

7강 시계열의 시각화 (2)

한국방송통신대학교 정보통계학과 이금희 교수

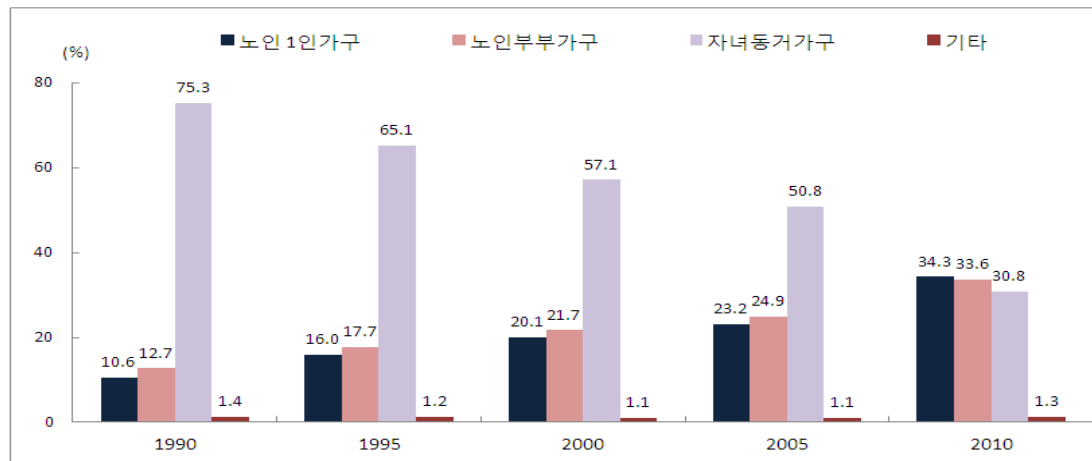
1. 막대 그래프의 작성
2. 누적 그래프의 작성
3. 경로 그래프의 작성
4. 채색달력 그래프의 작성

1. 막대 그래프의 작성

- 막대 그래프의 의미
- 막대 그래프의 작성

▶ 막대그래프의 의미

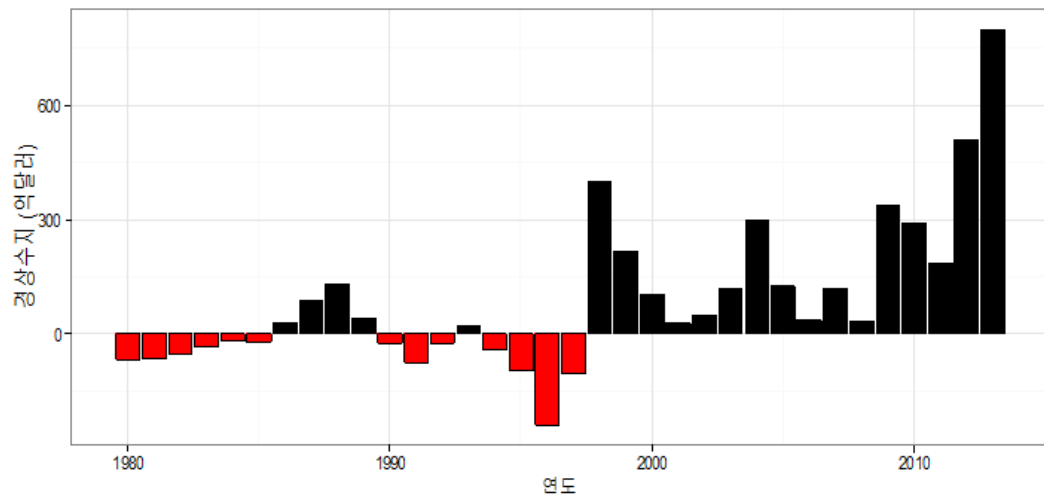
- 시간에 따른 크기 변화를 표현
- 노인가구 형태별 구성비율



출처 : 통계청(2013) 한국의 사회동향 2013

▶ 막대그래프의 작성

■ 우리나라 연도별 경상수지의 추이



▶ 막대그래프의 작성

```
library(ggplot2)
cb <- read.csv("cb.csv", header=TRUE)
cb$경상수지 <- cb$경상수지/100
cb$pos <- cb$경상수지 >=0
ggplot(cb, aes(x=연도, y=경상수지, fill=pos))
  + geom_bar(stat="identity", position="identity",
             colour="black", size=0.25)
  + scale_fill_manual(values=c("red", "black"),
                      guide=FALSE)
  + ylab("경상수지 (억달러)") + theme_bw()
```

2. 누적 그래프의 작성

- 누적 그래프의 의미
- 누적 그래프의 작성

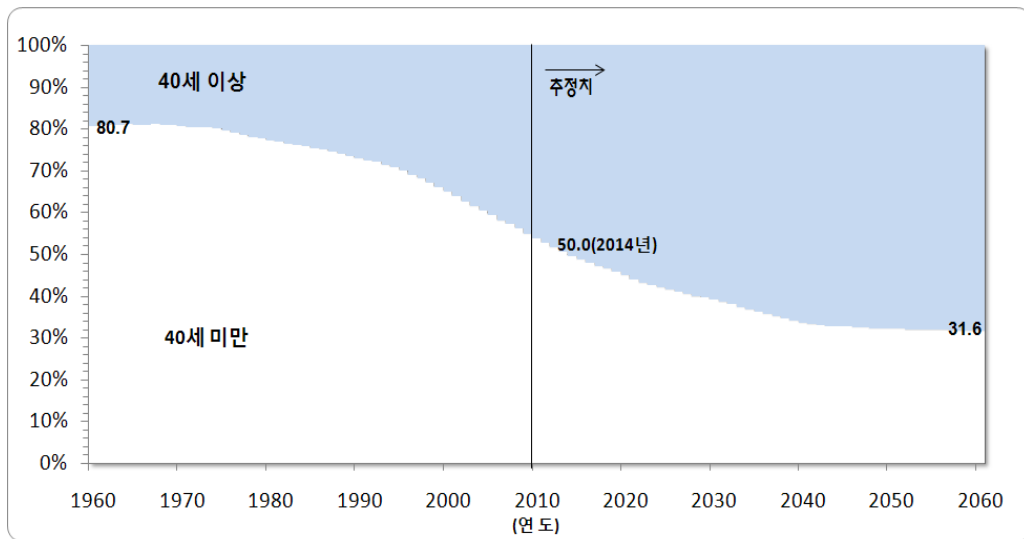


▶ 누적 그래프의 의미

- 누적 그래프 : 전체 시계열이 하부 시계열의 합으로 구성될 때
구성의 추이를 살펴보는데 이용되는 그래프
- 누적 그래프의 형태 : 면적 그래프, 막대 그래프

▶ 누적 그래프의 의미

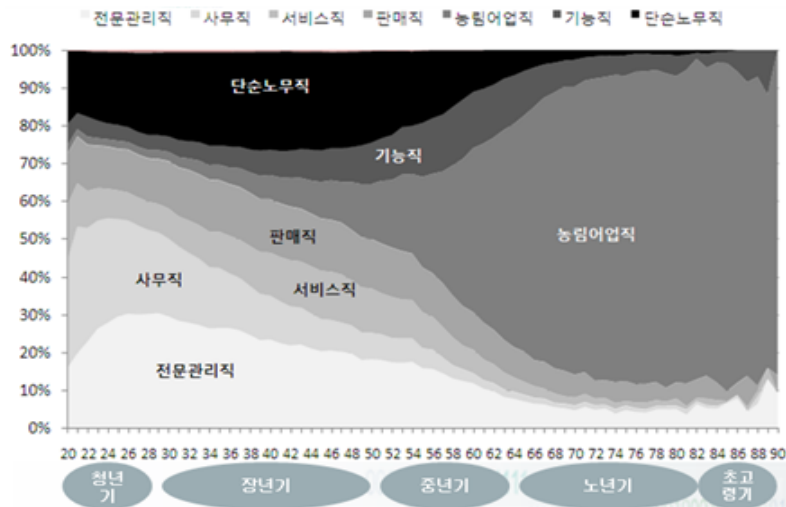
■ 우리나라 인구구성 추이



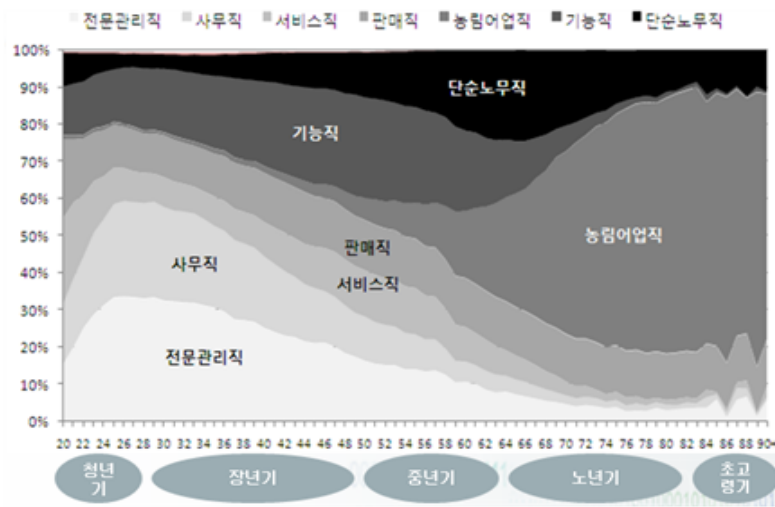
출처 : 통계청(2013) 2013 한국의 사회지표

▶ 누적그래프의 의미

< 2000년 >



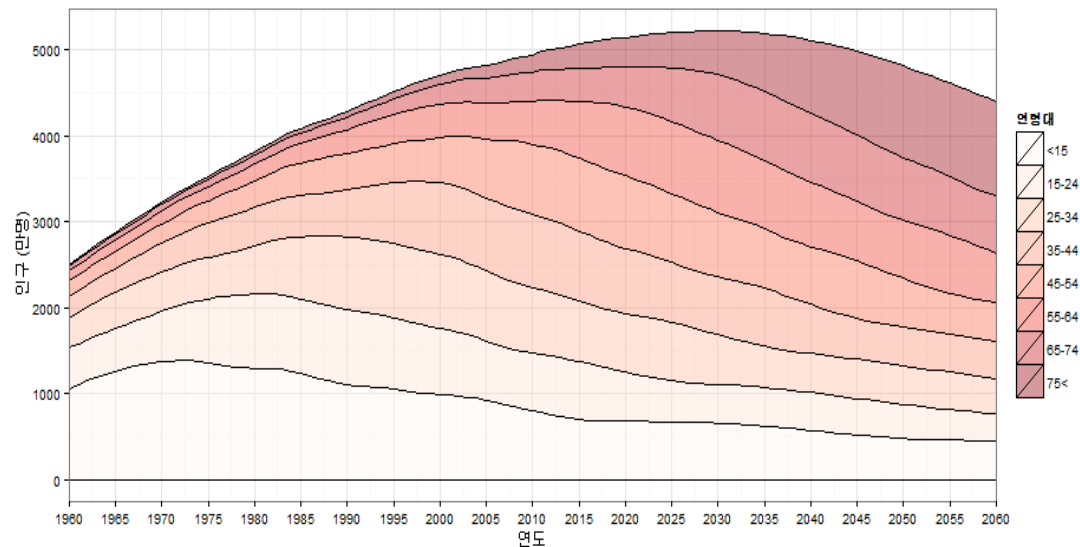
<2010년 >



출처 : 통계청(2014) 생애주기별 주요 특성 및 변화 분석

▶ 누적 그래프의 작성

■ 연령대별 우리나라 인구 추이



▶ 누적 그래프의 작성

```
library(ggplot2)
library(reshape2)
library(plyr)
# 데이터 불러오기
pop_kr1 <- read.csv("krpop.csv", header=TRUE)
# 데이터 재정렬
pop_kr <- melt(pop_kr1, id="연령대")
pop_kr$연도 <- as.numeric(substr(pop_kr$variable, 2, 5))
```

▶ 누적 그래프의 작성

```
> head(pop_kr1)
```

	연령대	X1960	X1961	X1962	X1963	X1964
1	<15	10587583	11065019	11472623	11850374	12215638
2	15-24	4741149	4750164	4803558	4872444	4930487
3	25-34	3489408	3638222	3785141	3936156	4058520
4	35-44	2495159	2531678	2585561	2656237	2746875
5	45-54	1825186	1863263	1904722	1927450	1939819
6	55-64	1147439	1166446	1178709	1202918	1247681

		X1965	X1966	X1967	X1968	X1969	X1970
1		12577740	12916059	13199493	13459383	13638232	13709367
2		4983871	5042592	5150344	5300639	5518950	5838400
3		4176545	4313326	4432958	4507526	4557803	4573716
4		2822132	2886676	2969920	3068255	3198132	3334111
5		1976105	2032004	2094918	2162172	2217509	2288704
6		1287198	1318280	1355426	1407390	1465097	1505221

```
> head(pop_kr)
```

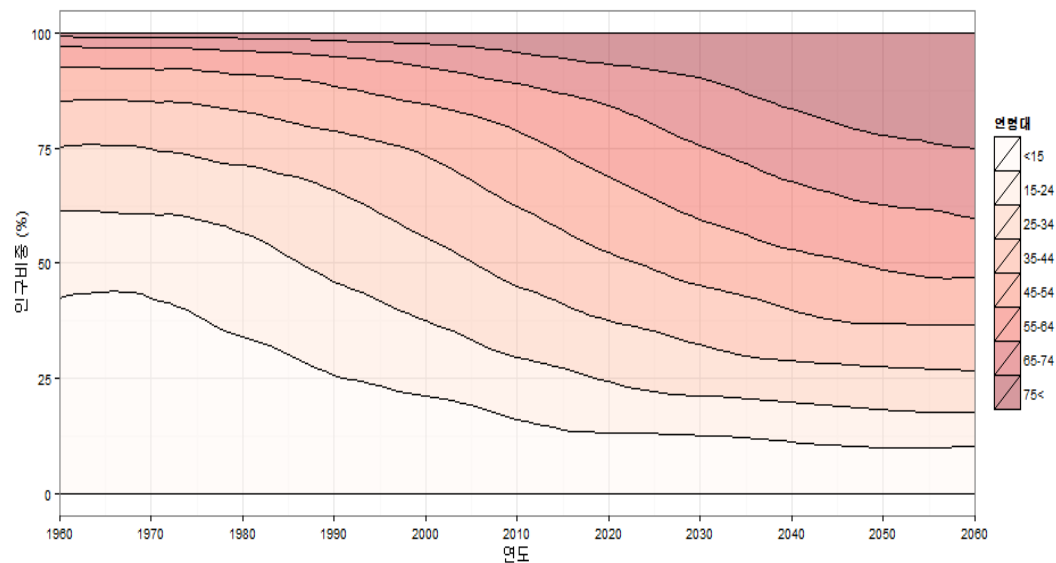
	연령대	variable	value	연도
1	<15	X1960	10587583	1960
2	15-24	X1960	4741149	1960
3	25-34	X1960	3489408	1960
4	35-44	X1960	2495159	1960
5	45-54	X1960	1825186	1960
6	55-64	X1960	1147439	1960

▶ 누적 그래프의 작성

연령대별 인구그래프 그리기

```
ggplot(pop_kr, aes(x=연도, y=value/10000, fill=연령대))  
  + geom_area(colour="black", size=.1, alpha=.4)  
  + scale_fill_brewer(palette="Reds")  
  + ylab("인구 (만명)")  
  + scale_x_continuous(breaks=seq(1960, 2060, 5),  
    expand=c(0,0))  
  + theme_bw()
```

연령대별 우리나라 비중 차이

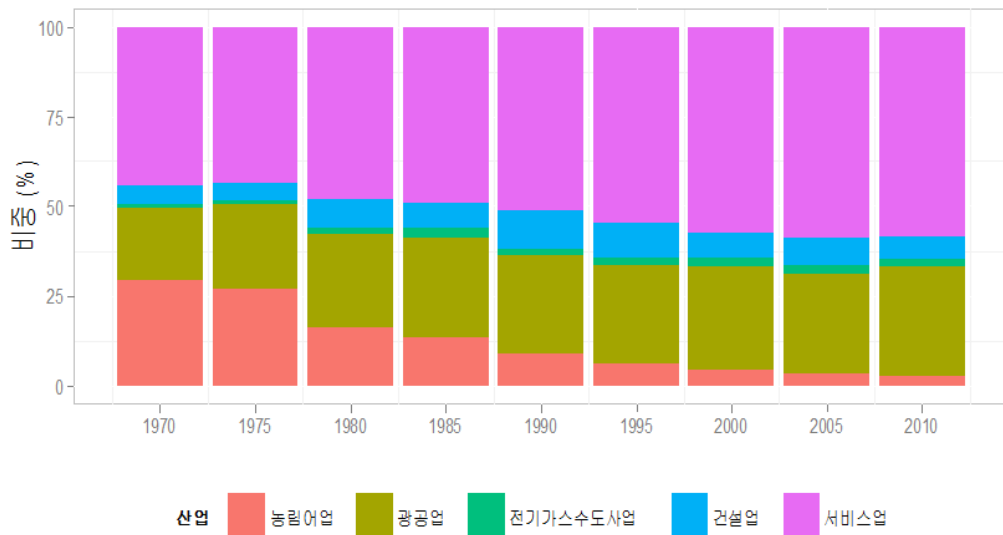


▶ 누적 그래프의 작성

```
pop_kr_p <- ddply(pop_kr, "연도", transform, 비중  
  = value/sum(value)*100)  
  
ggplot(pop_kr_p, aes(x=연도, y=비중, fill=연령대))  
  + geom_area(colour="black", size=.1, alpha=.4)  
  + scale_fill_brewer(palette="Reds")+ylab("인구비중  
    (%)")  
  + scale_x_continuous(breaks=seq(1960, 2060, 10),  
    expand=c(0,0))  
  + theme_bw()
```

▶ 누적 그래프의 작성

■ 우리나라 산업구조



▶ 누적 그래프의 작성

```
library(ggplot2)
library(reshape2)
gdp_s1 <- read.csv("gdp_sh.csv", header=TRUE)
gdp_s  <- melt(gdp_s1, id="연도")
names(gdp_s) <- c("연도", "산업", "비중")
```

▶ 누적 그래프의 작성

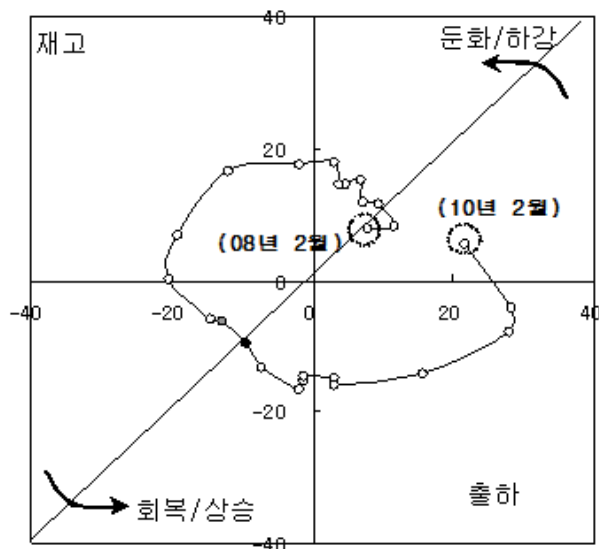
```
ggplot(gdp_s, aes(x=연도, y=비중, fill=산업))  
+ geom_bar(stat="identity")  
+ scale_x_continuous(breaks=seq(1970, 2010, 5))  
+ theme(panel.background = element_rect(fill = "white",  
      colour = "gray"), legend.position="bottom")  
+ ylab("비중 (%)")  
+ xlab("")
```

3. 경로 그래프의 작성

- 경로 그래프의 의미
- 경로 그래프의 작성

▶ 경로 그래프의 의미

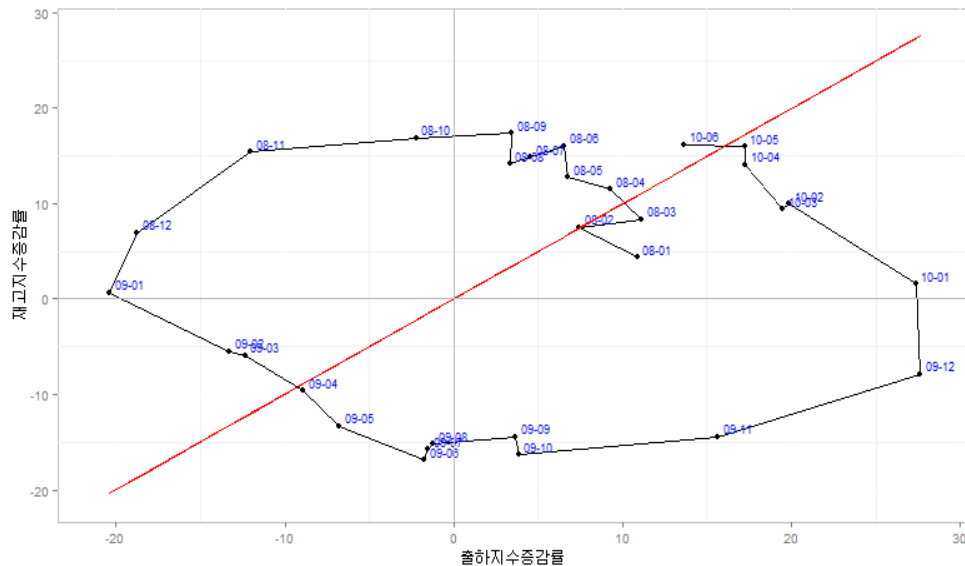
■ 좌표 점들을 시간에 따라서 연결해가는 그래프



출처 : 통계청(2010) 2010년 2월 산업활동동향

▶ 경로 그래프의 작성

■ 우리나라 제조업 재고출하순환도



▶ 경로 그래프의 작성

```
library(ggplot2)
library(scales)
library(xts)
inven1 <- read.csv("inven_cy.csv",header=TRUE)
연도 <- seq(as.Date("1980-01-01"), as.Date("2014-06-01"),
           "month")
inven <- xts(inven1[,2:3],연도)
inven$출하지수증감률 <- (inven$출하지수 - lag(inven$출하지수,12))
                        /lag(inven$출하지수,12)*100
inven$재고지수증감률 <- (inven$재고지수 - lag(inven$재고지수,12))
                        /lag(inven$재고지수,12)*100
inven_1 <- inven[337:366]
```

▶ 경로 그래프의 작성

```
ggplot(inven_1, aes(x=출하지수증감률, y=재고지수증감률))  
  + theme_bw() + geom_path()  
  + geom_point() + ylim(-21,28)+xlim(-21,28)  
  + geom_text(aes(label=substr(index(inven_1),3,7)),  
              size=3, hjust=-0.2, vjust=-0.3, colour="blue")  
  + geom_line(x=inven_1$출하지수증감률,y=inven_1$출하지수  
              증감률, colour="red")  
  + geom_hline(yintercept=0, colour="gray")  
  + geom_vline(xintercept=0, colour="gray")
```

4. 채색달력 그래프의 작성

- 채색달력 그래프의 의미
- 채색달력 그래프의 작성

▶ 채색달력 그래프의 의미

■ 일별 데이터 값에 색을 대응하여 달력에 표현한 그래프

● 일별 평균 대기오염도

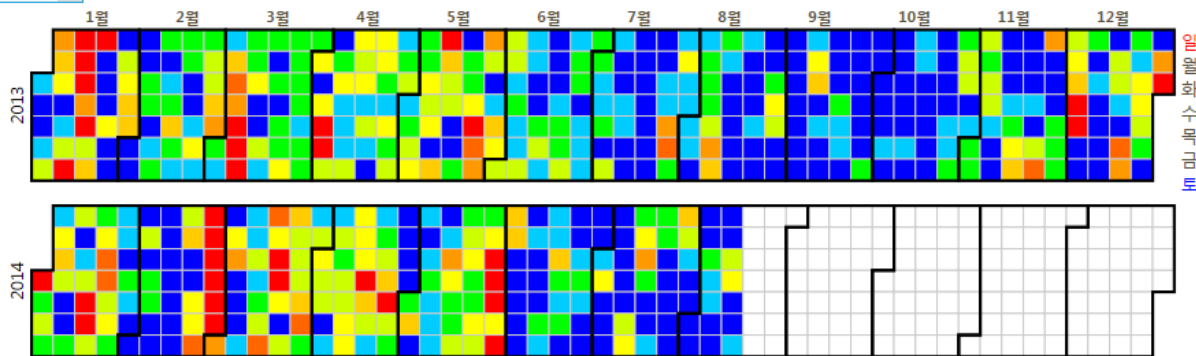
미세먼지 기준치: 24시간 평균 $100 \mu\text{g}/\text{m}^3$ 이하

2014 ▼

미세먼지 ▼

데이터 조회

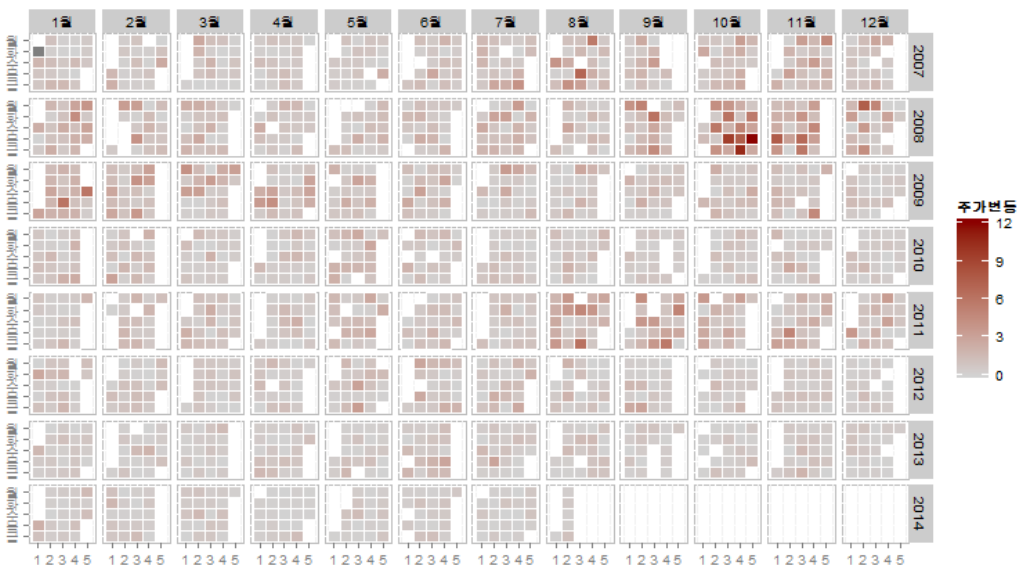
서울역 ▼



4 채색달력 그래프의 작성

▶ 채색달력 그래프의 작성

■ 우리나라 주가변동 달력 그래프



▶ 채색달력 그래프의 작성

```
library(quantmod)
library(ggplot2)
library(reshape2)
library(plyr)
library(scales)
# 야후 파이낸스 데이터베이스에서 가져오기
getSymbols("^KS11",src="yahoo")
KS11$주가변동 <- abs((KS11$KS11.Close -
  lag(KS11$KS11.Close,1)) /lag(KS11$KS11.Close,1)) *100
```

4 채색달력 그래프의 작성

▶ 채색달력 그래프의 작성

```
dat <- data.frame(date=index(KS11),KS11)
dat$year <- as.numeric(as.POSIXlt(dat$date)$year+1900)
dat$month <- as.numeric(as.POSIXlt(dat$date)$mon+1)
dat$monthf <- factor(dat$month,levels=as.character(1:12),
  labels=c("1월","2월","3월","4월","5월","6월",
    "7월","8월","9월","10월","11월","12월"),
  ordered=TRUE)
```

▶ 채색달력 그래프의 작성

```
dat$weekday = as.POSIXlt(dat$date)$wday
dat$weekdayf <-
  factor(dat$weekday, levels=rev(1:7),
    labels=rev(c("월", "화", "수", "목", "금", "토", "일")),
    ordered=TRUE)
dat$yearmonth <- as.yearmon(dat$date)
dat$yearmonthf <- factor(dat$yearmonth)
dat$week <- as.numeric(format(dat$date, "%W"))
dat <- ddply(dat, .(yearmonthf), transform,
  monthweek=1+week-min(week))
```

4 채색달력 그래프의 작성

▶ 채색달력 그래프의 작성

```
ggplot(dat, aes(monthweek, weekdayf, fill = 주가변동))  
  + geom_tile(colour = "white")  
  + facet_grid(year~monthf)  
  + scale_fill_gradient(limits=c(0, 12), low="lightgray",  
    high="darkred")  
  + xlab("") + ylab("")  
  + theme(panel.background = element_rect(fill = "white",  
    colour = "gray"))
```



다음시간안내

지리적 데이터의 시각화 1