

데이터과학을 위한 **R**프로그래밍

4주차. 데이터시각화



이혜선 교수

포항공과대학교 산업경영공학과



목차

4주차. 데이터시각화

1차시

R 그래픽 I (히스토그램)

2차시

R 그래픽 II (상자그림, 산점도)

3차시

R 그래픽 III (ggplot2 활용)

4차시

R 그래픽 IV (공간지도분석)



4주차

4차시

R 그래픽- 공간지도분석

R 추가패키지 설치

✓ 추가패키지 설치 (install.packages)

- maps - 세계의 지도 데이터베이스
- mapdata - maps보다 정교한 지도
- mapproj - 위도와 경도
- ggplot2 - R 그래픽

```
# lec4_4.R
# spatial analysis (2020)

#install.packages("dplyr")
library(dplyr)
# install.packages("ggplot2")
library(ggplot2)

# maps : world map
#install.packages("maps")
library(maps)

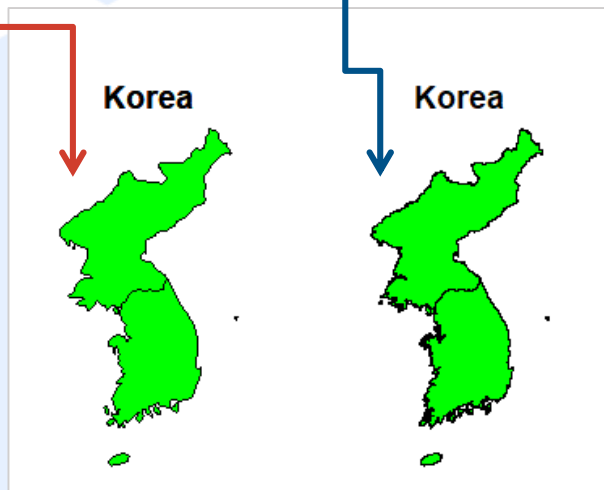
# mapdata : more world map
#install.packages("mapdata")
library(mapdata)

# mapproj : latitude and longitude
#install.packages("mapproj")|
library(mapproj)
```

● R그래픽 : 공간지도

☑ 한국지도 : maps와 mapdata를 이용한 지도

```
# using map package  
bar(mfrow = c(1, 2), mar=c(2,2,2,2))  
map(database = 'world', region = c('South Korea','North Korea'), col='green', fill = TRUE)  
title("Korea")  
# using mapdata package  
map(database = 'worldHires', region = c('South Korea','North Korea'), col='green', fill = TRUE)  
title("Korea")
```



● R그래픽 : 공간지도

☑ 지도 : 이탈리아

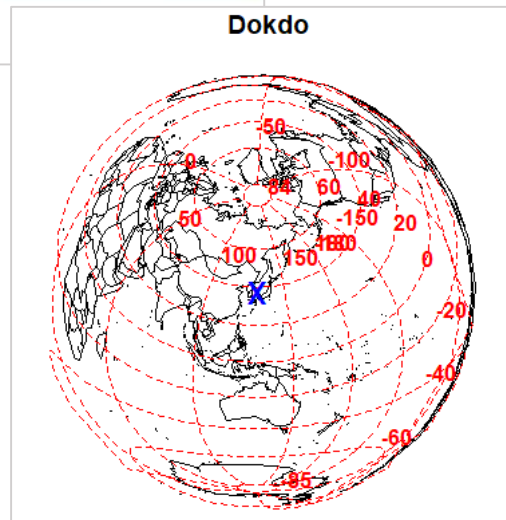
```
# 2.Italy  
par(mfrow = c(1, 1), mar=c(2,2,2,2))  
map(database = 'world', region = c('Italy'), col='coral', fill = TRUE)  
title("Italy")
```



● R그래픽 : 공간지도

☑ mapproj 패키지 - 위도, 경도 활용 (독도를 표시)

```
# Dokdo map using mapproj package
library(mapproj)
map('world', proj = 'azequalarea', orient = c(37.24223, 131.8643, 0))
map.grid(col = 2)
points(mapproject(list(y = 37.24223, x = 131.8643)), col = "blue", pch = "x", cex = 2)
title("Dokdo")
# for reading Korean : encoding to UTF-8
# file menu: Tools_global options_code_saving
```



공간지도분석 예제 : 국내공항 및 항공노선

☑ 국내 공항 및 노선 현황

```
# 4. Airport & route data (source : https://www.data.go.kr/)
airport <- read.csv("airport.csv")
route <- read.csv("route.csv")
head(airport)
head(route)
head(route[order(route$id),])
```

➤ airport.csv : 국내 공항 위치 정보

```
head(airport)
  airport iata lon lat
1   강릉  KAG 128.944 37.7536
2   광주  KWJ 126.809 35.1264
3   군산  KUV 126.616 35.9038
4   김포  GMP 126.791 37.5583
5   대구  TAE 128.659 35.8941
6   목포  MPK 126.380 34.7589
```

➤ route.csv : 국내선 노선 정보

```
  id airport lon lat
1  1   CJJ 127.499 36.7166
2  7   CJJ 127.499 36.7166
3 45   CJJ 127.499 36.7166
4 77   CJJ 127.499 36.7166
5  2   CJJ 127.499 36.7166
6  8   CJJ 127.499 36.7166
```

```
  id airport lon lat
1  1   CJJ 127.499 36.7166
83 1   CJU 126.493 33.5113
2  2   CJU 126.493 33.5113
84 2   CJJ 127.499 36.7166
3  3   CJU 126.493 33.5113
85 3   GMP 126.791 37.5583
```

id기준 정렬

항공노선

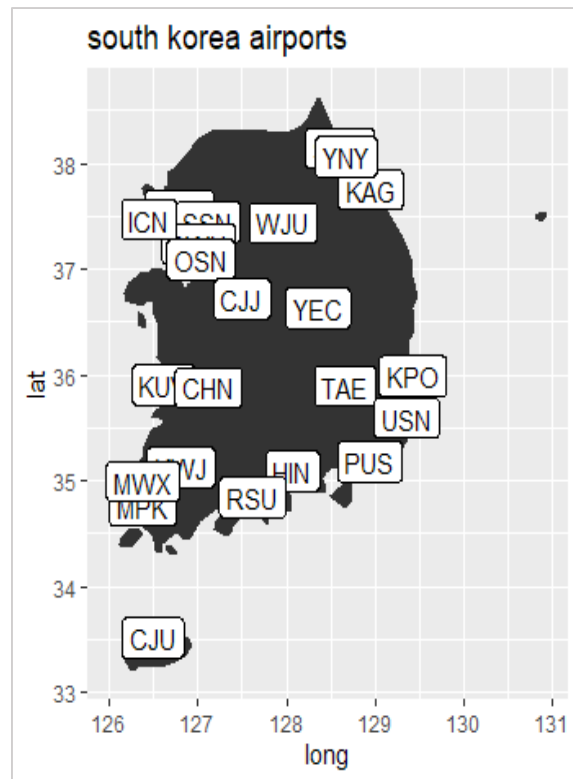
head(route[order(route\$id),])

공간지도분석 예제 1 : 국내공항 및 항공노선

☑ 국내 공항위치

```
# Korea map (kr.map)
world.map <- map_data("world")
kr.map <- world.map %>% filter(region == "South Korea")

# ----- #
# Korea map using ggplot
# ----- #
# 5. Domestic airport location
ggplot() +
  geom_polygon(data=kr.map, aes(x=long, y=lat, group=group)) +
  geom_label(data=airport, aes(x = lon, y = lat, label=iata)) +
  labs(title = "south korea airports")
```



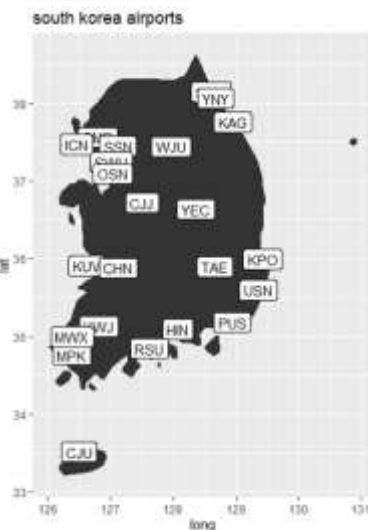
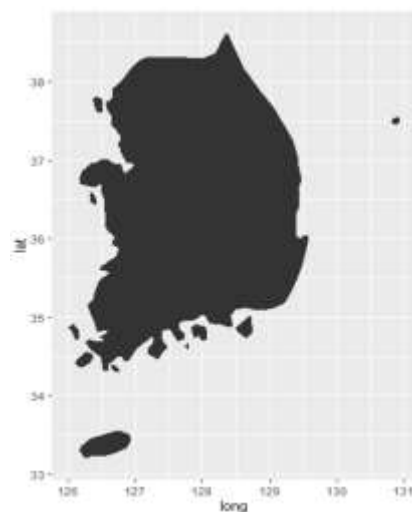
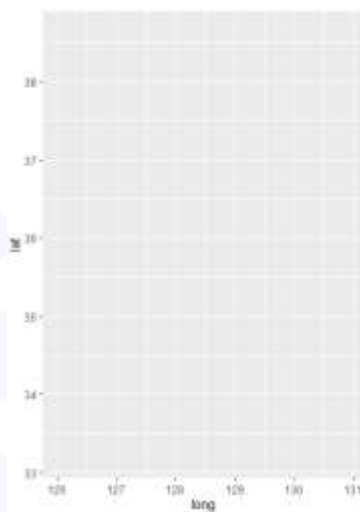
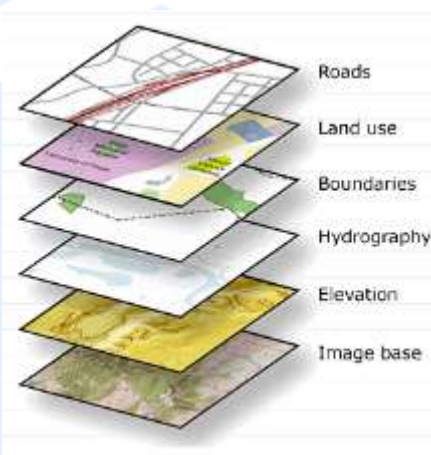
● 공간지도분석 예제 1 : 국내공항 및 항공노선

☑ ggplot 은 레이어를 추가하는 방식으로 그래픽을 구현

2 `ggplot(data=kr.map, aes(x=long, y=lat, group=group)) +
geom_polygon(fill="white", colour="black")`

3 `ggplot() +
geom_polygon(data=kr.map, aes(x=long, y=lat, group=group)) +
geom_label(data=airport, aes(x = lon, y = lat, label=iata)) +
labs(title = "south korea airports")`

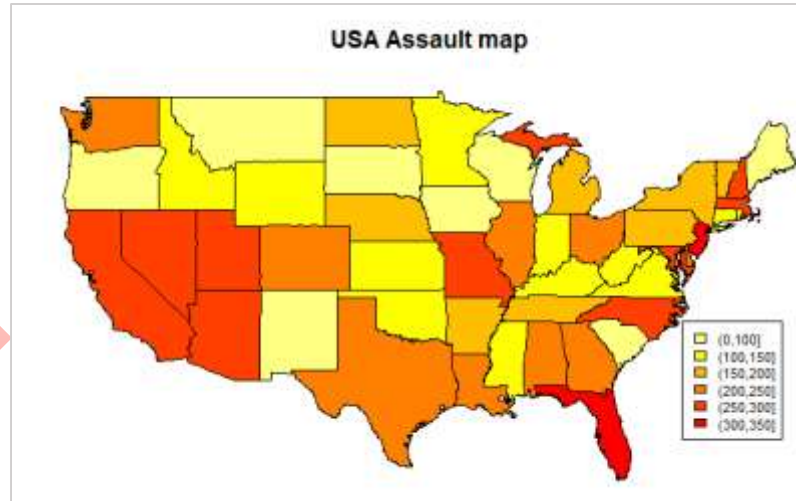
1 `ggplot(data=kr.map, aes(x=long, y=lat, group=group))`



● 공간지도분석 예제 2 : 미국 행정데이터

☑ 지도 데이터베이스와 행정자료 결합 : 미국 (1973) 범죄수 지도

```
# Assault in US (1973)
par(mfrow = c(1, 1))
library(maps)
# excluding Alaska, Hawaii
sub.usa <- subset(USArrests, !rownames(USArrests) %in% c("Alaska", "Hawaii"))
# data with State name, Assault count
usa.data <- data.frame(states = rownames(sub.usa), Assault = sub.usa$Assault)
# legend
col.level <- cut(sub.usa[, 2], c(0, 100, 150, 200, 250, 300, 350))
legends <- levels(col.level)
# displaying color for the size
levels(col.level) <- sort(heat.colors(6), decreasing = TRUE)
usa.data <- data.frame(usa.data, col.level = col.level)
# Map
map('state', region = usa.data$states, fill = TRUE, col = as.character(usa.data$col.level),
    title = "USA Assault map")
legend(-76, 35, legends, fill = sort(heat.colors(6), decreasing = TRUE), cex = 0.5)
```



공간지도분석 예제 2 : 미국 행정데이터

✓ 행정자료 (USArrests 데이터)

```
help(USArrests)  
head(USArrests)
```

Description

This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

Usage

USArrests

Format

A data frame with 50 observations on 4 variables.

- [,1] Murder numeric Murder arrests (per 100,000)
- [,2] Assault numeric Assault arrests (per 100,000)
- [,3] UrbanPop numeric Percent urban population
- [,4] Rape numeric Rape arrests (per 100,000)

```
> head(USArrests)
```

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7