

# BIGDATA X CAMPUS 프로젝트 보고서

지진발생에 따른 효율적인 대피방안 제시

Team

권기태  
김나영  
조수진  
한지우

# contents

1. 분석 배경

2. 활용 데이터 및 기법

3. 분석 및 시각화

4. 기대효과





## 지진이란?

- 정 의 : 지구 내부에서 급격한 지각변동이 생겨 그 충격으로 생긴 파동,  
즉. **지진파가 지표면까지 전해져 지반을 진동**시키는 것
- 발생 원인 : 암석권에 있는 판의 움직임
- 특 징 : 진앙에서 흔들림이 가장 세고,  
진앙에서 멀어지면 흔들림이 점점 약해져서  
어느 한계점을 지나면 느끼지 못하게 됨







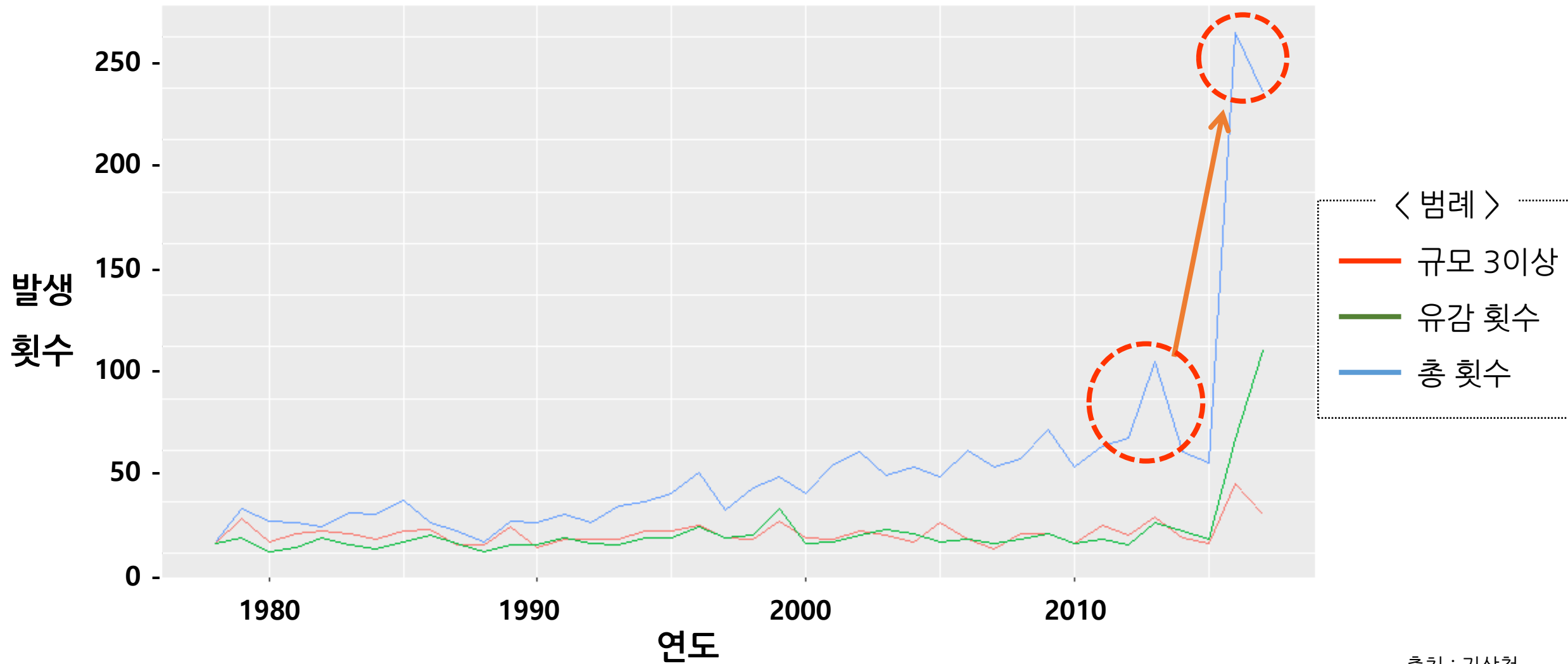
2016년 경주 지진(좌)



2017년 포항 지진(우)

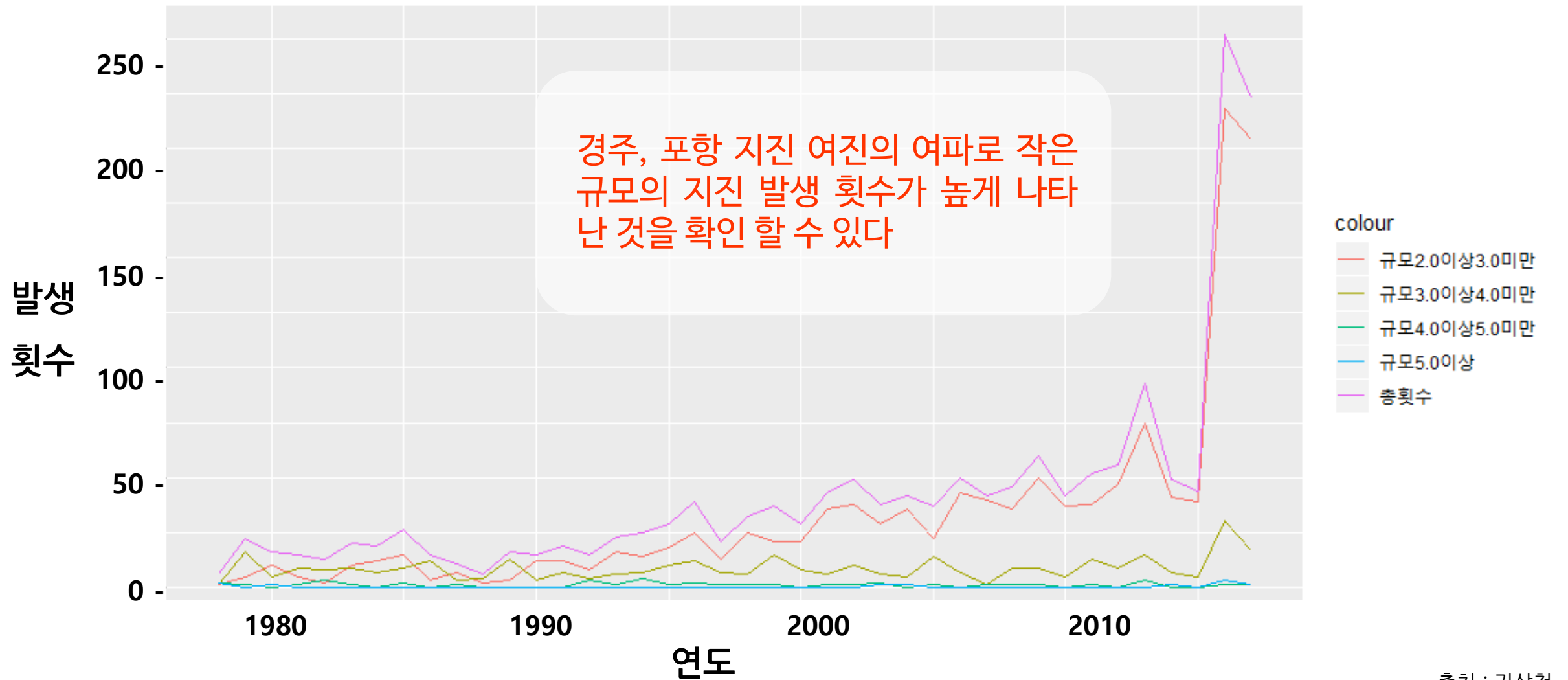
# 국내 지진 발생 추이

분석배경



# 국내 지진 발생 추이 (규모별)

분석배경





## 데이터 검색



## 데이터 전처리

### [지진데이터]

- 국내 지진데이터 1978~2018년 (날씨누리)
- 아시아 지진데이터 2003~2018년 (독일 지구과학연구소, GFZ)

### [지진피해액]

- 해외 사례 1980~2018년 (미국해양대기청, NOAA)

### [지진대피소현황]

- 실내 구호소 및 옥외대피소 (행정안전부)
- 전국지진해일대피소 표준데이터 (공공데이터포털)

### [인구수 데이터]

- 전국 및 지역별 인구수 데이터 (국가통계포털, KOSIS)

### [지진관련데이터]

- 결측치 제거
- 필요한 데이터 csv 형식으로 가공

### [지진대피소현황]

- 결측치 제거 또는 보완 (위도, 경도)
- 회귀분석을 통한 대피소수용인원 추정





다음 지진의 위치를 어떻게 하면 예측 할 수 있을까?

## 기존 방식

시간의 흐름에 따른 지진의 발생 위치와  
진도, 강도에 대한 데이터를 활용하여  
인공신경망 모델 내에서  
지각의 변동에 대한 내용을 학습케 함

단, 학습 변수가 부족하고,  
재현성이 낮다는 문제가 있음

## 새로운 방식 제안

기존 지진 사례의 피해 정도에 대한  
데이터를 활용해 진원의 깊이/강도/진도의  
변수에 따른 지진의 피해 정도를  
예측할 수 있도록 **Machine Learning,**  
**KNN ,c50, SVM 분석기법을 활용**



	location	lat	lon	depth	mag	int	damage
1	ROMANIA: CARPATHIAN FOLD, VRANCEA	45.700	26.600	150	7.4	10	2
2	PERU: LIMA	-12.100	-77.100	30	7.5	8	1
3	GEORGIA: KAKHETIYA	41.800	45.200	7	4.2	7	2
4	VENEZUELA: LA GUAIRA	10.600	-66.900	33	7.7	10	3
5	GUATEMALA: SOLOMA	15.500	-91.500	33	7.5	9	3

독립변수 X's

종속변수 Y

회귀식 :  $(\text{Depth}) + (\text{Mag}) + (\text{Int}) = (\text{Damage})$

진원의 깊이      규모      진도      피해정도 (1~4)

Call:

```
lm(formula = damage ~ mag + int + depth, data = damage_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.35802	-0.69855	-0.09472	0.68330	2.37809

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.08376	0.20116	0.416	0.6772
mag	0.07832	0.03492	2.243	0.0251 *
int	0.24266	0.02000	12.132	<2e-16 ***
depth	-0.00237	0.00103	-2.302	0.0215 *

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.935 on 897 degrees of freedom

Multiple R-squared: 0.2093, Adjusted R-squared: 0.2067

F-statistic: 79.16 on 3 and 897 DF, p-value: < 2.2e-16

0.05 유의수준 하에서 모든 독립변수가  
종속변수 Y ( 피해정도) 에 영향을 미친다.

### 추정 회귀식

$$(\text{damage}) = 0.08(\text{mag}) + 0.24(\text{int}) - 0.002(\text{depth}) + 0.08$$

## ① KNN

Step1.

데이터 셋을 train 데이터와  
test 데이터로 나눈다.

Step2.

train 데이터로 KNN모델을 구축하고  
test 데이터의 예측 값을 구한다.

Step3.

예측 값과 test데이터의  
target 변수를 비교하여  
정확도를 계산한다.

901개 관측치의 데이터 셋

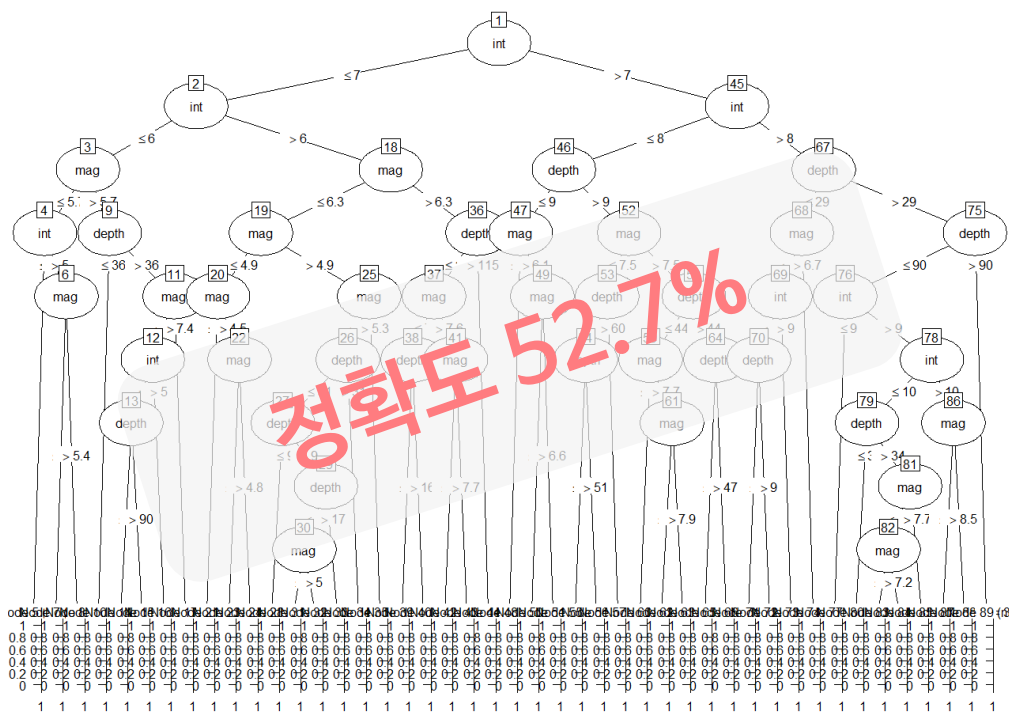


750개 관측치의 train 데이터  
151개 관측치의 test 데이터

```
knn(damage_train[,-4],damage_test[-4],  
cl=damage_train[,4], k=3)
```

**정확도 66.9%**

## ② C50



## ③ SVM

```
install.packages("e1071")
library(e1071)
model <- svm(damage ~., data=damage_train)
pred <- predict(model, newdata = damage_test[, -4])
mean(round(pred)==damage_test[,4])
```

정확도 43.3%





피해정도 예측 분석 결과, 학습 데이터의 개수와 랜덤 추출에 따라  
모델의 성능 편차가 너무 컸고, 평균적으로 50~60% 정도의 정확도로 낮은 성능을 보였다.  
또한 지진 발생지의 내진 설계 등 예방 정도를 고려하지 못했다는 점에서 한계가 명확했다.  
이에 따라 예측보다는 **예방과 대비가 더 중요**할 것이라고 판단하였고,  
현재 우리나라의 지진 대비 정도를 중심으로 시각화해보고,  
**최적의 대피방안을 제시**하는 것으로 방향을 수정했다.



## 1. Load packages

## 2. Load map data

## 3. Transform coordinate (UTMK → WGS84)

## 4. Load data (Population, Shelter info)

## 5. Visualize data

### 1. Load packages

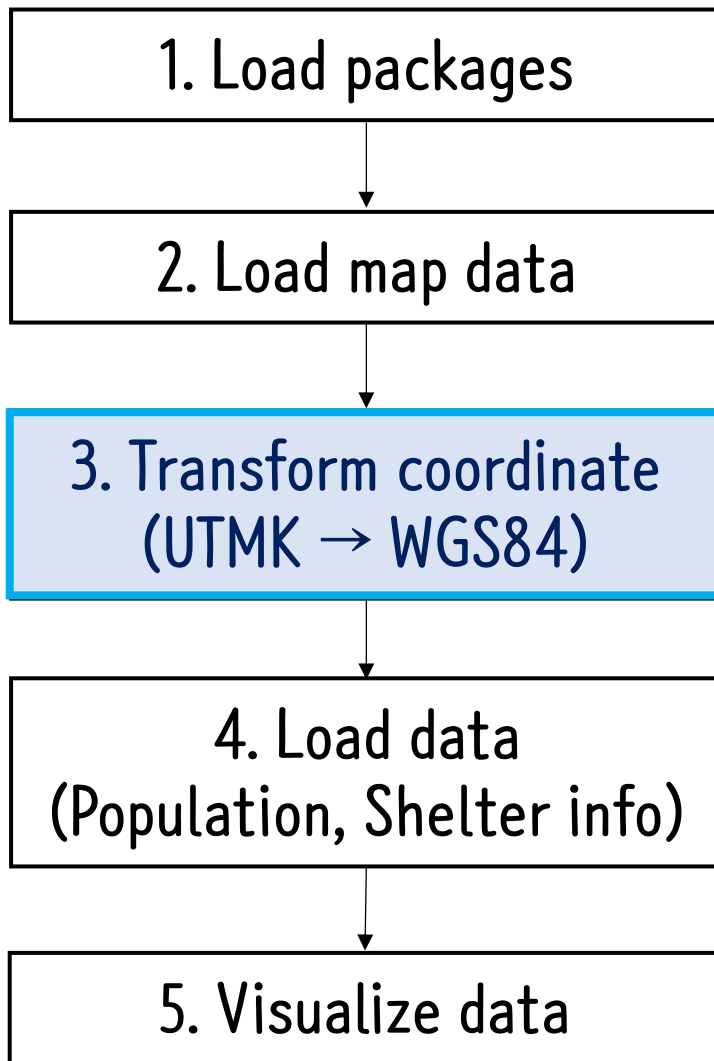
시각화에 필요한 패키지들을 설치하고, 불러온다.

```
ggmap, ggplot2, raster, rgeos, maptools, rgdal, sp
```

### 2. Load map data

원하는 구역의 위치정보 (위도, 경도)와 행정구역 정보를 불러온다.

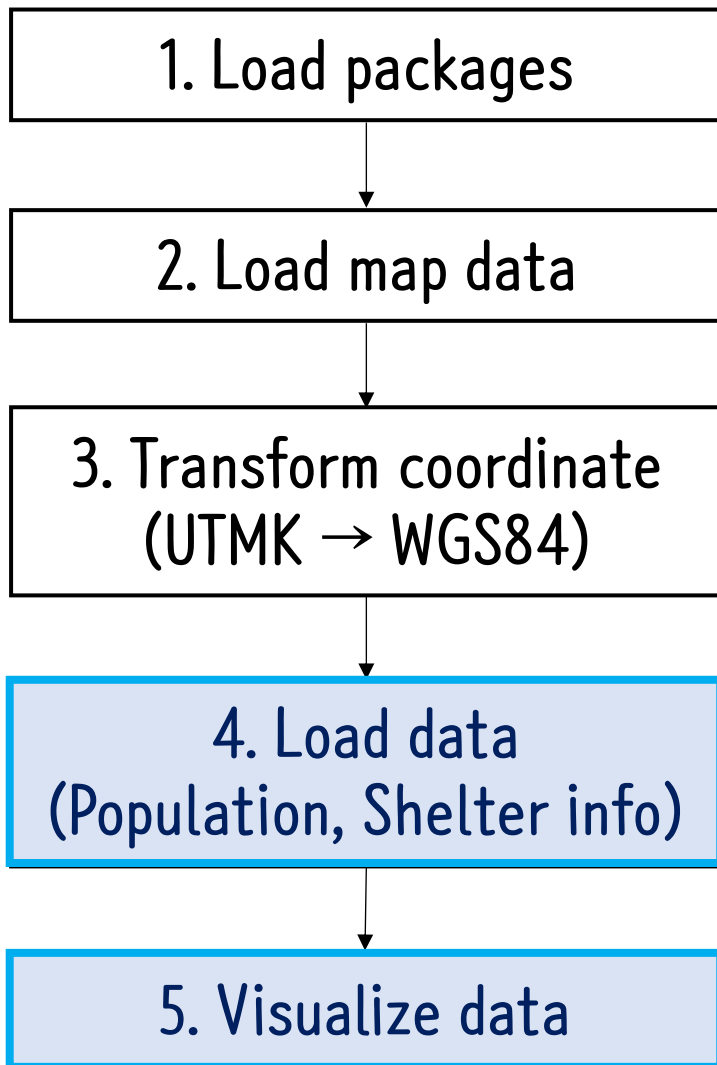
```
shp1 <- readShapePoly("TL_SCC0_CTPRVN.shp") # 시도별  
map1 <- fortify(shp1)
```



### 3. Transform coordinate (UTMK → WGS84)

불러온 행정구역지도 정보는 UTM-K 좌표계를 가지고 있기때문에 spCoordinateTrans 함수를 이용하여 WGS84 좌표계로 변경해준다.

```
spCoordinateTrans <- function(long, lat, from.crs = "+proj=tmerc  
+lat_0=38 +lon_0=127.5 +k=0.9996 +x_0=1000000 +y_0=2000000  
+ellps=GRS80 +units=m +no_defs", to.crs = "+proj=longlat +ellps=WGS84  
+datum=WGS84 +no_defs") {  
  xy <- data.frame(long = long, lat = lat)  
  sp::coordinates(xy) <- ~ long + lat  
  f_coordinates <- sp::SpatialPoints(xy, sp::CRS(from.crs))  
  res <- sp::spTransform(f_coordinates, sp::CRS(to.crs))  
  return(as.data.frame(res))  
}
```



#### 4. Load data (Population, Shelter info)

시각화하고자 하는 지역의 인구수 또는 지진대피소 정보를 불러온다.

```
pop_all <- read.csv("SID_population.csv", header = T) # 전국 인구수
```

#### 5. Visualize data

행정구역지도에 불러온 정보를 매칭시켜 시각화 한다.

```
merge_data <- merge(map1, pop_all, by = "id")
```

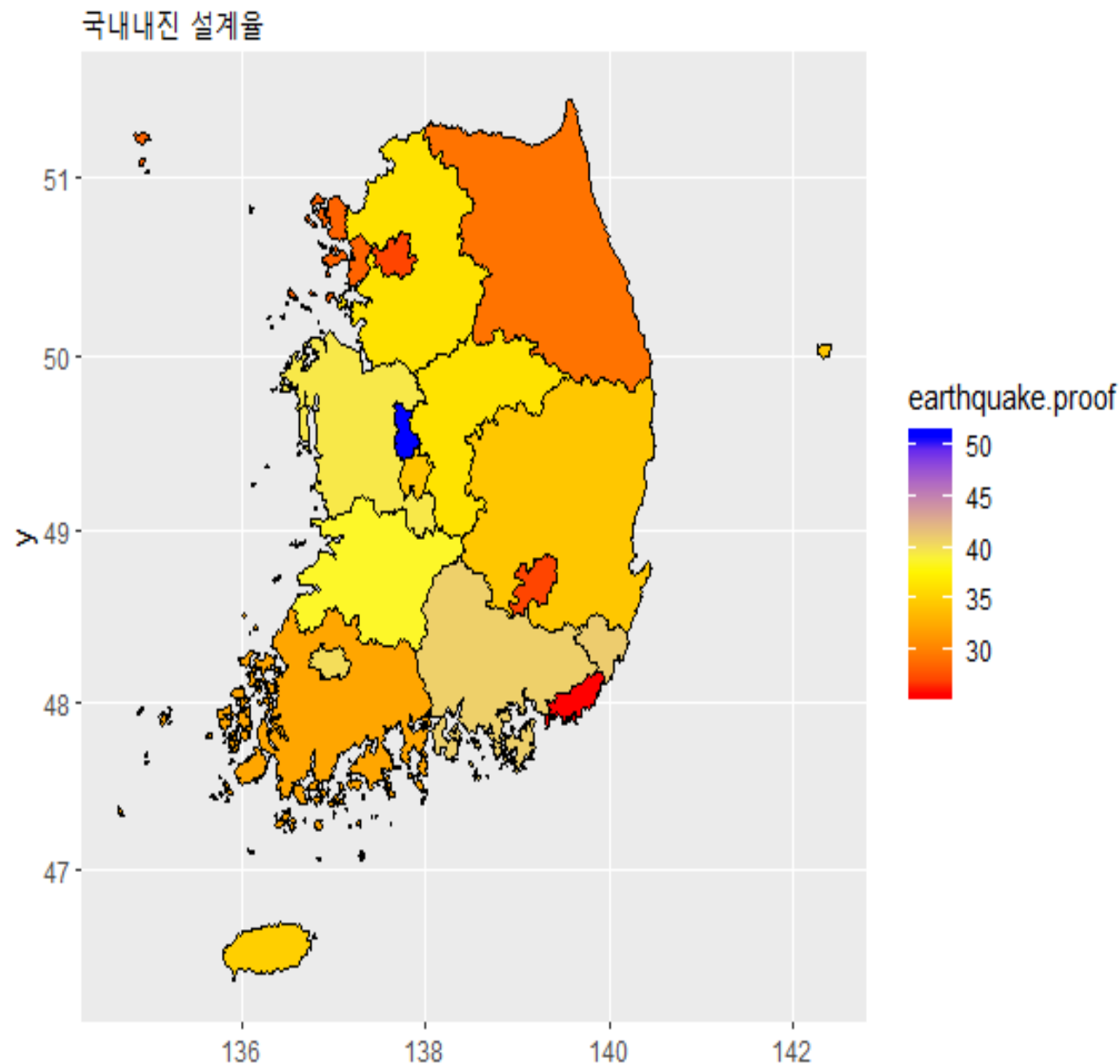
```
map <- ggplot() + geom_polygon(data = merge_data, aes(x = long, y = lat,  
group = group, fill = Population)) + labs(fill = "Population of Korea")
```

```
map + scale_fill_gradient(low = "pink", high = "red")
```

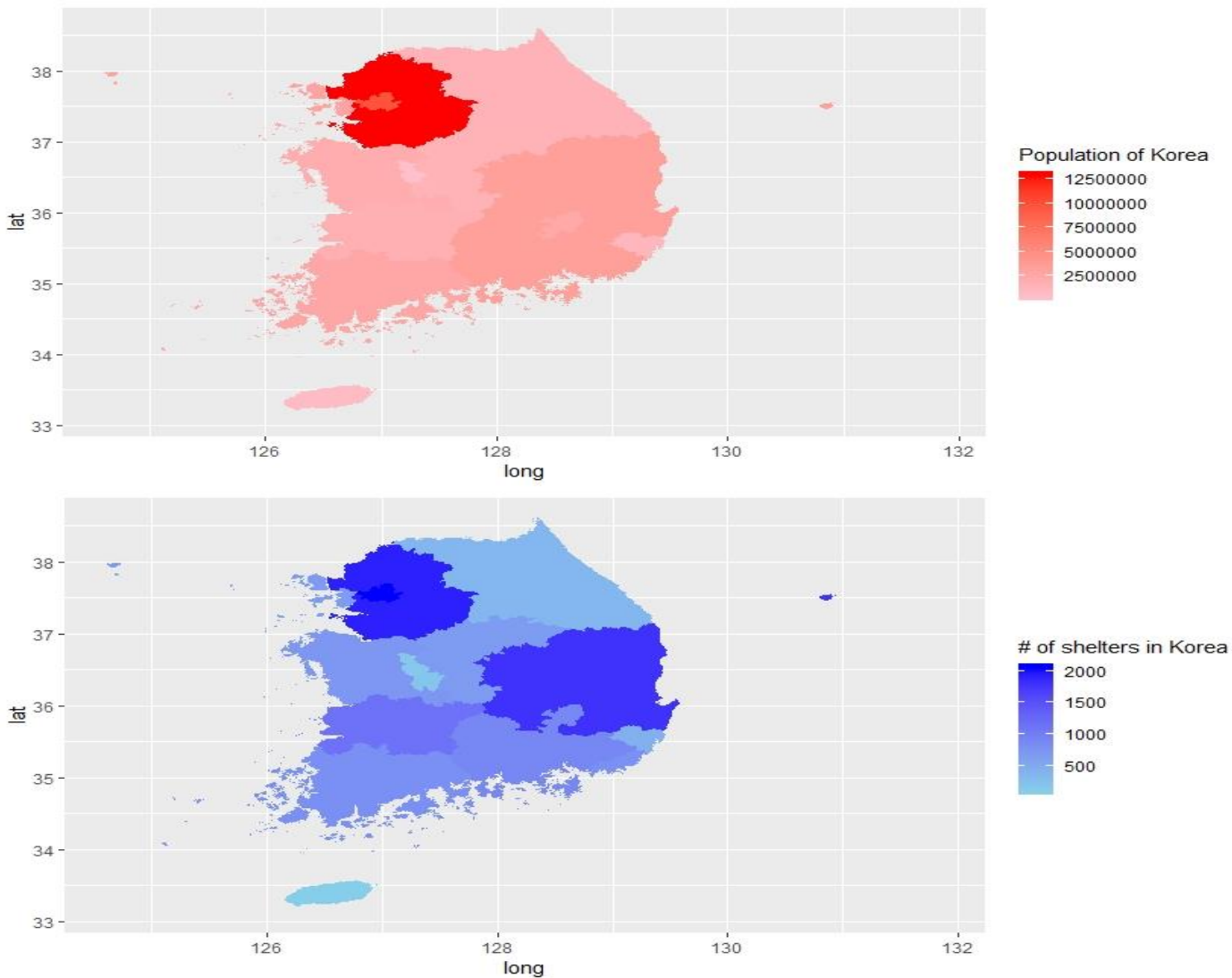


## 국내 건물 내진 설계율

- ✓ 비교적 최근에 도시가 조성된  
세종특별자치시는 내진율이 높게 나타났고,  
부산광역시와 대구광역시, 서울특별시 등  
노후화된 건물이 많은 지역은  
내진 설계율이 낮은 것을 볼 수 있다.



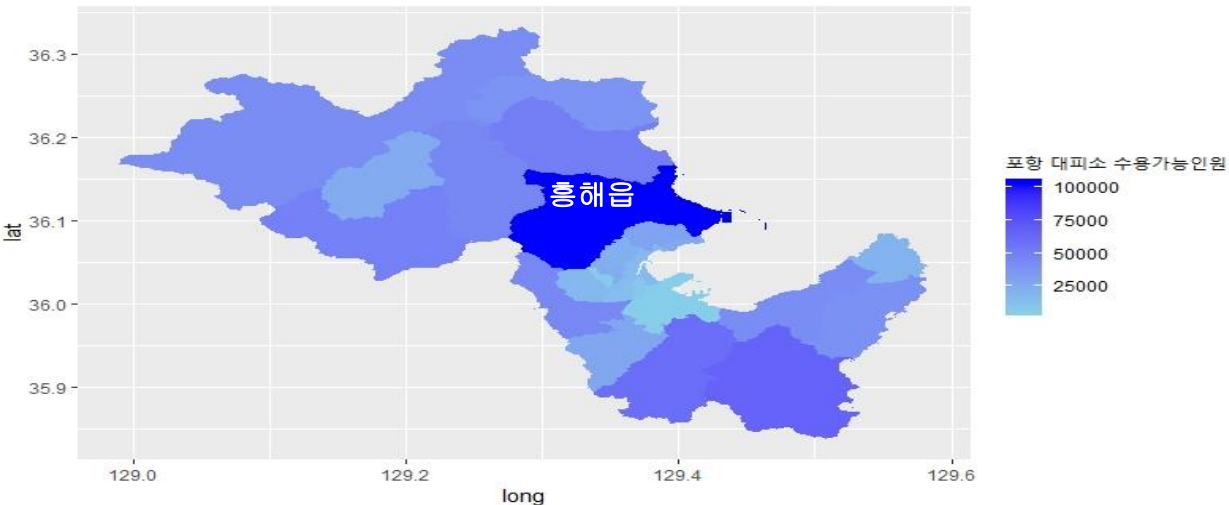
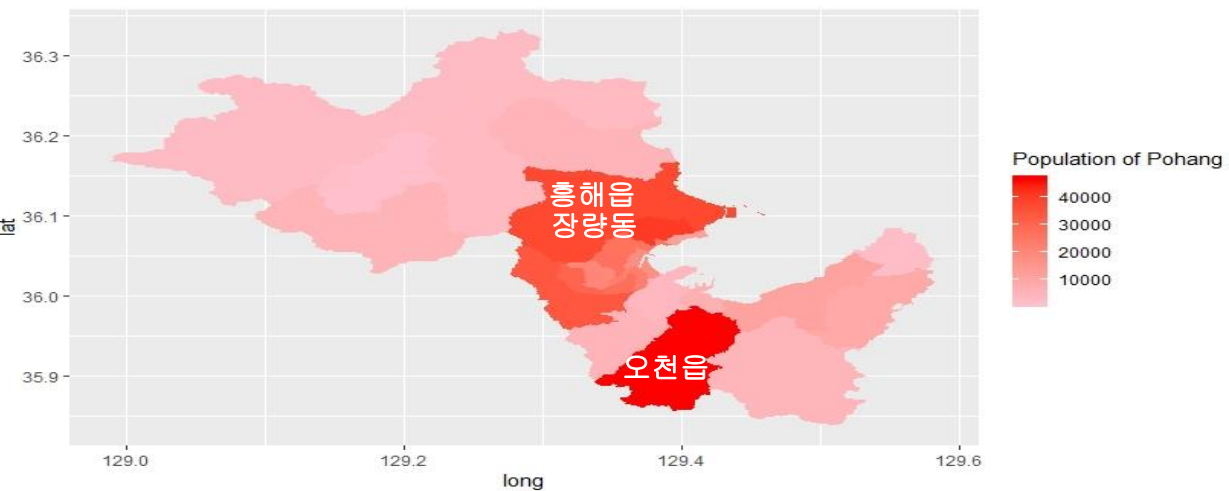
## ① 전국 시도별 인구수 및 대피소 현황



시도별	인구수 (명)	지진대피소 개수
서울특별시	9,806,538	2,070
부산광역시	3,453,198	798
대구광역시	2,468,824	893
인천광역시	2,953,841	671
광주광역시	1,461,203	182
대전광역시	1,495,029	180
울산광역시	1,158,847	420
세종특별자치시	302,703	156
경기도	12,993,165	1,933
강원도	1,544,258	359
충청북도	1,596,161	659
충청남도	2,122,531	730
전라북도	1,843,009	1,149
전라남도	2,679,546	805
경상북도	3,377,248	1,805
경상남도	3,377,483	902
제주특별자치도	664,579	89



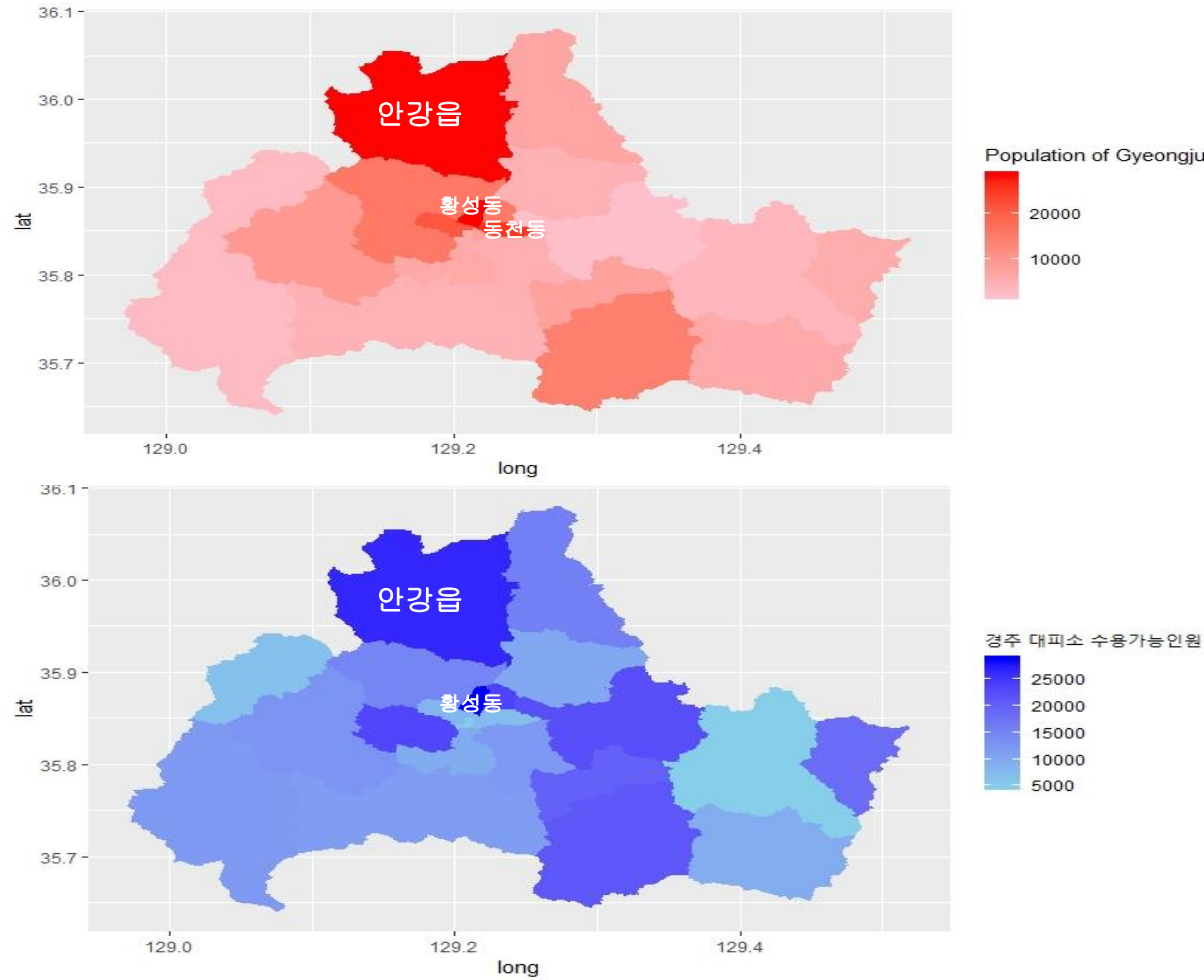
② 포항 읍면동별 인구수 및 대피소 현황



읍면동별	인구수	대피소 수용인원	읍면동별	인구수	대피소 수용인원
상대동	28,151	14,220	양학동	20,010	7,540
해도동	23,216	8,020	죽도동	23,102	16,345
송도동	18,078	16,266	용흥동	27,165	24,109
청림동	6,847	5,253	우창동	26,762	21,787
제철동	3,885	5,515	두호동	25,994	17,900
효곡동	30,818	17,943	장량동	40,725	29,357
대이동	20,823	17,136	환여동	11,951	26,492
구룡포읍	9,343	39,823	흥해읍	37,928	103,995
연일읍	34,408	44,745	신광면	3,325	45,841
오천읍	46,928	60,513	청하면	5,456	50,740
대송면	5,434	27,890	송라면	3,194	38,933
동해면	11,193	42,877	기계면	5,579	49,463
장기면	4,830	66,256	죽장면	3,003	42,640
호미곶면	2,196	20,222	기북면	1,222	25,710
중앙동	18,377	16,260			



③ 경주 읍면동별 인구수 및 대피소 현황



읍면동별	인구수	대피소 수용인원	읍면동별	인구수	대피소 수용인원
중부동	5,712	5,597	안강읍	28,499	26,689
황오동	8,391	7,424	건천읍	9,734	13,001
성건동	21,424	8,071	외동읍	14,511	21,103
황남동	6,616	9,569	양북면	3,666	4,787
월성동	5,237	12,247	양남면	6,327	9,401
선도동	15,654	23,361	내남면	4,765	11,638
황성동	28,554	28,894	산내면	2,838	12,051
용강동	15,577	23,684	서면	3,020	6,732
동천동	25,197	7,099	현곡면	15,766	14,956
불국동	8,120	19,915	강동면	7,234	15,791
보덕동	1,850	22,289	천북면	4,796	10,138
감포읍	5,913	18,367			





## 최단거리 대피소를 찾아주는 프로그램

‘약 13,000개의  
실내, 옥외 대피소 위치 데이터’와  
‘KNN(  $k=1$  ) 기법’을 활용한다면  
현 위치에서 가장 가까운  
대피소를 찾을 수 있을 것!

Step 1. 입력 받은 주소 값의 위도, 경도 값을 찾는다.

“mutate\_geocode” 함수

Step 2. 전국 실내, 옥외 대피소 데이터와 현재 위치 값을  
비교해 가장 가까운 대피소의 이름과 위치를 찾는다.

KNN(  $k=1$  ) 기법

Step 3. “ggmap” 패키지를 활용해 현 위치와  
최단거리 대피소를 표시한다 .



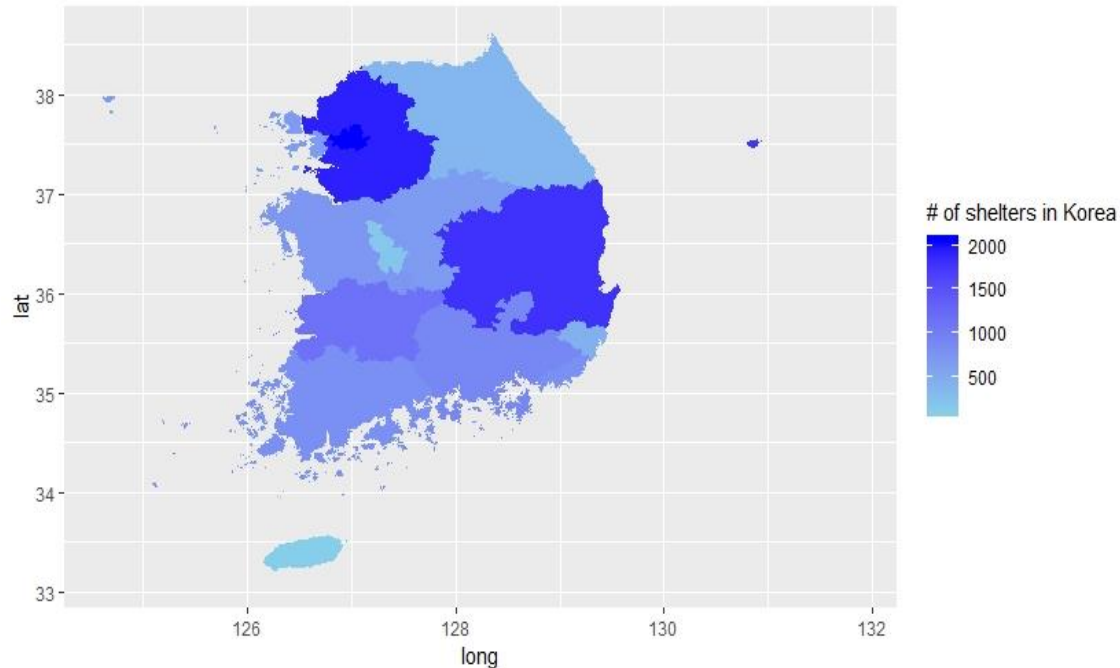
⚙️ For example

단국대학교를 입력했을 때의 결과창

가장 가까운 옥외대피소인

“현암초등학교 운동장”을 찾아주고,  
그 위치를 표시

- 분석한 결과를 토대로 대피소가 부족한 지역에 대피소를 추가할 수 있을 것으로 예상된다.
- 집에서 가까운 대피소를 미리 파악하는 등 평소에 대비하는 마음가짐이 중요할 것이다.



전국대피소현황



행정부 “안전디딤돌”

*Thank you*