

6강 시계열의 시각화 (1)

한국방송통신대학교 정보통계학과 이금희 교수

1. 시계열의 개요
2. 시계열과 R
3. 선 그래프의 작성

1. 시계열의 개요

- 시계열의 정의
- 시계열 도표의 필요성
- 시계열 도표의 예

1 시계열의 개요

▶ 시계열

- 시간에 따라 측정된 데이터
- 월별, 분기별, 연도별로 수집된 국가통계가 대표적인 시계열
 - 경제 시계열 : 경제총생산(GDP), 소비자 물가지수 등
 - 물리 시계열 : 일일 강수량, 기온, 지진 발생 건수 등
 - 경영 시계열 : 상품의 판매량, 매출액
 - 인구 시계열 : 총인구, 농가수
 - 사회 시계열 : 교통사고 건수, 범죄발생 수

1 시계열의 개요

▶ 우리나라 총 인구와 인구성장률 추이

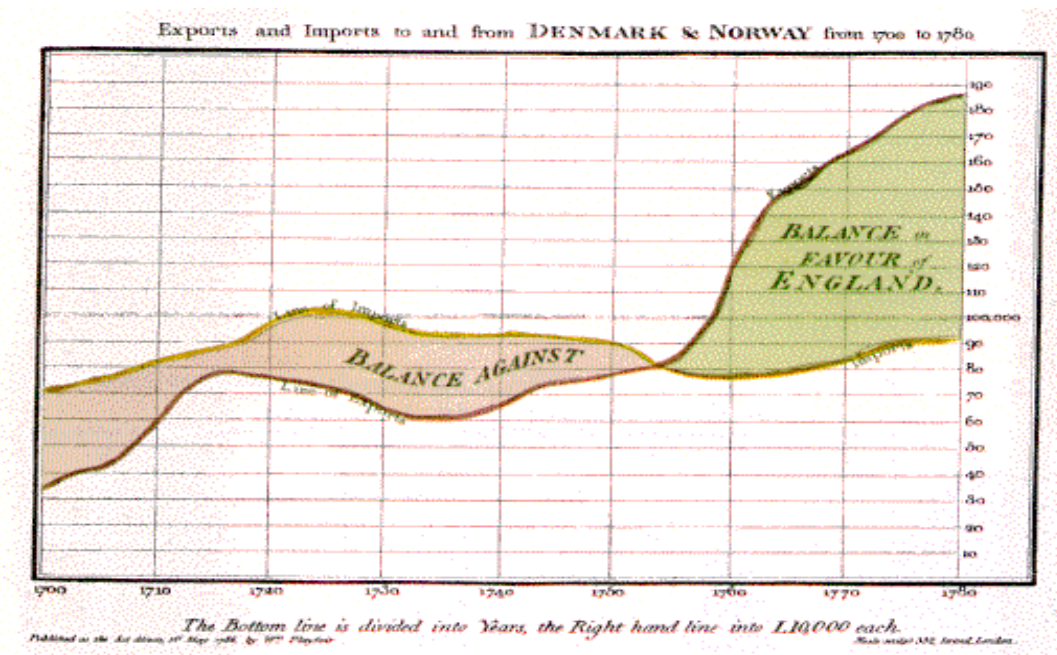
우리나라 총인구와 인구성장률 추이

연도	총인구	인구성장률
1990	42,869	0.99
1995	45,093	1.01
2000	47,008	0.84
2005	48,138	0.21
2010	49,410	0.46
2012	50,004	0.45
2013	50,220	0.43
2030	52,160	0.01
2040	51,091	-0.39

출처 : 통계청, 「장래인구추계」 2011.12., 2013 한국의 사회지표

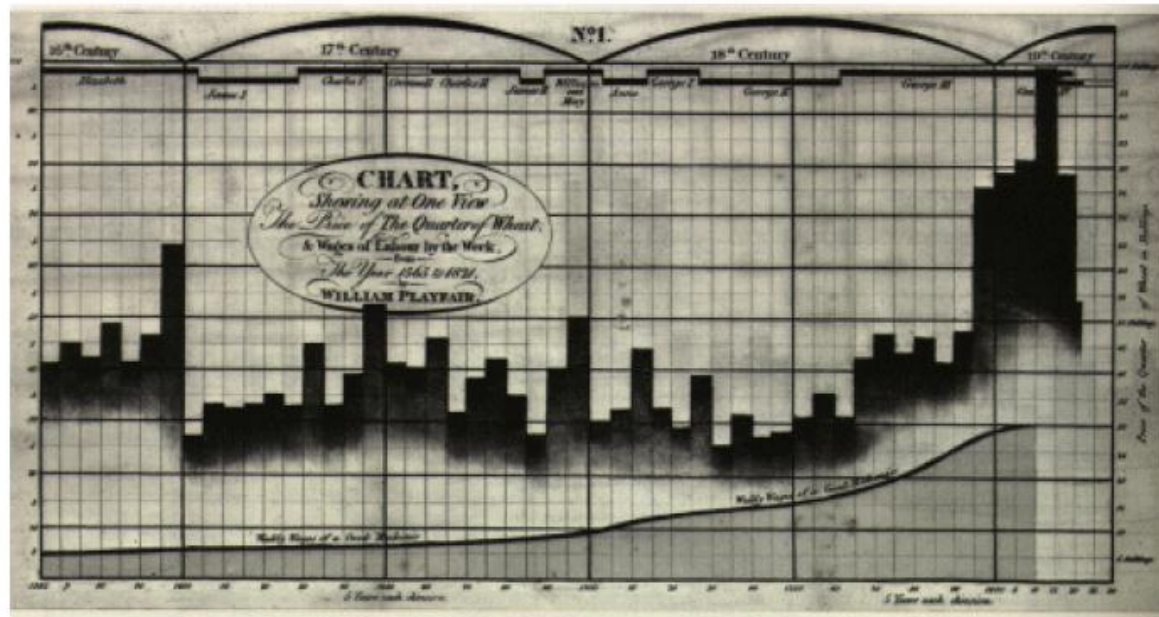
1 시계열의 개요

▶ playfair의 그래프



출처 : Playfair (1821), <http://www.datavis.ca/gallery/missed.php>

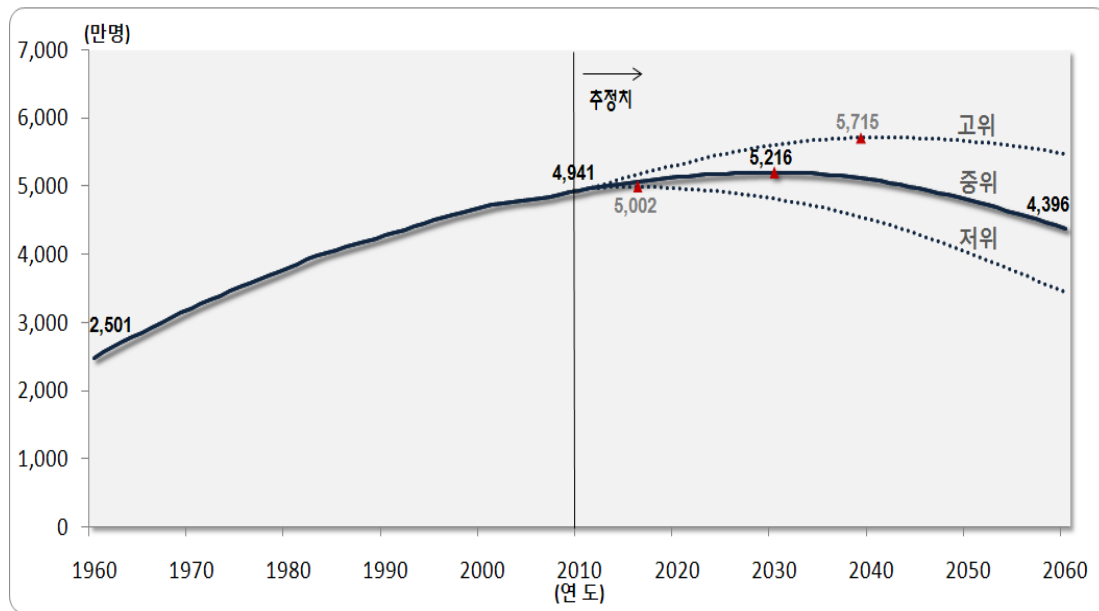
▶ playfair의 그래프



출처 : Playfair (1821), <http://www.datavis.ca/gallery/missed.ph>

1 시계열의 개요

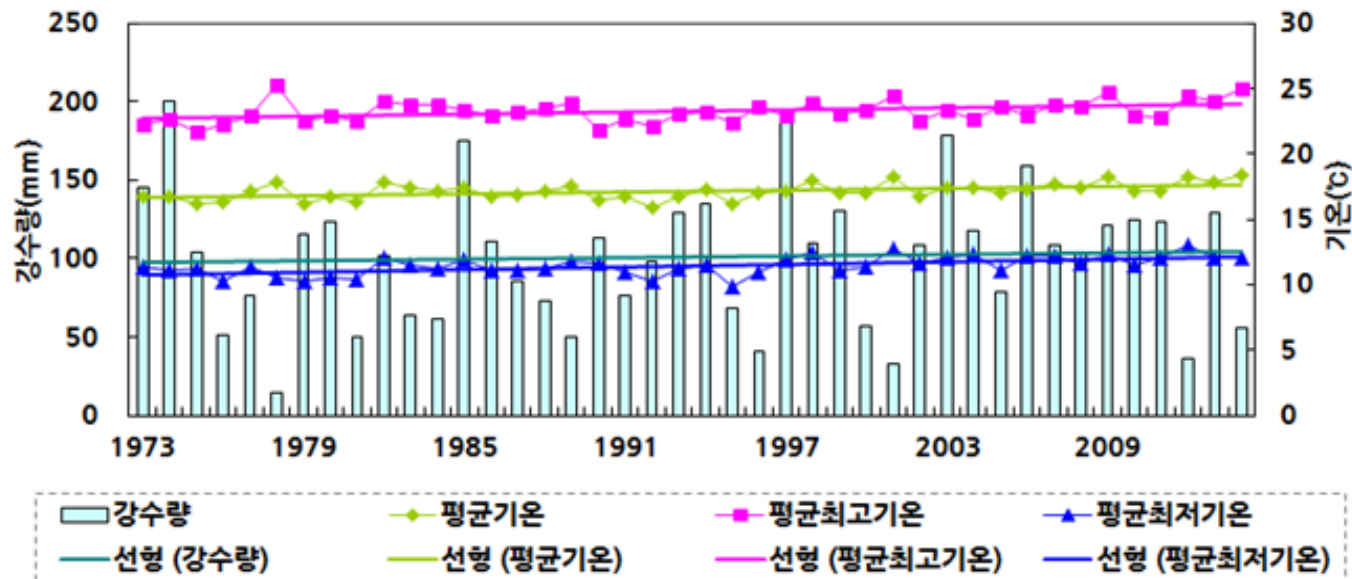
▶ <예 4-1> 우리나라 인구추계



출처 : 통계청(2012) 장래인구추계

1 시계열의 개요

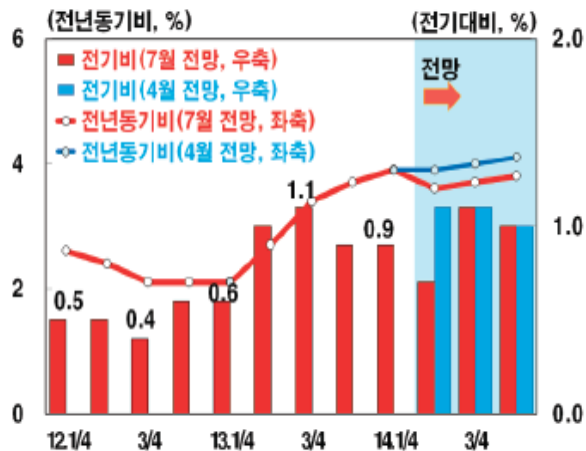
▶ <예 4-2> 전국 5월 기상자료의 특성



1 시계열의 개요

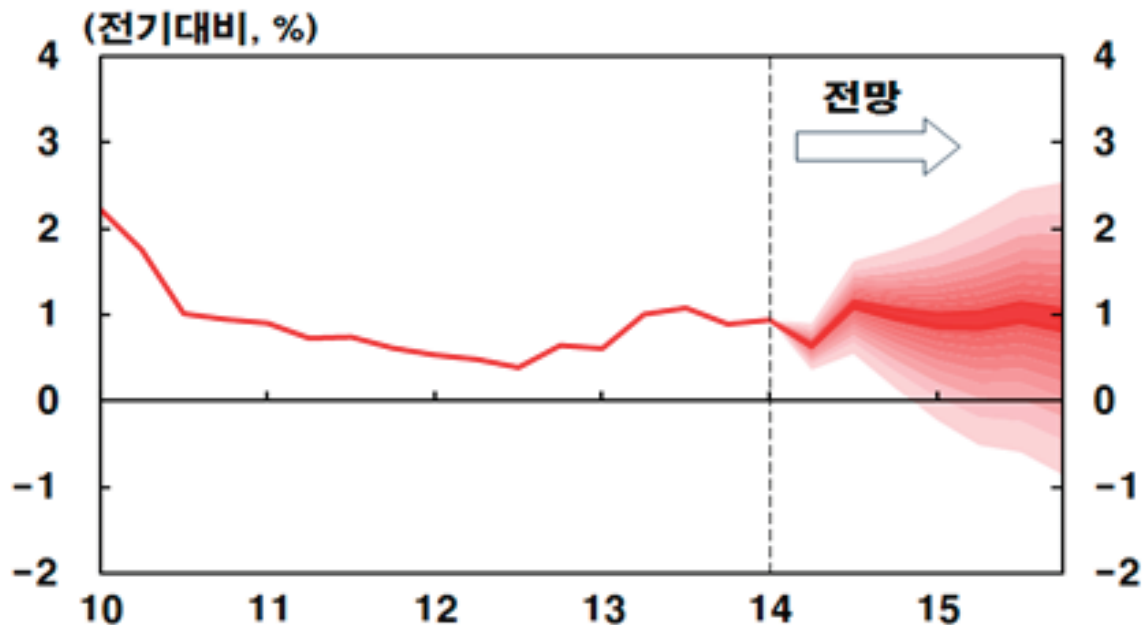
▶ <예 4-3> 한국은행의 경제전망

GDP 전망경로



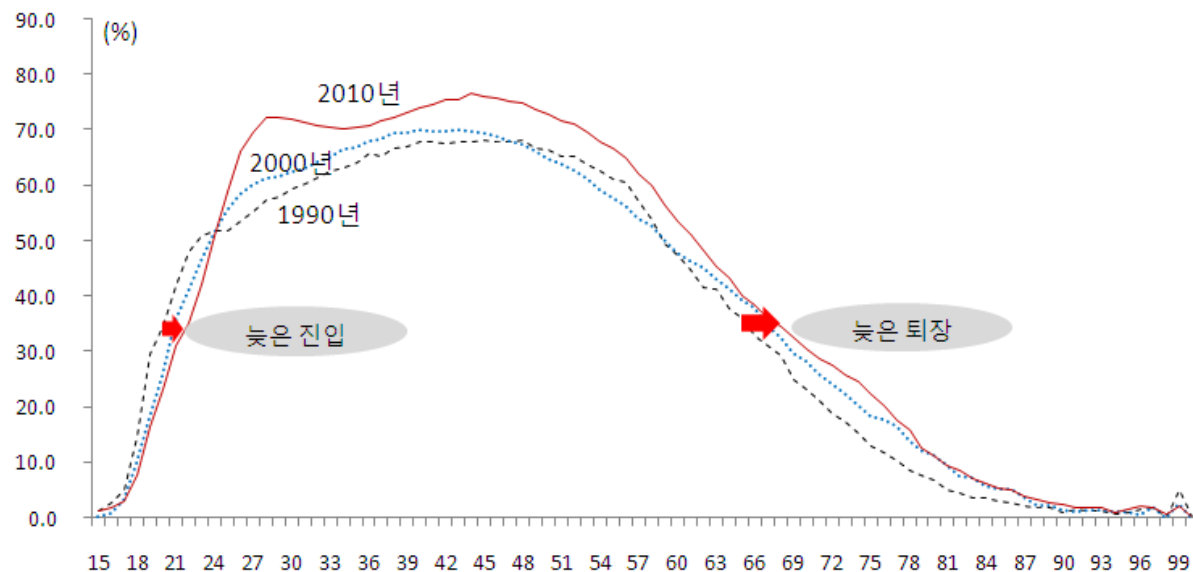
1 시계열의 개요

▶ 〈예 4-3〉 한국은행의 경제전망



1 시계열의 개요

▶ <예 4-4> 우리나라 취업인구 비중



2. 시계열과 R

- 시계열과 R
- ggplot2 패키지

▶ R과 시계열 객체

- stats 패키지 : ts 함수
- zoo 패키지
- xts 패키지

▶ R시계열 객체의 생성

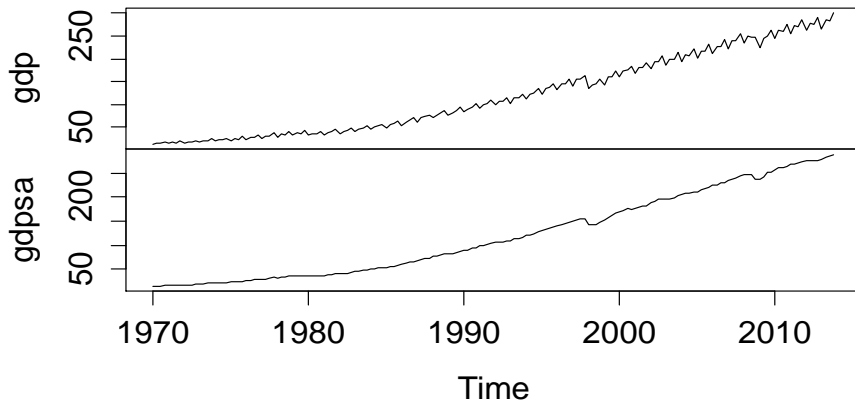
■ GDP 원계열과 계절조정계열의 그래프 작성

```
econ1 =  
  read.csv("i:/R/datavis/gdp.csv", header=TRUE)  
econ.ts = ts(econ1, start=1970, freq=4)  
plot(econ.ts/1000, main="")
```

2 시계열과 R

▶ R시계열 객체의 생성

■ 우리나라 GDP 원계열과 계절조정계열의 차이



▶ R시계열 객체의 생성

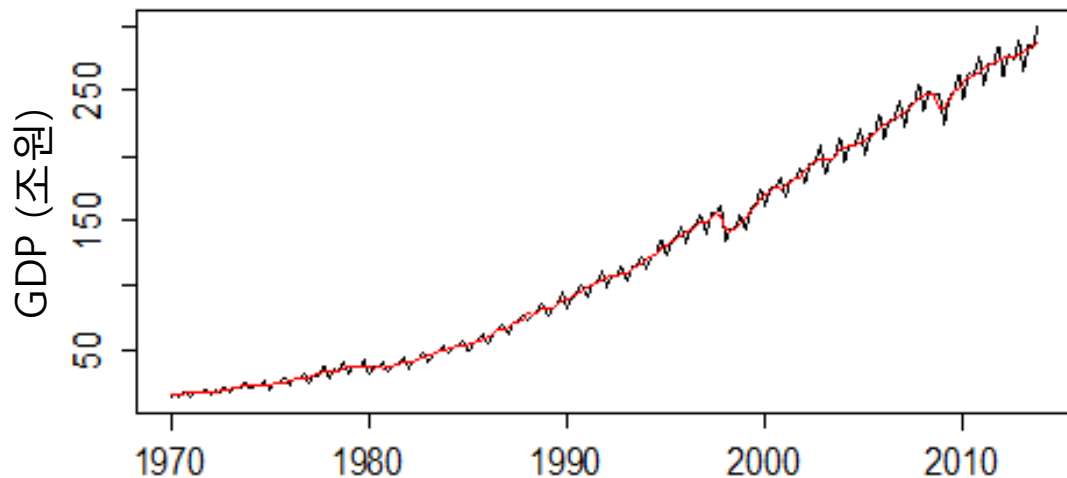
■ GDP 원계열과 계절조정계열의 그래프 작성

```
setwd("c:/work/data/")
library(zoo)
econ1 = read.csv("gdp.csv", header=TRUE)
연도 = seq(as.Date("1970-01-01"),
           as.Date("2013-12-01"), "quarter")
econ = zoo(econ1, 연도)
plot(econ/1000, ylab="GDP(조원)", xlab="",
      col=1:2, screens=1)
```


2 시계열과 R

▶ R시계열 객체의 생성

■ 우리나라 GDP 원계열과 계절조정계열의 차이



▶ ggplot2 패키지

- Hadley Wickham : 그래프 문법을 바탕으로 패키지 개발

Grammar of Graphics (Wilkinson, 2005)

- `install.packages("ggplot2")`

`library(ggplot2)`

▶ ggplot2 패키지

■ 그래프 작성방법

- ① qplot
- ② ggplot+geom_xxx
- ③ ggplot+layer

▶ ggplot2 패키지

■ diamond data

```
small<-diamonds[sample(nrow(diamonds),1000),]  
head(small)
```

	carat	cut	color	clarity	depth	table	price	x	y	z
50481	0.70	Premium	H	SI1	62.2	57	2268	5.74	5.71	3.56
35586	0.35	Ideal	D	VS2	60.5	56	907	4.63	4.60	2.79
46585	0.55	Ideal	G	VS1	60.9	55	1786	5.29	5.32	3.23
31286	0.32	Ideal	G	SI2	62.2	53	450	4.40	4.44	2.75
52548	0.60	Ideal	G	IF	62.2	55	2533	5.36	5.44	3.36
32233	0.31	Ideal	G	VVS1	61.6	57	789	4.34	4.36	2.68

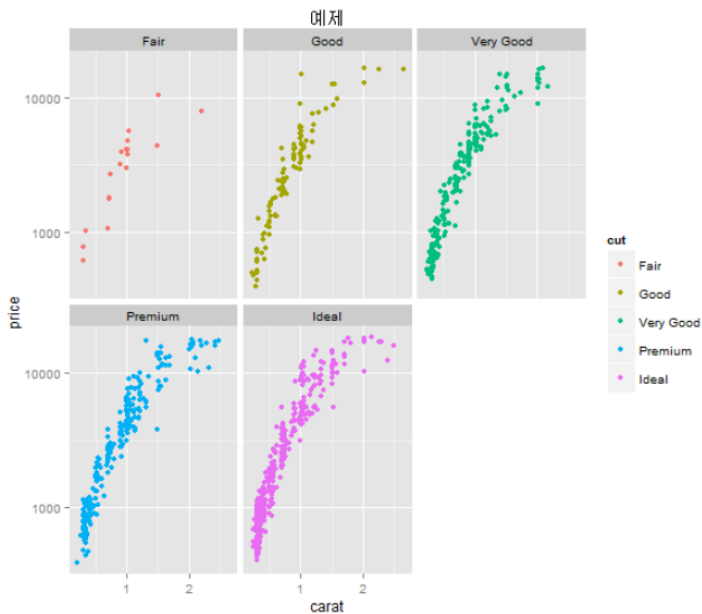
▶ ggplot2 패키지

■ diamond data

```
ggplot (small)  
+ geom_point (aes (x=carat, y=price, colour=cut))  
+ scale_y_log10()  
+ facet_wrap (~cut)  
+ ggtitle ("예제")
```

▶ ggplot2 패키지

■ diamond data

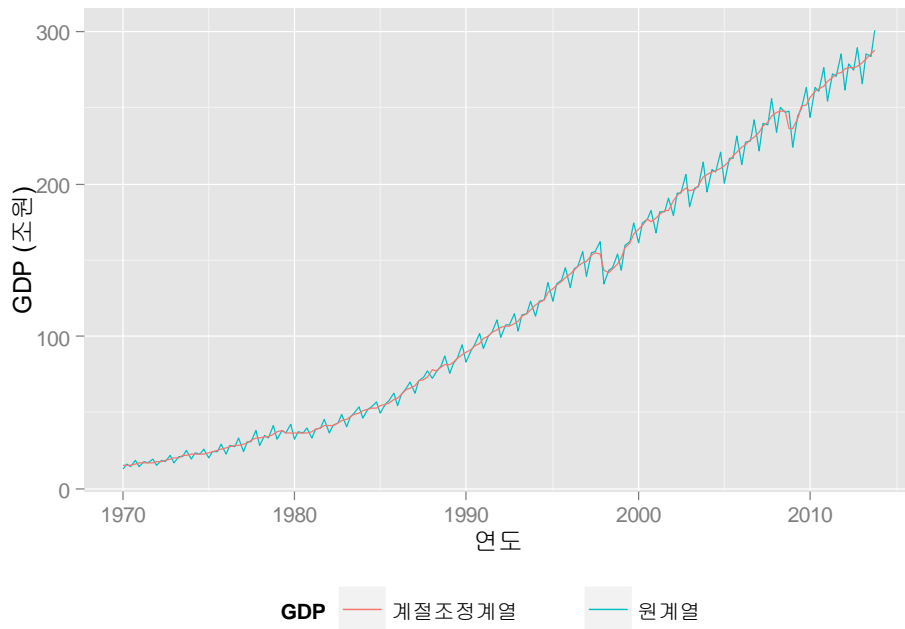


▶ ggplot2 패키지

■ ggplot2를 이용한 선 그래프의 작성

```
ggplot(data=gdp_kr, aes(x=연도))  
+ geom_line(aes(y = gdp/1000, colour="원계열"))  
+ geom_line(aes(y = gdpsa/1000, colour = "계절  
조정계열"))  
+ ylab("GDP (조원) ")  
+ scale_color_hue("GDP")  
+ theme(legend.position="bottom")
```

ggplot2 패키지

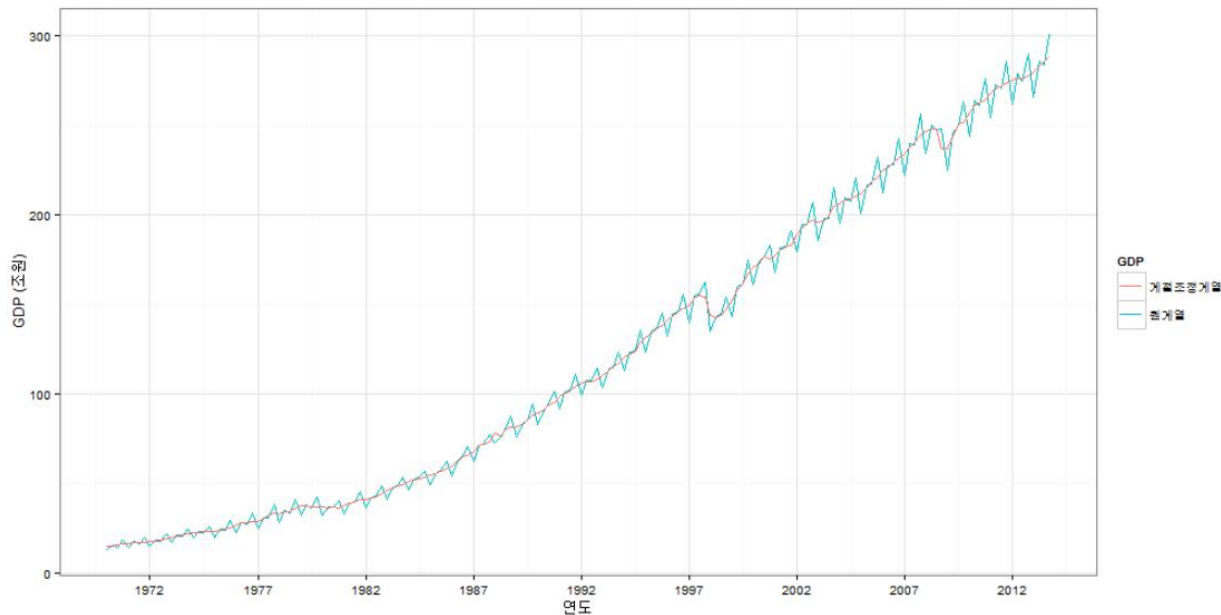


3. 선 그래프의 작성

- 시계열도표의 의미
- 시계열의 이동평균
- 시계열 변동요인

▶ 시계열도표의 의미

■ x축을 시간, y축을 데이터값으로 두고 선을 연결한 그래프



▶ 시계열도표

```
ggplot(gdp_kr, aes(x=연도))  
+ geom_line(aes(y = gdp/1000, colour="원계열"))  
+ geom_line(aes(y = gdpsa/1000, colour = "계절  
조정계열"))  
+ scale_x_date(breaks="5 years",  
labels=date_format("%Y"))  
+ labs(x="연도", y="GDP (조원)")  
+ scale_colour_hue("GDP")  
+ theme_bw()
```

▶ 시계열의 이동평균

■ 시계열 = 신호 + 소음

■ 중심화 이동평균 : $cma(2k + 1)_t = \frac{1}{(2k + 1)} \sum_{i=-k}^k y_{t-i}$

– 장점 : 시계열과 시차구조가 일치

– 단점 : 미래 데이터가 없어 최근 이동평균값
산출할 수 없음

▶ 시계열의 이동평균

- 후방 이동평균 : $bma(k)_t = \frac{1}{(2k+1)} \sum_{i=0}^{2k} y_{t-i}$
 - 장점 : 최근 이동평균값 산출 가능
 - 단점 : 중심화 이동평균값보다 후행

▶ 시계열의 이동평균

■ 우리나라 종합주가지수의 추이



▶ 시계열의 이동평균

```
library(ggplot2)
library(quantmod)
kospi <- getSymbols("^KS11", auto.assign = FALSE)[, 4]
kospi$ma <- runMean(kospi, n = 200)
colnames(kospi) <- c("종합주가지수", "200일 이평선")
autoplot(kospi, facets = NULL)
+ xlab("연도")
+ theme_bw(panel.background = element_rect(fill =
"white", colour = "gray"), legend.position = "none")
```

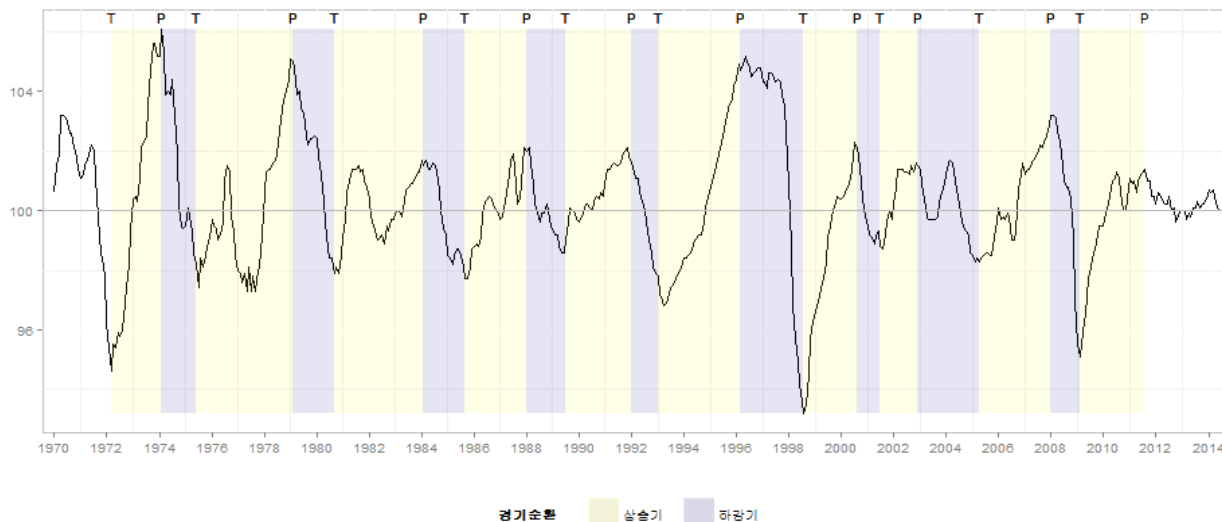
▶ 시계열의 변동요인

■ 시계열은 4개 변동으로 구성

- 추세변동
- 순환변동
- 계절변동
- 불규칙변동

▶ 시계열의 변동요인

■ 우리나라 경기동행지수 순환변동치의 추이



▶ 시계열의 변동요인

■ 우리나라 경기동행지수 순환변동치의 선그래프 작성

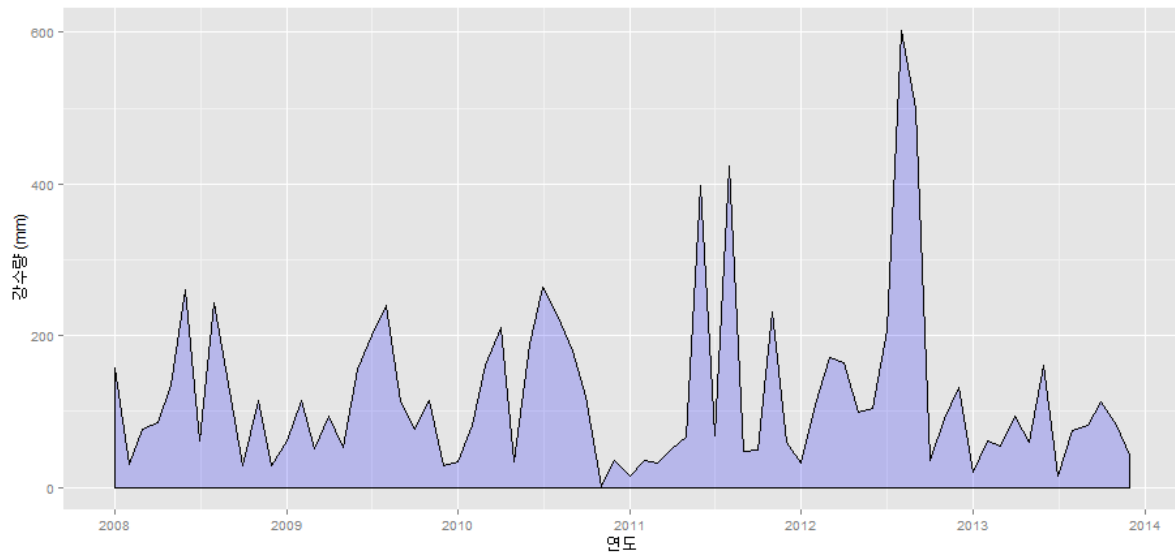
```
library(ggplot2)
library(scales)
library(xts)
cycle1 <- read.csv("cycle.csv",header=TRUE)
연도 <- seq(as.Date("1970-01-01"), as.Date("2014-06-01"),
            "month")
cycle <- xts(cycle1[,2],연도)
refdate <- read.csv("refdate1.csv",header=TRUE)
yrng <- range(cycle)
datebreaks <- seq(as.Date("1970-01-01"), as.Date("2014-01-01"), "2 year")
```

▶ 시계열의 변동요인

```
p <- autoplot(cycle, facets = NULL)
+ theme(panel.background = element_rect(fill = "white",
  colour = "gray"), legend.position = "bottom")
+ geom_rect(aes(NULL, NULL, xmin = as.Date(start),
  xmax = as.Date(end), fill = 경기순환), ymin = yrng[1],
  ymax = yrng[2], data = refdate)
+ scale_fill_manual(values = alpha(c("yellow","darkblue"),0.1))
+ ylab("") + xlab("")
+ geom_hline(yintercept=100, colour="gray")
+ geom_text(aes(x = as.Date(start), y = yrng[2], label =
  name1),data = refdate, size = 4, hjust = 0.5, vjust = -0.5)
+ geom_text(aes(x = as.Date(end), y = yrng[2], label = name2),
  data = refdate, size = 4, hjust = 0.5, vjust = -0.5)
p + scale_x_date(breaks=datebreaks, labels=date_format("%Y"),
  expand=c(0.01,0.01))
```

▶ 시계열의 변동요인

■ 제주시 강수량의 추이



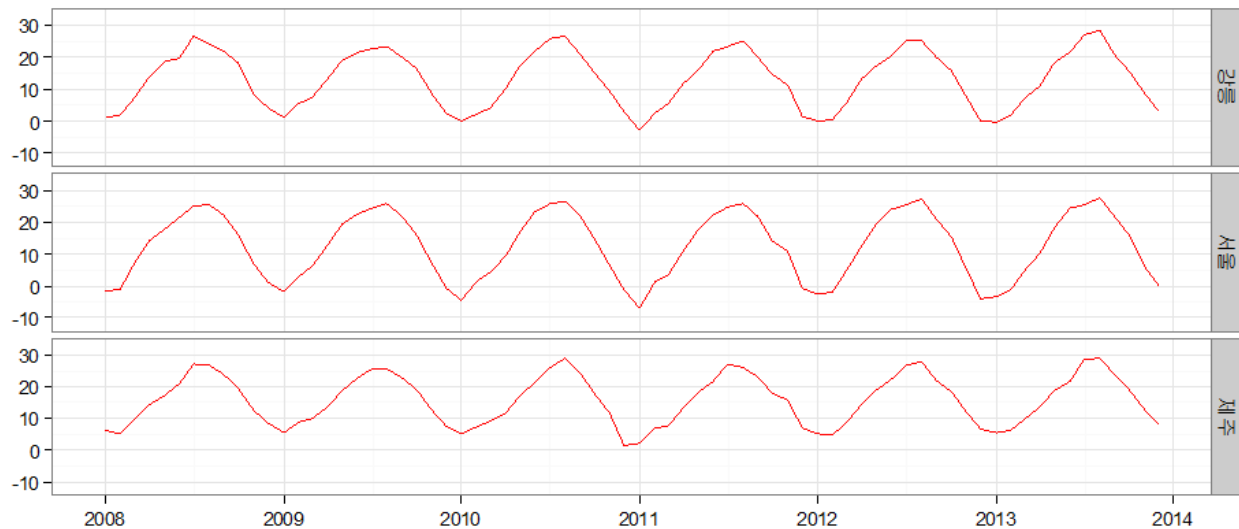
▶ 시계열의 변동요인

■ 제주시 강수량의 선그래프 작성

```
library(ggplot2)
climate_kr1 <- read.csv("climate.csv", header=TRUE)
연도 <- seq(as.Date("2008-01-01"), as.Date("2013-12-01"), "month")
climate_kr <- cbind(연도, climate_kr1)
ggplot(climate_kr, aes(x=연도, y=강수량))
  + geom_area(colour="black", fill="blue",
    alpha=.2) + ylab("강수량 (mm)")
```

▶ 시계열의 변동요인

■ 강릉시, 서울시, 제주시, 기온의 추이



▶ 시계열의 변동요인

```
library(ggplot2)
climate_kr1 <- read.csv("climate1.csv", header=TRUE)
연도 <- rep(seq(as.Date("2008-01-01"), as.Date("2013-12-01"), "month"), 3)
climate_kr <- cbind(연도, climate_kr1)
h <- ggplot(climate_kr, aes(x=연도))+facet_grid(지역 ~.)
  + geom_ribbon(aes(ymin=최저, ymax=최고), fill="pink",
    alpha=.7)
  + geom_line(aes(y=평균), colour="red") + theme_bw()
h + ylab("") + xlab("") + ylim(-12,33)
```



다음시간안내

시계열의 시각화 2