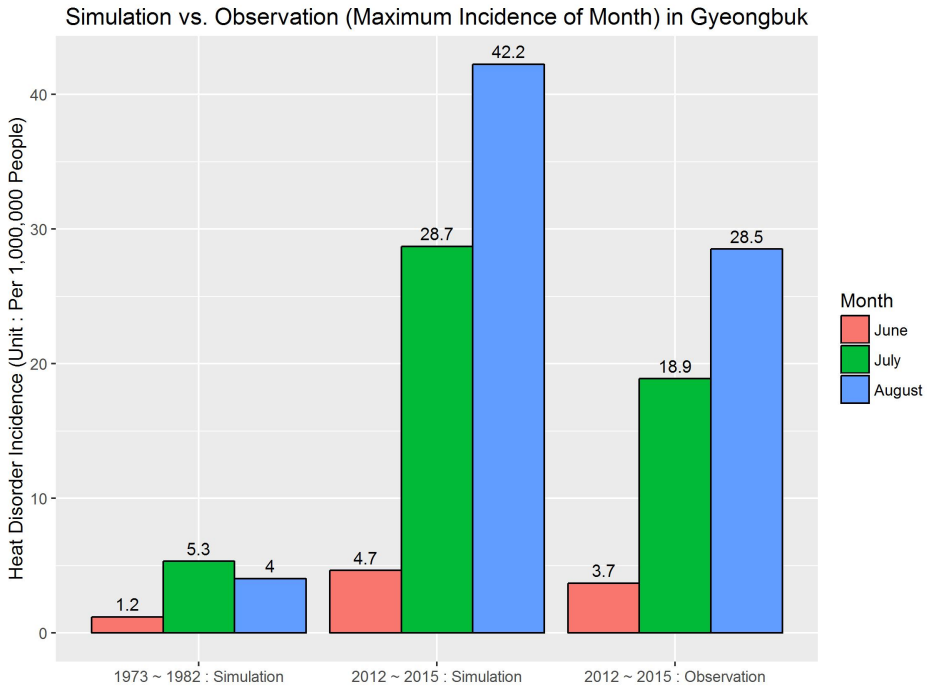

Simulation vs. Observation (Maximum Incidence of Month) in Daegu
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
3
6
9
2.3
10.5
8.9
1.2
8.1
10.8
3.6
10.9
9.6
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

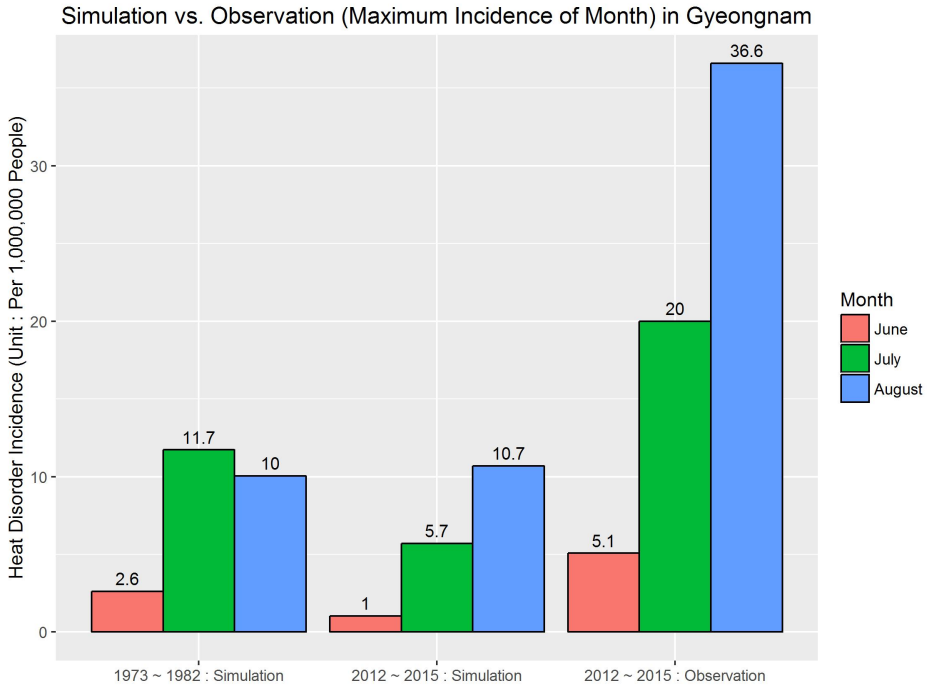

Simulation vs. Observation (Maximum Incidence of Month) in Daejeon
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
5
10
15
2.2
7.8
7.1
2.5
7.7
15.5
2.6
8.6
9.8
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

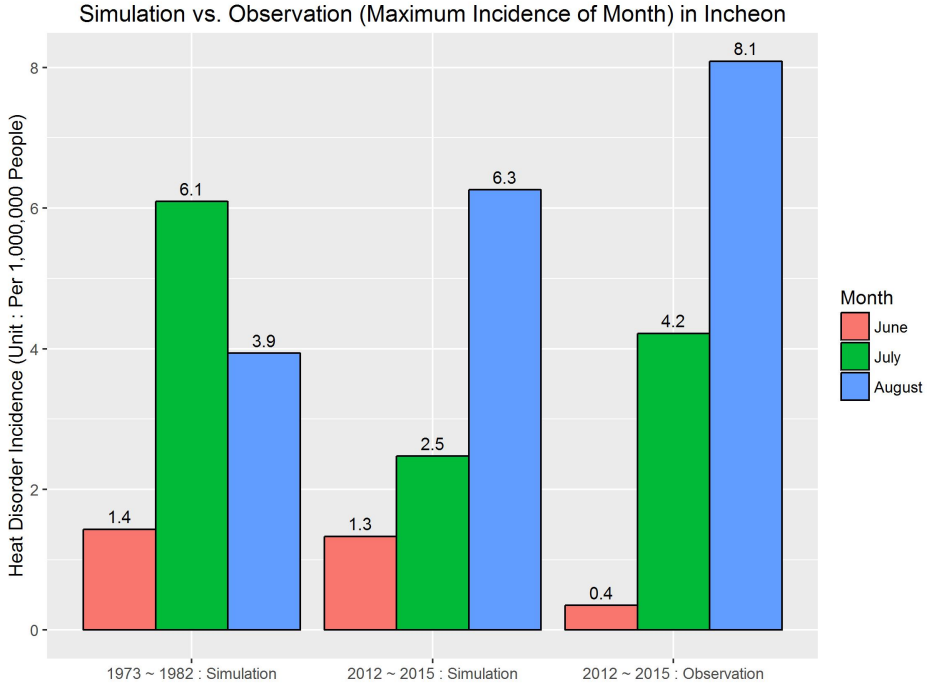

Simulation vs. Observation (Maximum Incidence of Month) in Gangwon
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
5
10
15
20
25
2.1
9.4
6.2
5
14.8
19
4.5
16.1
24.5
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

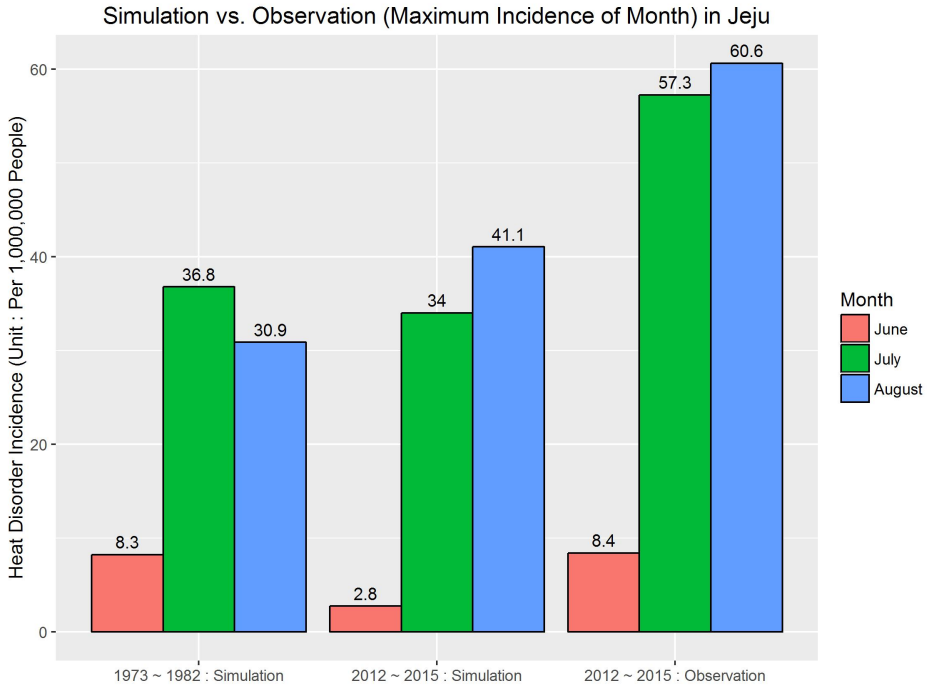

Simulation vs. Observation (Maximum Incidence of Month) in Gwangju
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
10
20
3.1
13.1
12.1
1.8
7.9
13.6
2.7
21.1
28.5
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

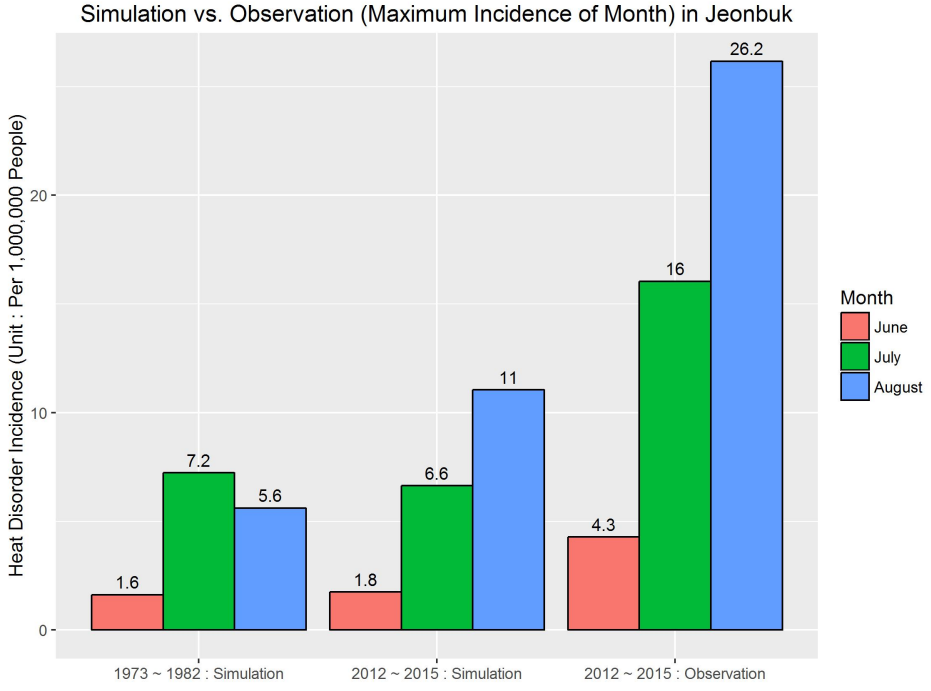

Simulation vs. Observation (Maximum Incidence of Month) in Gyeonggi
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
2
4
6
0.6
2.2
1.7
1
3
5.5
2
3.2
7.2
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

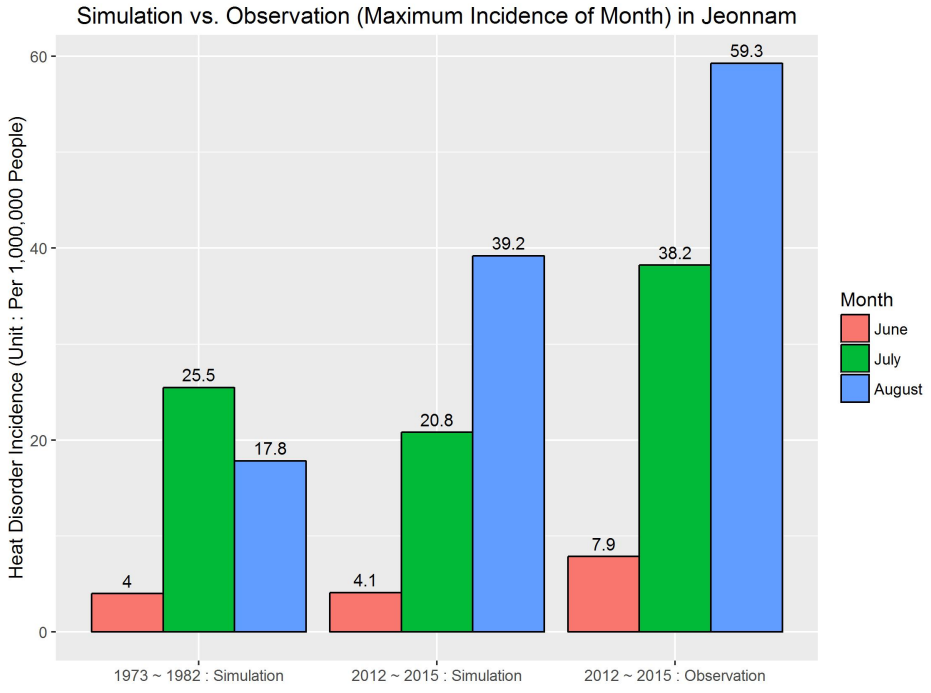

Simulation vs. Observation (Maximum Incidence of Month) in Gyeongbuk
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
10
20
30
40
1.2
5.3
4
4.7
28.7
42.2
3.7
18.9
28.5
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

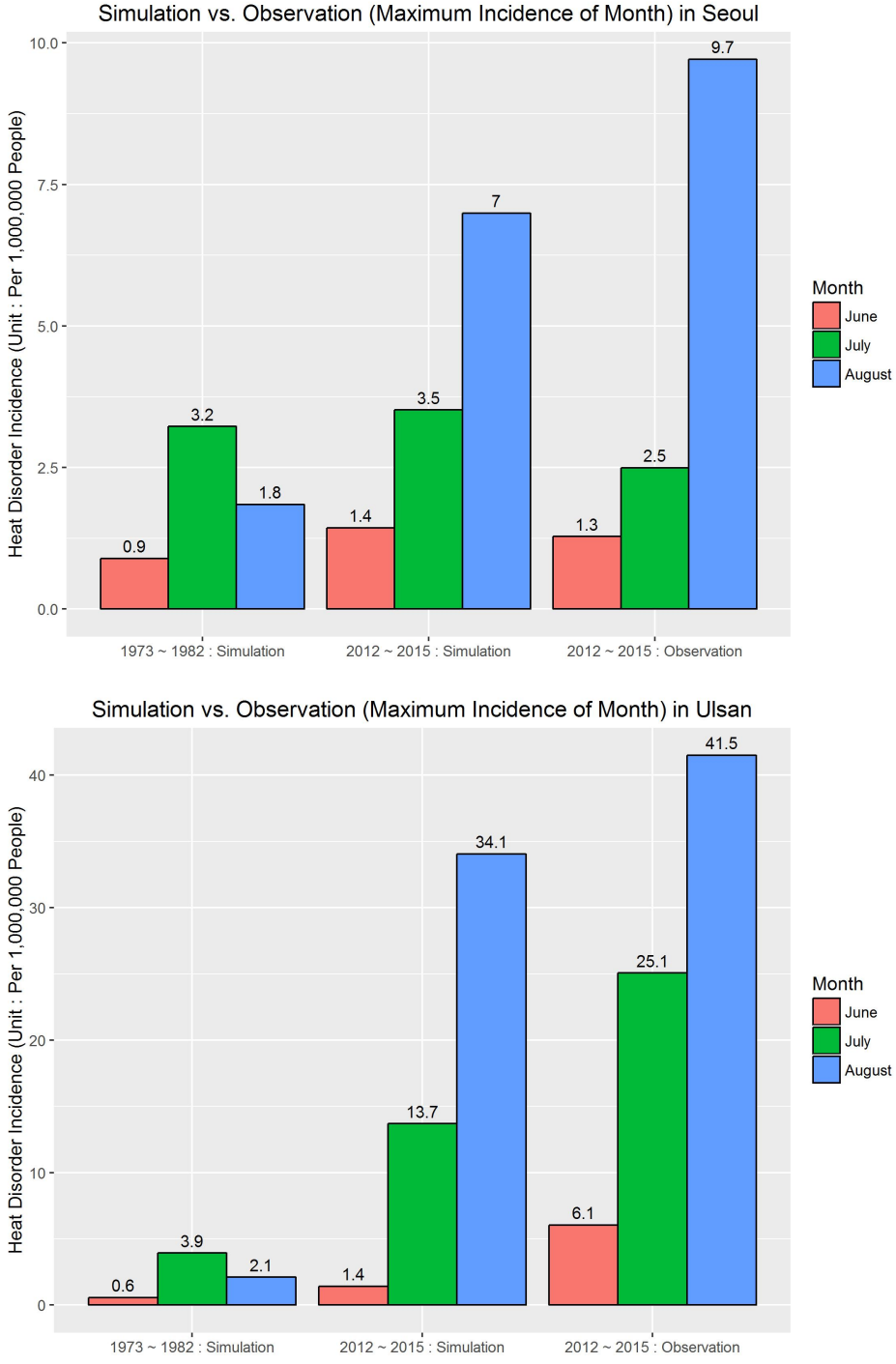

Simulation vs. Observation (Maximum Incidence of Month) in Gyeongnam
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
10
20
30
2.6
11.7
10
1
5.7
10.7
5.1
20
36.6
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August


Simulation vs. Observation (Maximum Incidence of Month) in Incheon
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
2
4
6
8
1.4
6.1
3.9
1.3
2.5
6.3
0.4
4.2
8.1
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August


Simulation vs. Observation (Maximum Incidence of Month) in Jeju
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
20
40
60
8.3
36.8
30.9
2.8
34
41.1
8.4
57.3
60.6
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August


Simulation vs. Observation (Maximum Incidence of Month) in Jeonbuk
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
10
20
1.6
7.2
5.6
1.8
6.6
11
4.3
16
26.2
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August


Simulation vs. Observation (Maximum Incidence of Month) in Jeonnam
Heat Disorder Incidence (Unit : Per 1,000,000 People)
0
20
40
60
4
25.5
17.8
4.1
20.8
39.2
7.9
38.2
59.3
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August


Simulation vs. Observation (Maximum Incidence of Month) in Seoul
Heat Disorder Incidence (Unit : Per 1,000,000 People)
10.0
7.5
5.0
2.5
0.0
0.9
3.2
1.8
1.4
3.5
7
1.3
2.5
9.7
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August
Simulation vs. Observation (Maximum Incidence of Month) in Ulsan
Heat Disorder Incidence (Unit : Per 1,000,000 People)
40
30
20
10
0
0.6
3.9
2.1
1.4
13.7
34.1
6.1
25.1
41.5
1973 ~ 1982 : Simulation
2012 ~ 2015 : Simulation
2012 ~ 2015 : Observation
Month
June
July
August

Our model does not consider socioeconomic features but only considers the natural conditions for heat disorder incidence. KCDC does not share these features, such as age or sex, of heat disorder incidence with its daily data. If we add these features, for example the age of patients or the place where the patients were transported to hospitals, we could improve the prediction power of our model.

This simulation is the solution for our second question. This can be a measure of heat disorder incidence risk. We provided the possible maximum heat disorder incidence by month and province and the comparisons. This might be helpful to identify *risky* provinces and expect the level of heat disorder incidence in the future.

Chapter 5

Conclusion

We started our study with two questions below.

(1) *First, might there be any visible proof of climate change in South Korea in terms of probability?*

(2) *Second, how can we calculate the risk of climate as an effect to health?*

For the first question, we calculated empirical (probability) density functions of temperature and humidity variables. Then, we compared the least updated density functions (only used data between 1972 and 1982) and the most updated one (used all 43-year data) to capture any temporal change which might be a proof of climate change. Most updated empirical (probability) density functions of temperature variables showed a slight move to a hotter zone. During this process, we discovered that several regions had a twin-peak in their density function lines of temperature variables. This appearance of twin-peck implied us an abrupt difference of temperature by

month so, the month variable was included in our GLM eqations.

Moreover, we demonstrated two kinds of simple hypothesis testing results calculated with variability variables in 1973 (the first year of our study period) and 2015 (the last year of our study period). We used range and standard deviation. The statistics of July and August showed an increase by time and we could easily reject the null hypotheses that the probability distribution of range and standard deviation in 1973 were same to those of 2015.

In general, the stationarity of statistics is often assumed in many time-series data analysis. This means the statistics of time-series variable do not change by time. And, it is well known that this assumption is not well applicable to the real world. Concerning of this, our analysis might not suggest definitive proofs of climate change. But, over the past 43 years, daily temperature variables show consistent increases in their variabilities and empirical (probability) density functions that would have caused an increase in the heat disorder incidence risk. Furthermore, we found that the variabilities of temperature variables positively increased especially in July and August.

These results raise the necessity to presume the situation in that the weather conditions threaten more the public health.

Connecting this concern to our model analysis, we could answer the second question. On the basis of Chapter 3. Basic Analysis, we chose 5 variables: the regional mean of daily highest temperature of each month and year (the level of temperature), the regional standard deviation of daily average temperature of each month and year (the variability of temperature), the regional

mean of daily average relative humidity of each month and year (the level of relative humidity), month which varied from June to August as a nominal variable and province name as another nominal variable. An interaction term of the regional mean of daily highest temperature and province was inserted into Model 4.1, Model 4.2 and Model 4.3. The population variable was included in Model 4.2. We tested whether there might be a linear relationship between the variables relevant with temperature or not using the correlation procedure and we did not find any relationship.

We conducted GLMs and HGLMs which allowed us consider a simple random effect (Model 4.5) and a special type of random effect conditioning spatial correlation structure (Model 4.6). We affirmed the possibility of random effect conditioning spatial correlation structure to improve the result of modeling without complex and ambigous interpretation.

At first, we calculated the linear models with the original count as response variable and then did the same process again with incidence count per 1 million. We could discover one possibility to reduce cAICs and deviances to gain better goodness of fits. The response variable of incidence count per 1 million showed a better goodness of fit.

These linear models allowed us to discover the significant estimates of the regional mean of daily highest temperature, the regional standard deviation of daily average temperature and month (especially July compared to June). Not only was the level of temperature (mean of daily highest temperature) helpful to predict heat disorder incidence, but also the variability (standard deviation of daily average temperature) of temperature.

Among our linear models, Model 4.6 assumed an spatial correlation struc-

ture as a random effect gave us the estimates to review more thoroughly. According to this model, Jeju and Jeonnam were the regions which had the significant estimates of province. The empirical (probability) density functions of them showed a twin-pick and, moreover, only these two regions had a higher second peak. This twin-peak implies the existence of multiple dominant probability density functions. We observed this phenomenon on the density plots in Section 3.1 (daily highest temperature) and the similar density plots in Appendix (daily average temperature). The estimates and their t-values from Model 4.6 helped us detect this characteristic, too. This detection led us to try to explain heat disorder incidence more with the explanatory variables of natural conditions and not to rely too much on individual unknown characteristics of each province by using nominal province variable. We may think spatial correlation structure as a random effect instead of nominal province variable.

Comparing the 10-year simulation results with the observation, we focused on the under-prediction of our model. August, the hottest month of year in South Korea, showed more under-predictions than other months. Then, we identified more dangerous regions such as Gwangju, Gyeongbuk, Gyeongnam, Jeju, Jeonbuk, Jeonnam and Ulsan. These regions showed bigger gaps between the simulated maximum heat disorder incidence from 1973 to 1982 and the observed maximum heat disorder incidence from 2012 to 2015. This is the solution for our second question.

If we find better solutions to coordinate the spatial scales of KMA and KCDC and discover more socioeconomic features of heat disorder incidence, we might use this process to predict the risk of heat disorder incidence.

Reference

[1] Agresti, A. (2013) *Categorical data analysis* (3rd). Hoboken: John Wiley & Sons.

[2] Ahn, J. (2011). *Data analysis for biginners*. Seoul: Hannarae. (Korean)

[3] Golub, G.H. and van Loan, C.F. (2013). *Matrix Computations* (4th). Baltimore: Johns Hopkins University.

[4] Intergovernmental Panel of Climate Change (IPCC). (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation (summary for policy makers).*

[5] Korea Centers for Disease Control & Prevention (KCDC). (2016). *Annual report on the notified patients with heat-related illness in Korea (2015).* (Korean)

[6] Korea Meteorological Administration (KMA). (2014). *Korean climate change assessement report 2014*. (Korean)

[7] Lee, S. (2015). *The association between heat waves and emergency department visits from NEDIS in South Korea*. Seoul National University (Master's Thesis) (Korean)

[8] Lee, T., Heo, M., Lee, J., Lee, G. (2015). *Data visulation*. Seoul: KNOU

Press. (Korean)

[9] Lee, Y. and Nelder, J.A. (1996). Hierarchical generalised linear models (with discussion). *Journal of the Royal Statistical Society* **B**, 58, 619-656.

[10] Lee, Y., Nelder, J.A. and Pawitan, Y. (2006). *Generalized linear models with random effects*. Boca Raton: Chapman & Hall/CRC.

[11] Noh, M., *et al.* (2013). *Are there spatial and temporal correlations in the incidence distribution of scrub typhus in Korea?*. Korea Centers Disease Control & Prevention (KCDC). Elsevier Korea LLC.

[12] Noh, M. and Lee, Y. (2015), dhglm package: Double Hierarchical Generalized Linear Models for R Language. Busan: South Korea.

[13] Pebesma, E. J., *et al.* (2016). sp packages: Classes and methods for spatial data in R for R Language. Münster: Germany.

[14] R Core Team. (2016), R: A Language and Environment for Statistical Computing, Vienna: Austria.

[15] R Core Team. (2016), stats package for R: A Language and Environment for Statistical Computing, Vienna: Austria.

[16] Rönnegård, L., Shen, X. and Alam, M. (2010). hglm: a package for fitting hierarchical generalized linear models. *The R Journal* Vol. 2/2, December 2010.

[17] Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, **8**, 2, 158-183.

[18] Wickham, H. and Cheong, W. (2016), ggplot2 package: Elegant Graphics for Data Analysis for R Language. Huston: USA.

[19] Wickham, H. (2015), plyr package: The Split-Apply-Combine Strategy for Data Analysis for R Language, Huston: USA.

[20] Wickham, H. (2014), reshape2 package: Flexibly Reshape Data: A Reboot of the Reshape Package. Huston: USA.

Appendix

[1] The Details on KMA Stations

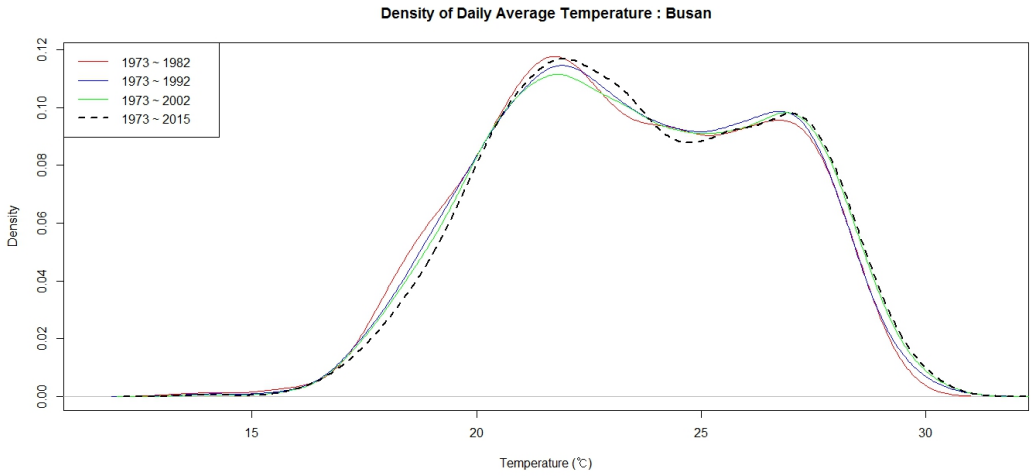
[2] Density Functions of Daily Average Temperature by Province

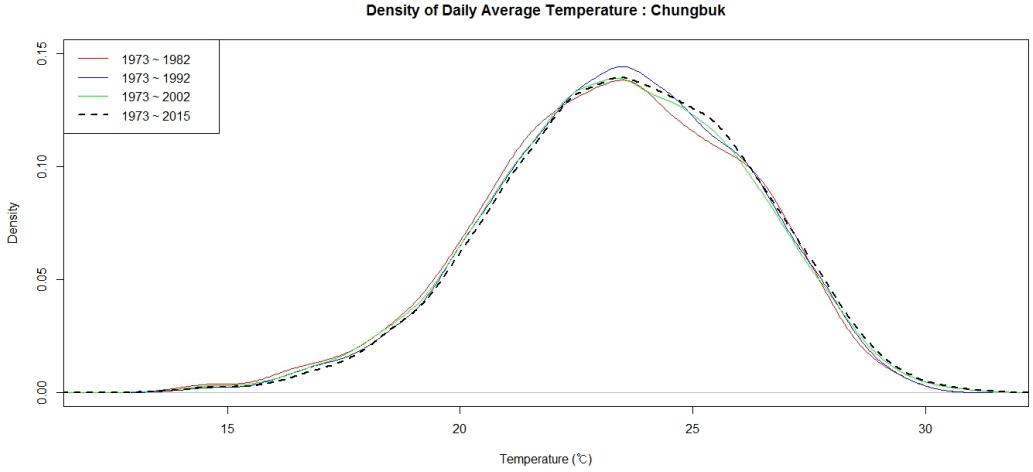
[3] Scatter plots (Climate vs. Heat Disorder Incidence) by Province

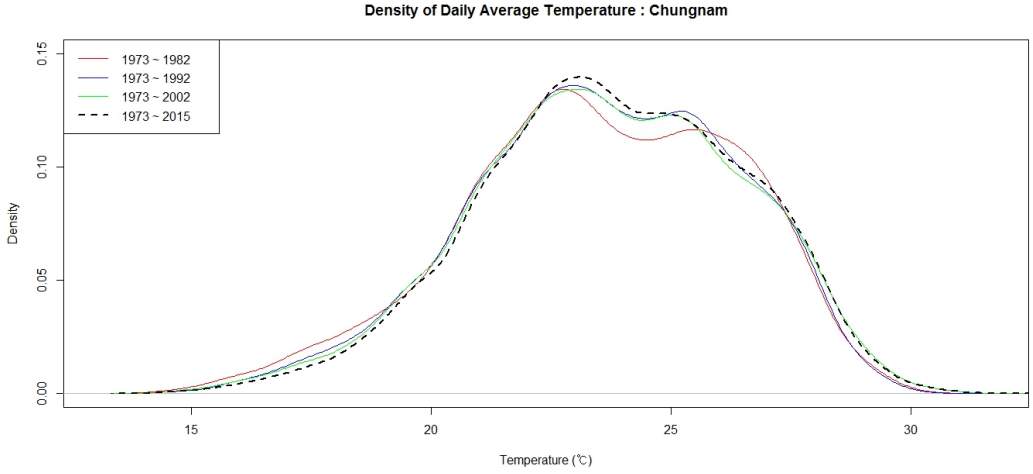
Province	Station	Lattitude	Longitude
Busan	Busan	35.06	129.01
Chungbuk	Chungju	36.58	127.57
	Cheongju	36.38	127.26
	Jecheon	37.09	128.11
	Boeun	36.29	127.44
Chungnam	Seosan	36.46	126.29
	Cheonan	36.46	127.07
	Boryeong	36.19	126.33
	Buyeo	36.16	126.55
	Geumsan	36.06	127.29
Daegu	Daegu	35.53	128.37
Daejeon	Daejeon	36.22	127.22
Gangwon	Sokcho	38.15	128.33
	Daegwallyeong	37.40	128.43
	Chuncheon	37.54	127.44
	Gangneung	37.45	128.53
	Wonju	37.20	127.56
	Inje	38.03	128.10
	Hongcheon	37.41	127.52
Gwnagju	Gwnagju	35.10	126.53

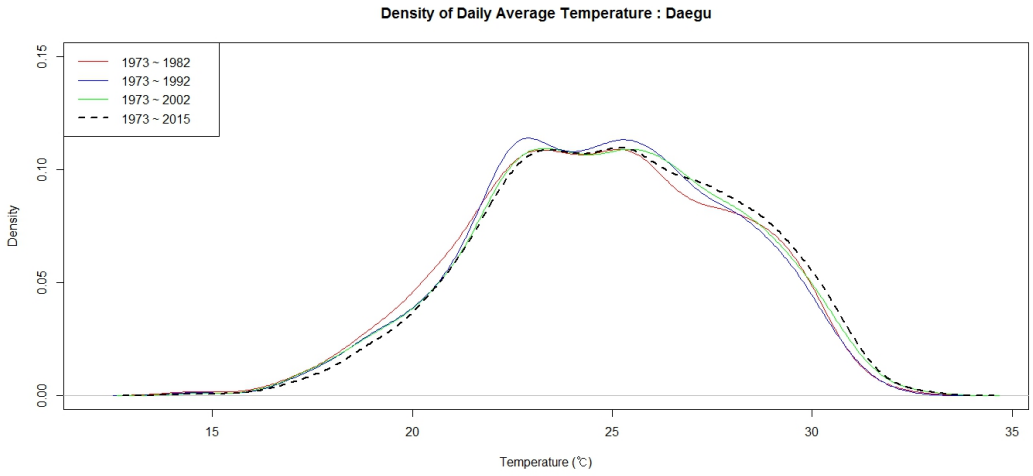
Province	Station	Lattitude	Longitude
Gyeonggi	Suwon	37.16	126.59
	Yangpyeong	37.29	127.29
	Icheon	37.15	127.29
Gyeongbuk	Ulleungdo	37.28	130.53
	Uljin	36.59	129.54
	Chupungnyeong	36.13	127.59
	Pohang	36.01	129.22
	Yeongju	36.52	128.31
	Mungyeong	36.37	128.08
	Yeongdeok	36.31	129.24
	Uiseong	36.21	128.41
	Gumi	36.07	128.19
	Yeongcheon	35.58	128.57
Gyeongnam	Tongyeong	34.50	128.26
	Jinju	35.09	128.02
	Geochang	35.40	127.54
	Hapcheon	35.33	128.10
	Miryang	35.29	128.44
	Sancheong	35.24	127.52
	Geoje	34.53	128.36
	Namhae	34.48	127.55
Incheon	Incheon	37.28	126.37
	Ganghwa	37.42	126.26

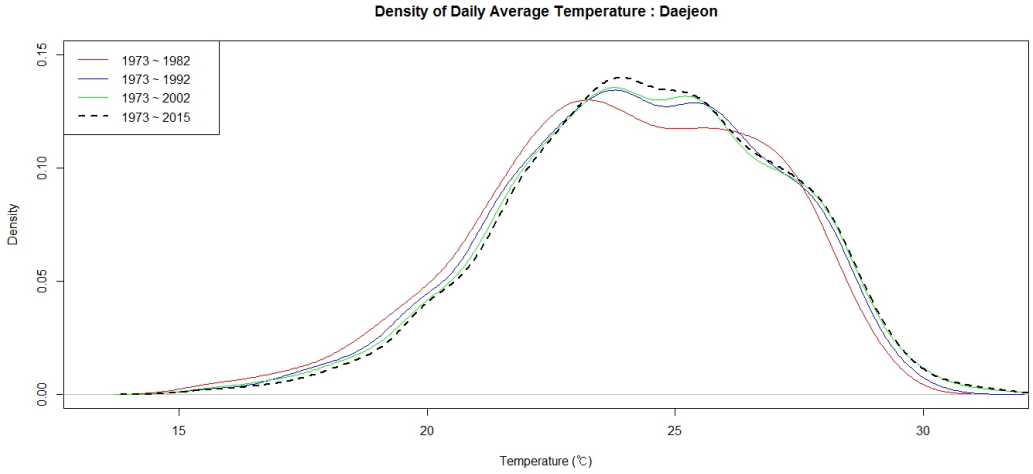
Province	Station	Lattitude	Longitude
Jeju	Jeju	33.30	126.31
	Seongsan	33.23	126.52
	Seogwipo	33.14	126.33
Jeonbuk	Gunsan	36.00	126.45
	Jeonju	35.49	127.09
	Buan	35.43	126.42
	Imsil	35.36	127.17
	Jeongeup	35.33	126.51
	Namwon	35.24	127.19
Jeonnam	Mokpo	34.49	126.22
	Yeosu	34.44	127.44
	Wando	34.23	126.42
	Jangheung	34.41	126.55
	Haenam	34.33	126.34
	Goheung	34.37	127.16
Seoul	Seoul	37.34	126.57
Ulsan	Ulsan	35.33	129.19

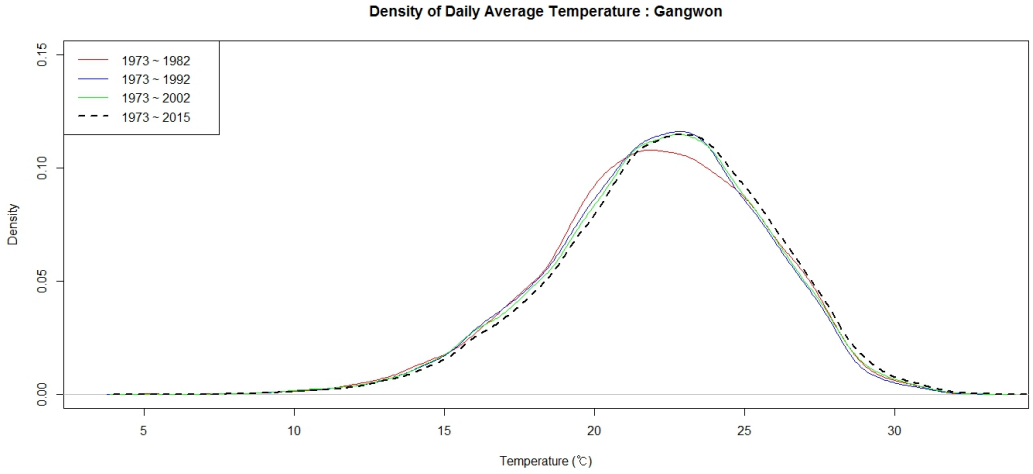

Density of Daily Average Temperature : Busan
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

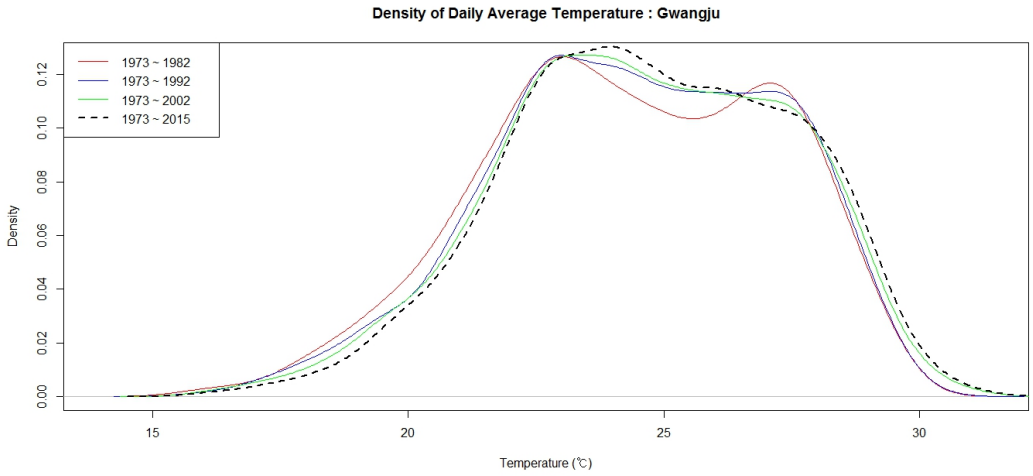

Density of Daily Average Temperature : Chungbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

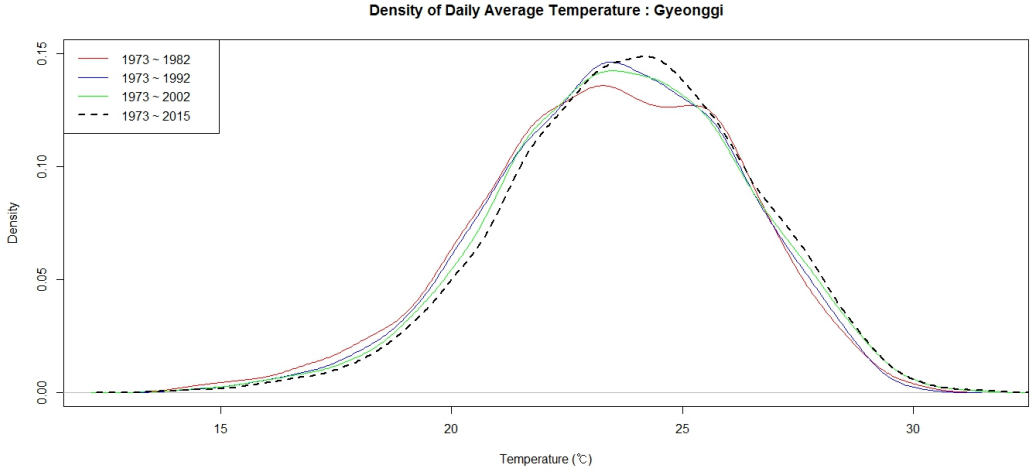

Density of Daily Average Temperature : Chungnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

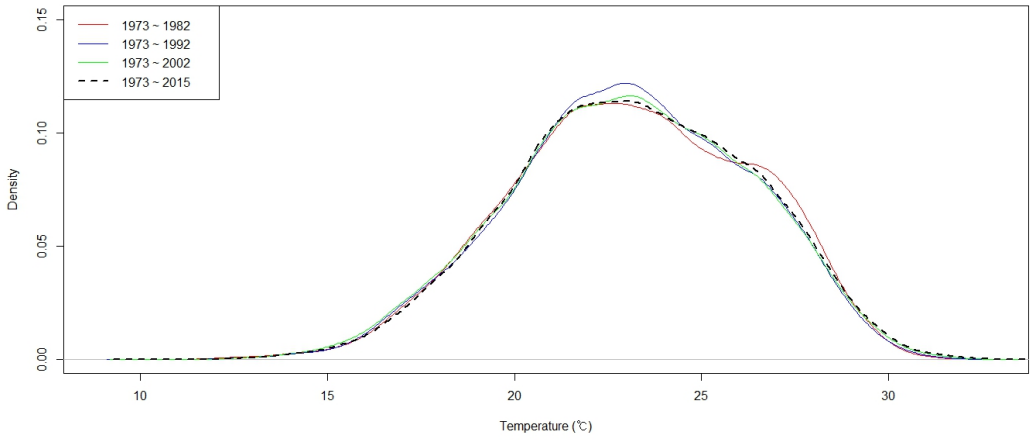

Density of Daily Average Temperature : Daegu
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
35
Temperature (℃)

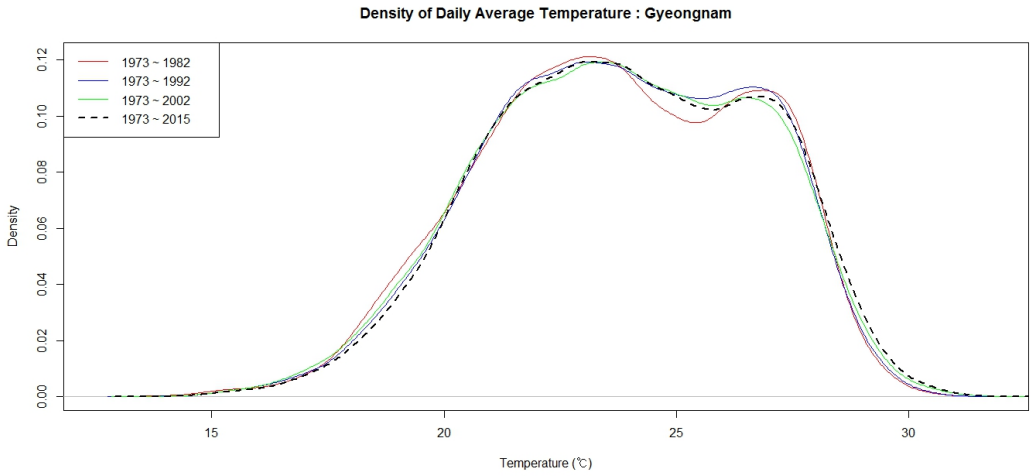

Density of Daily Average Temperature : Daejeon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

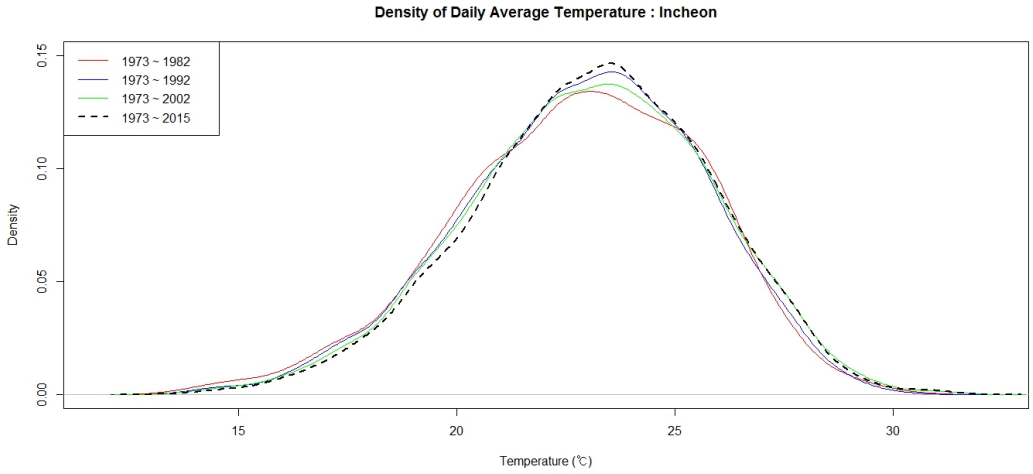

Density of Daily Average Temperature : Gangwon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
5
10
15
20
25
30
Temperature (℃)

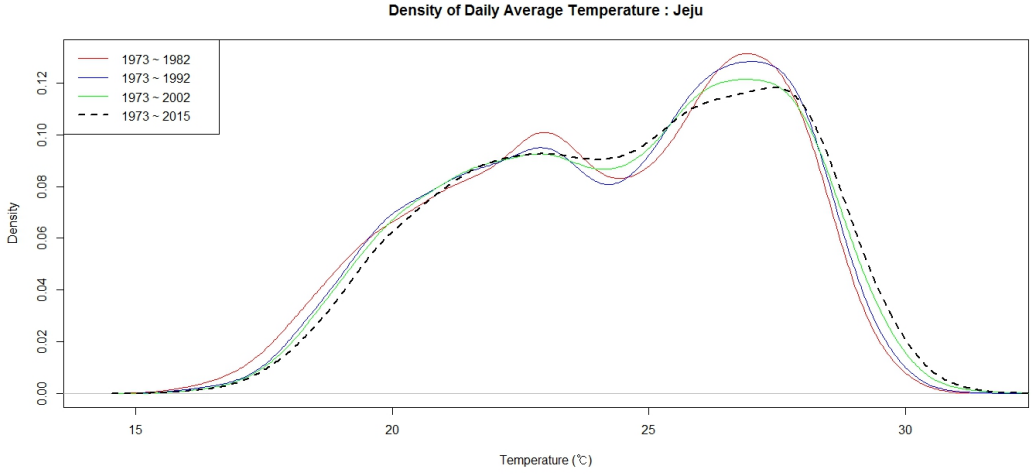

Density of Daily Average Temperature : Gwangju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

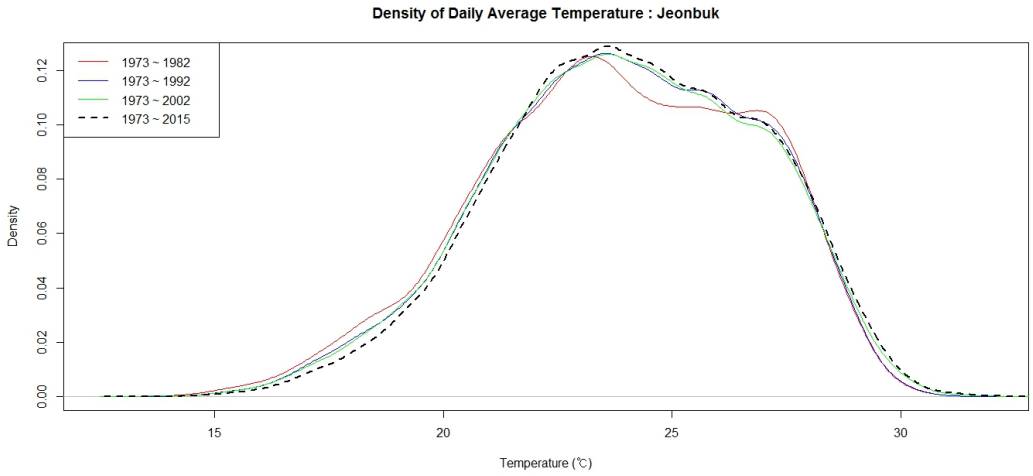

Density of Daily Average Temperature : Gyeonggi
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

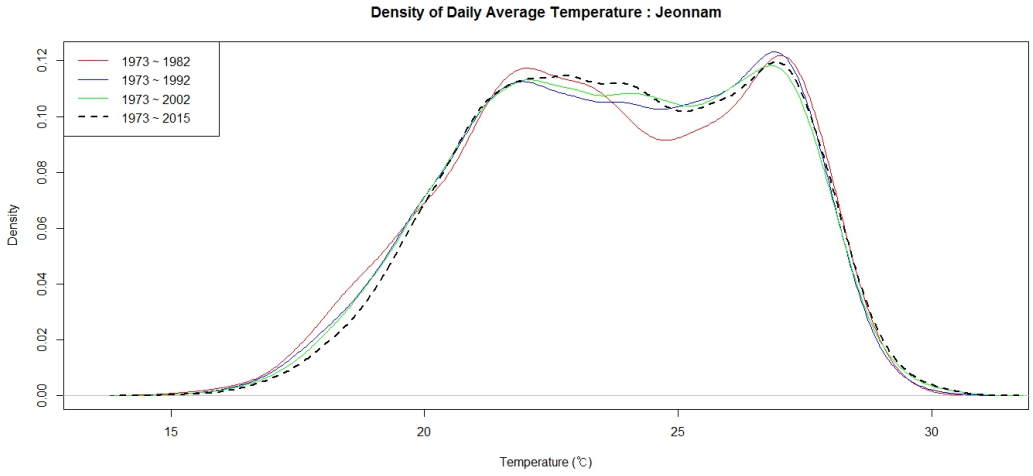

Density of Daily Average Temperature : Gyeongbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
10
15
20
25
30
Temperature (℃)

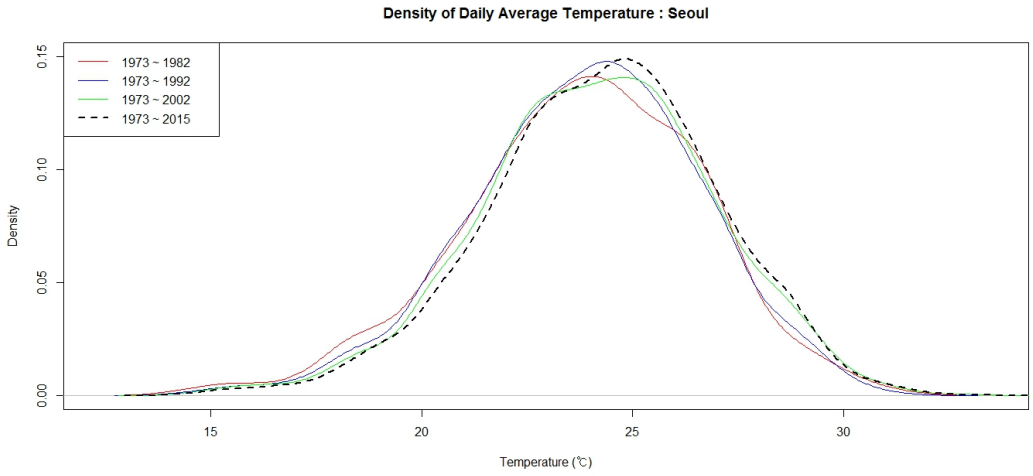

Density of Daily Average Temperature : Gyeongnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

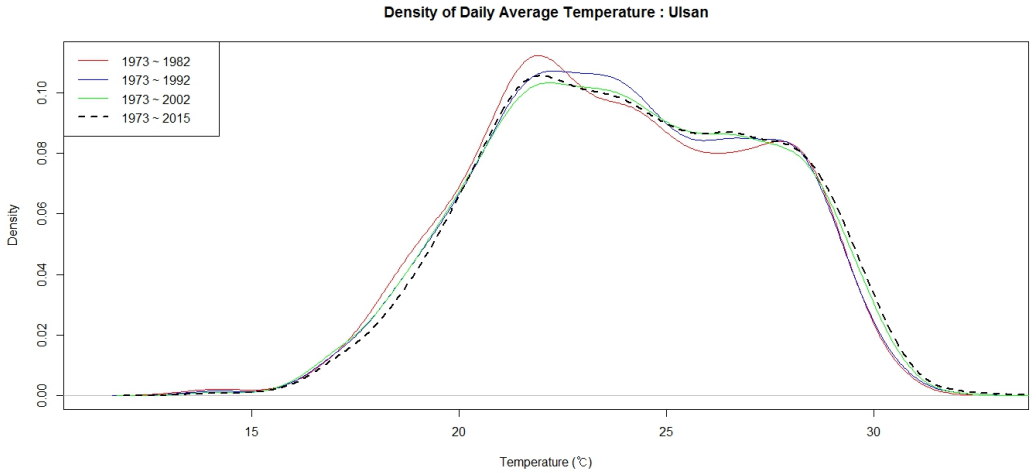

Density of Daily Average Temperature : Incheon
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

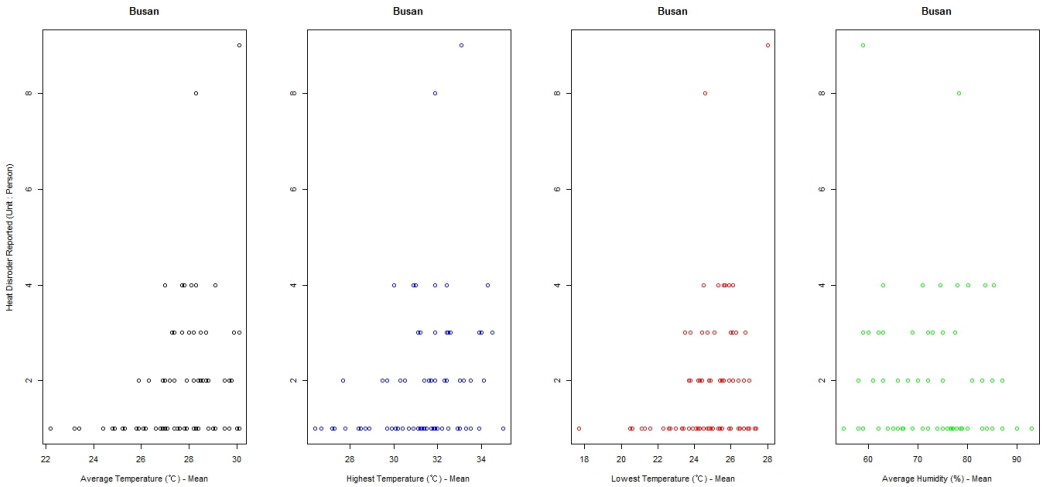

Density of Daily Average Temperature : Jeju
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

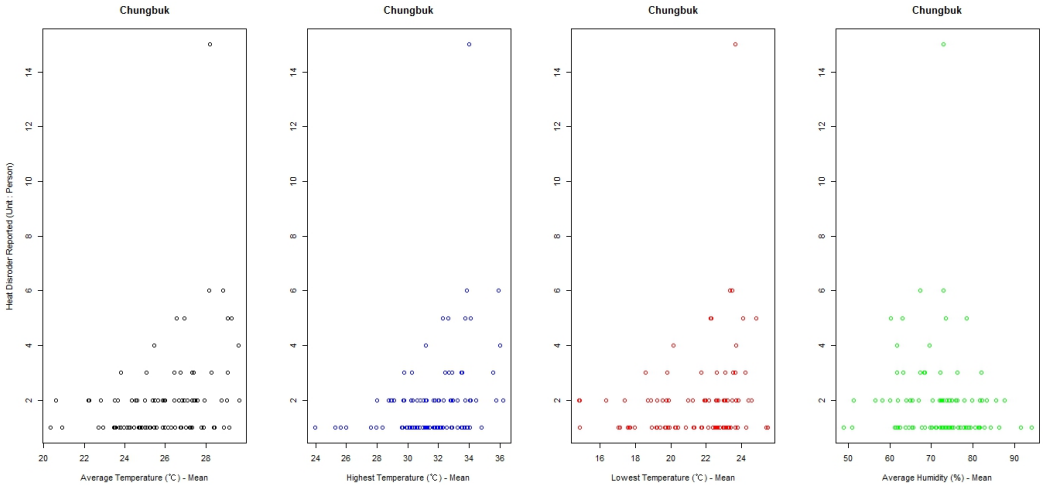

Density of Daily Average Temperature : Jeonbuk
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

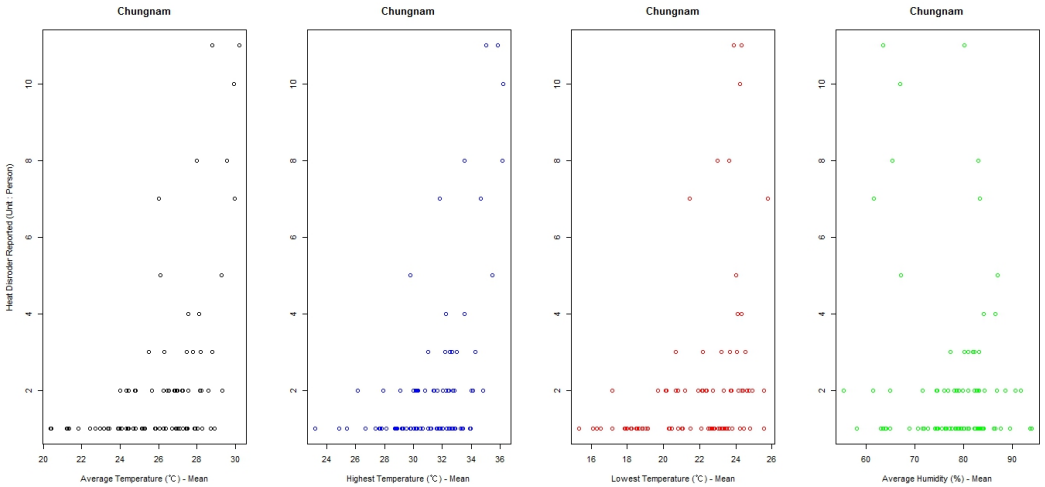

Density of Daily Average Temperature : Jeonnam
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
15
20
25
30
Temperature (℃)

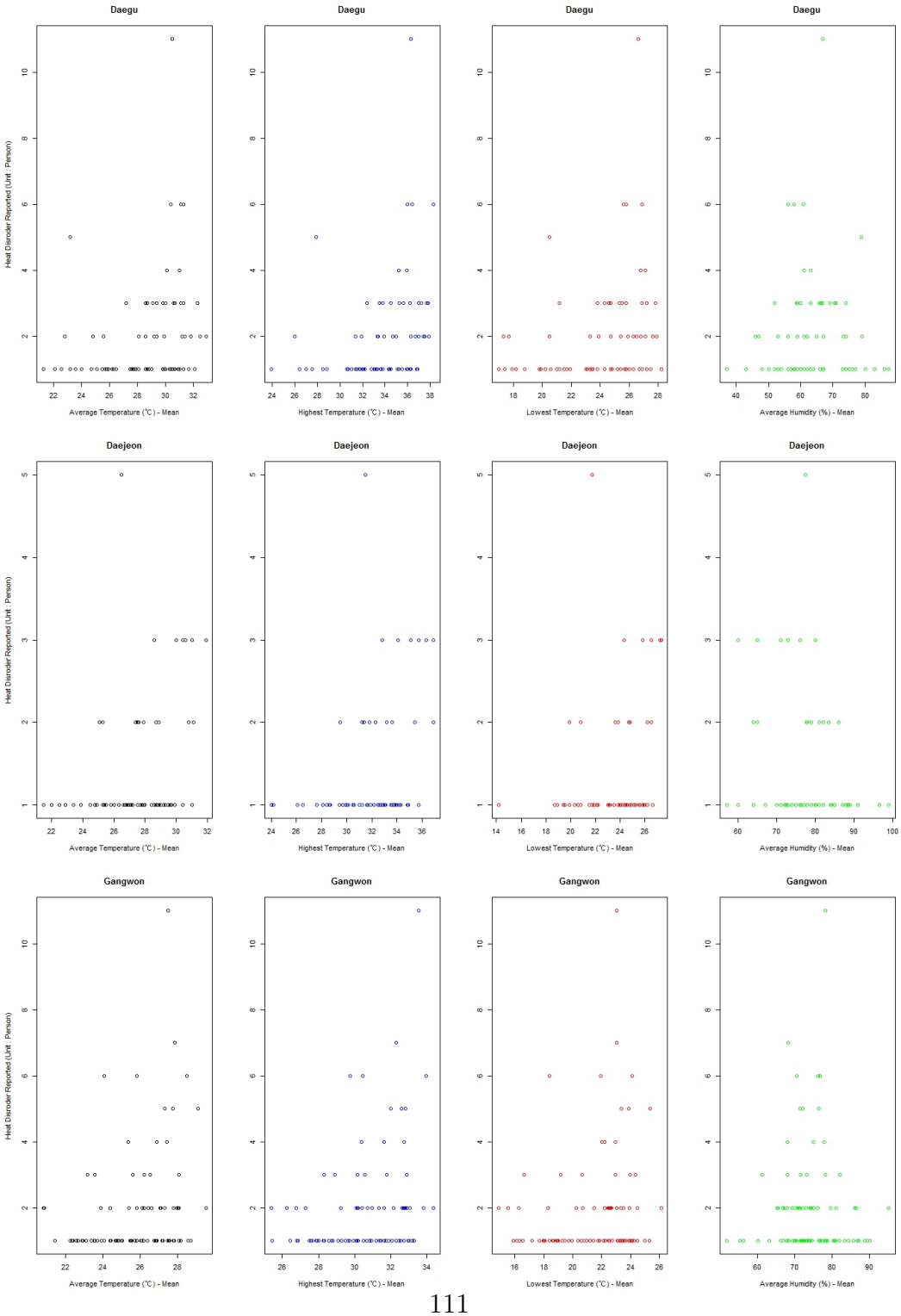

Density of Daily Average Temperature : Seoul
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.05
0.10
0.15
15
20
25
30
Temperature (℃)

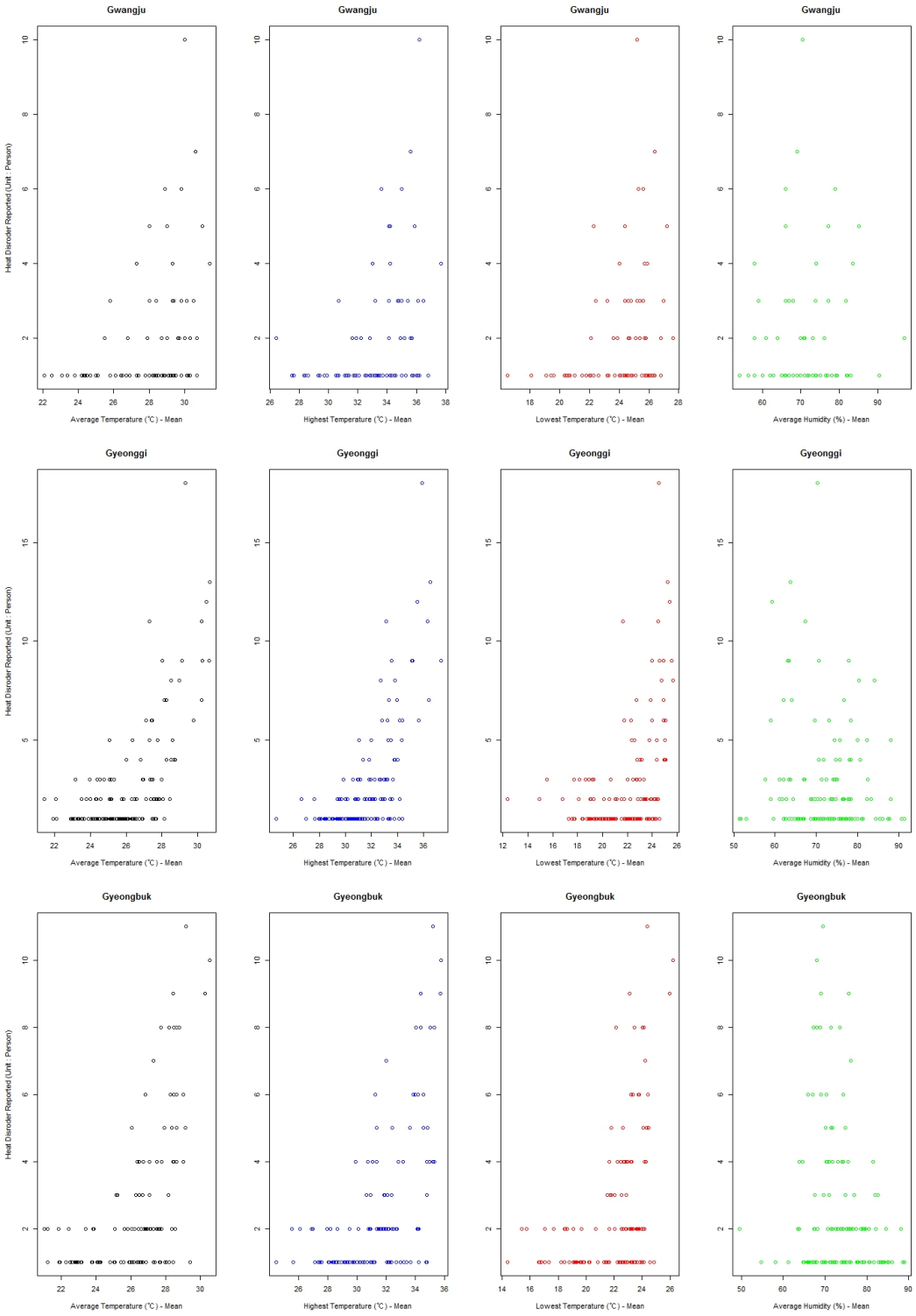

Density of Daily Average Temperature : Ulsan
1973 ~ 1982
1973 ~ 1992
1973 ~ 2002
1973 ~ 2015
Density
0.00
0.02
0.04
0.06
0.08
0.10
15
20
25
30
Temperature (℃)

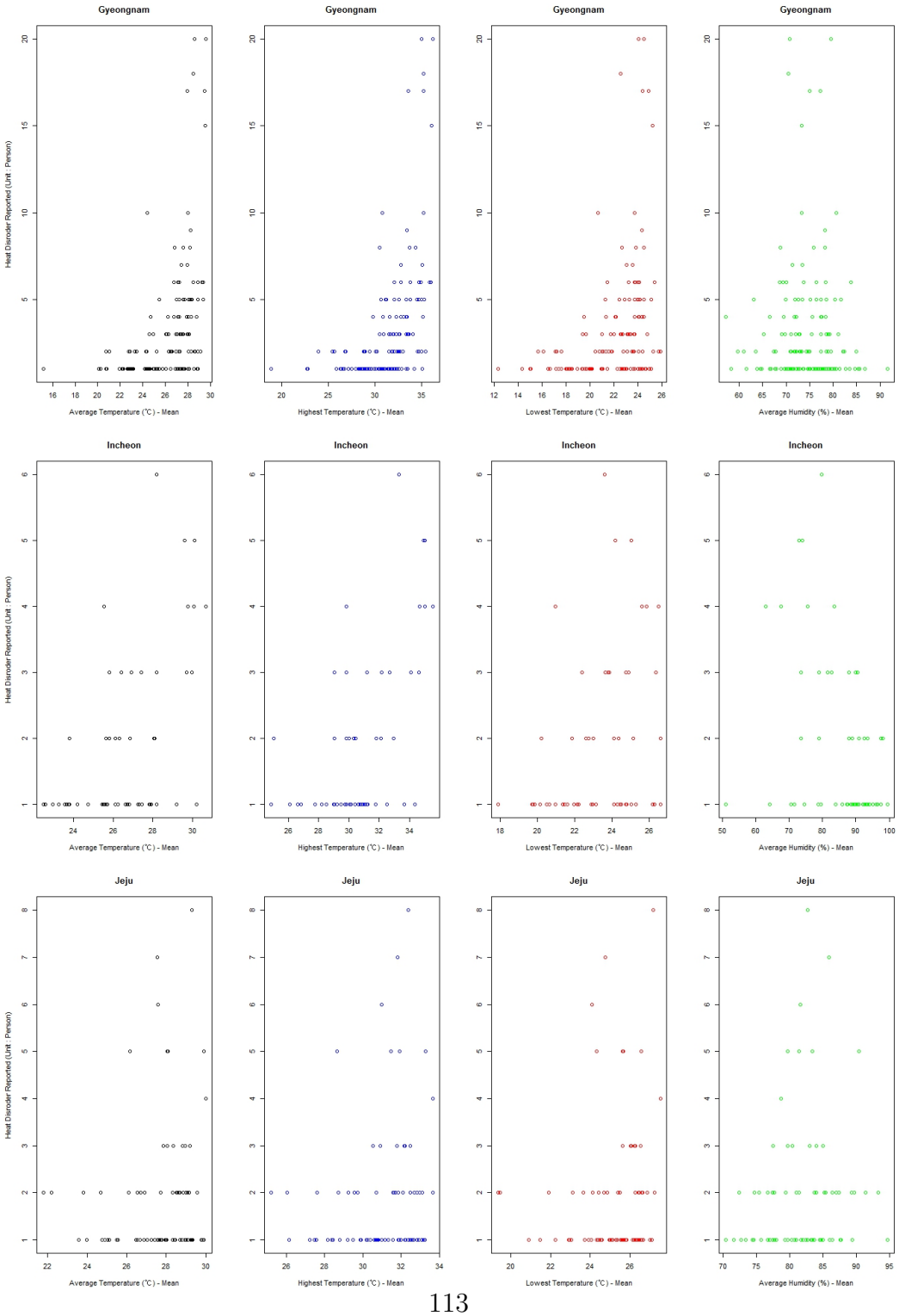

Busan
Busan
Busan
Busan
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean

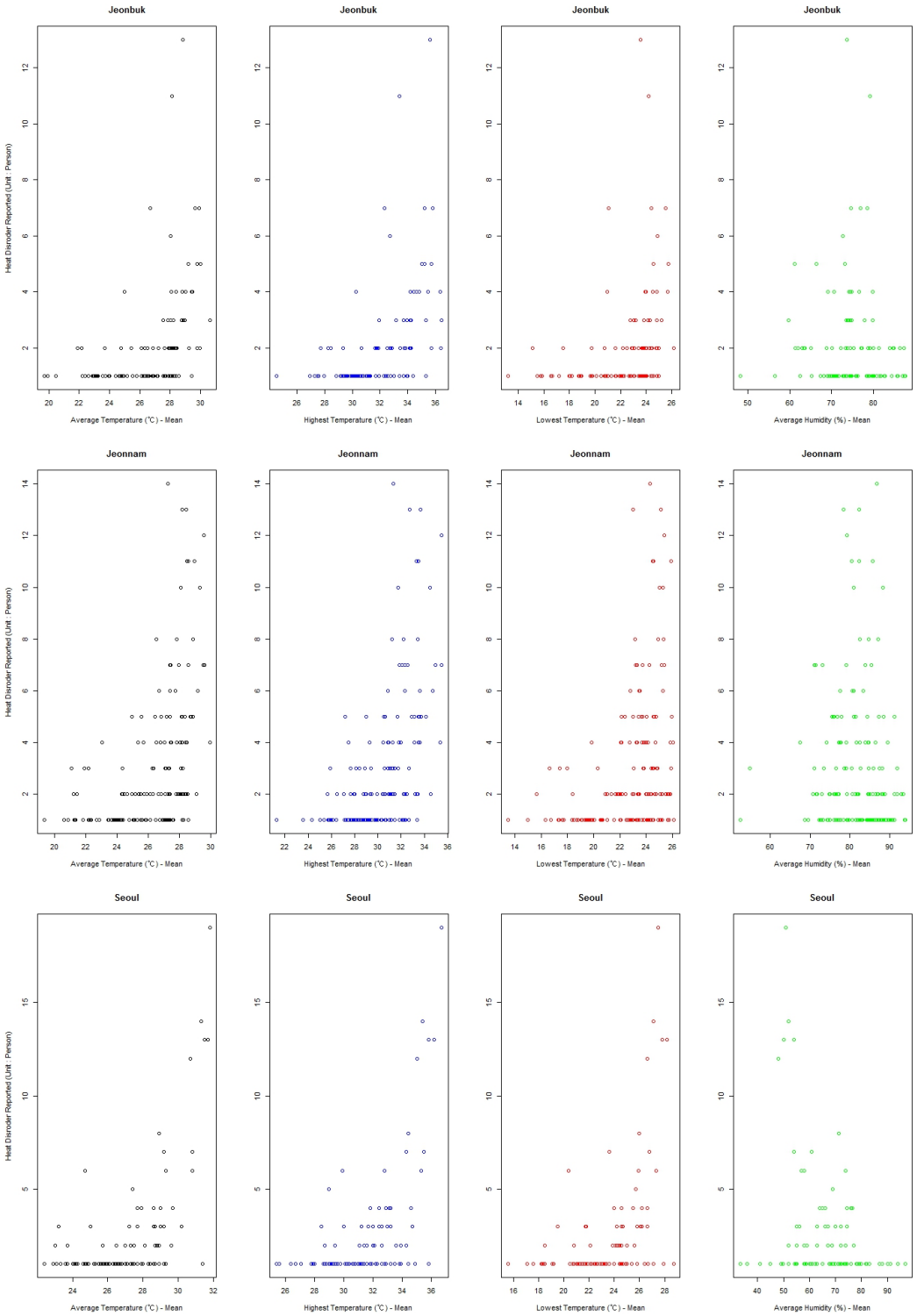

Chungbuk
Chungbuk
Chungbuk
Chungbuk
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean

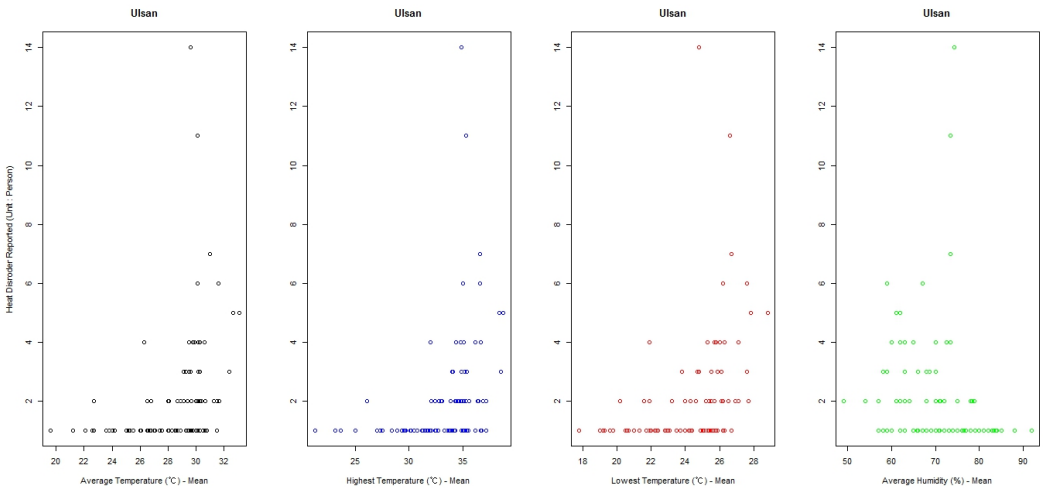

Chungnam
Chungnam
Chungnam
Chungnam
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean


Daegu
Daegu
Daegu
Daegu
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean
Daejeon
Daejeon
Daejeon
Daejeon
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean
Gangwon
Gangwon
Gangwon
Gangwon
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean


Gwangju
Gwangju
Gwangju
Gwangju
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean
Gyeonggi
Gyeonggi
Gyeonggi
Gyeonggi
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean
Gyeongbuk
Gyeongbuk
Gyeongbuk
Gyeongbuk
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean


Gyeongnam
Gyeongnam
Gyeongnam
Gyeongnam
Incheon
Incheon
Incheon
Incheon
Jeju
Jeju
Jeju
Jeju
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean


Jeonbuk
Jeonbuk
Jeonbuk
Jeonbuk
Jeonnam
Jeonnam
Jeonnam
Jeonnam
Seoul
Seoul
Seoul
Seoul
Heat Disroder Reported (Unit : Person)
Average Temperature (℃) - Mean
Highest Temperature (℃) - Mean
Lowest Temperature (℃) - Mean
Average Humidity (%) - Mean


Ulsan
Heat Disorder Reported (Unit : Person)
Average Temperature (℃) - Mean
Ulsan
Highest Temperature (℃) - Mean
Ulsan
Lowest Temperature (℃) - Mean
Ulsan
Average Humidity (%) - Mean

국문초록

이진희
협동과정 계산과학
자연과학대학
서울대학교

본 논문은 기후 현상(기온, 상대습도)과 온열질환(일사병 등) 발생 사이 관계에 대하여 선형 모형으로 분석한 내용을 제공한다. 우선, 여러 기술 통계량들로 기후 현상의 변화를 분석하였다. 분석 대상으로 선택된 기후 현상은 기온과 상대습도이다. 이 현상들의 확률 밀도 함수를 여러 기간 별로 시뮬레이션하여 각 변수 별로 시간에 따른 확률 분포 변화 여부를 보였는지 확인하였다. 이 과정에서 연구 대상 기간 내 첫 해(1973년)와 마지막 해(2015년)에 나타난 범위(range)와 표준편차(standard deviation)들로 두 해의 통계량들이 같은 확률 분포에서 나타난다는 영가설에 대하여 간단한 가설 검정 결과를 제공한다.

이런 기본 통계 분석으로 각 변수들을 이해한 후에, 온열질환 발생 신고수를 6월, 7월, 8월로 구분하여 월별 신고수와 100만 명당 신고 수를 종속변수로 두고 다단계 일반화 선형모형으로 기후 현상과의 관계를 분석하였다. 이 논문에서는 기후 및 온열질환 발생 수 데이터에서 나타나는 인접 지역간 상관형관계를 랜덤 효과로 모델에 반영하여 분석하였다. 랜덤 효과가 선형 모형 분석 결과를 어떻게 바꾸는지 확인하였다. 이 모델 결과를 활용하여 관찰이 이루어지지 않은 기간 동안 발생했을 온열질환 발생 수(리스크)를 계산한 후 현재와 비교하였다.

Keyword : *기후, 온열질환 발생, 다단계 일반화 선형모형, 랜덤 효과(random effect), 공간 상관관계*

Student Number : 2013-23011