

14강

웹을 이용한 동적 · 대화형 데이터 시각화(1)

한국방송통신대학교 정보통계학과 이금희 교수

1. 동적 · 대화형 시각화
2. 웹을 이용한 동적 · 대화형 시각화의 사례
3. ggvis 패키지를 이용한 데이터 시각화
4. shiny 패키지

1. 동적 · 대화형 시각화

- 동적 · 대화형 시각화

1 동적 · 대화형 시각화

▶ 동적 · 대화형 시각화의 필요성

- 정적 그래프로 데이터의 수많은 특성을 표현하는데 제약
→ 동적, 대화형 시각화가 필요
- 동적 시각화 : 시간, 연속변수 변화 등에 따른 변화 파악
- 대화형 시각화 : 데이터의 보다 많은 특성을 파악

1 동적 · 대화형 시각화

▶ 동적 · 대화형 시각화의 배경

- 스마트폰 또는 패드 등의 기기확산, 다양한 웹 기술의 개발, 무선인터넷 등으로 활성화
- 자바스크립트, CSS 적용으로 웹 페이지에서 동적 시각화와 대화형 시각화 가능
- 구글차트도구, D3 등 다양한 데이터 시각화 도구들이 개발
- R에서도 ggvis, shiny 패키지와, 자바스크립트 라이브러리와 연동되는 googleVis, rCharts 패키지 등이 개발

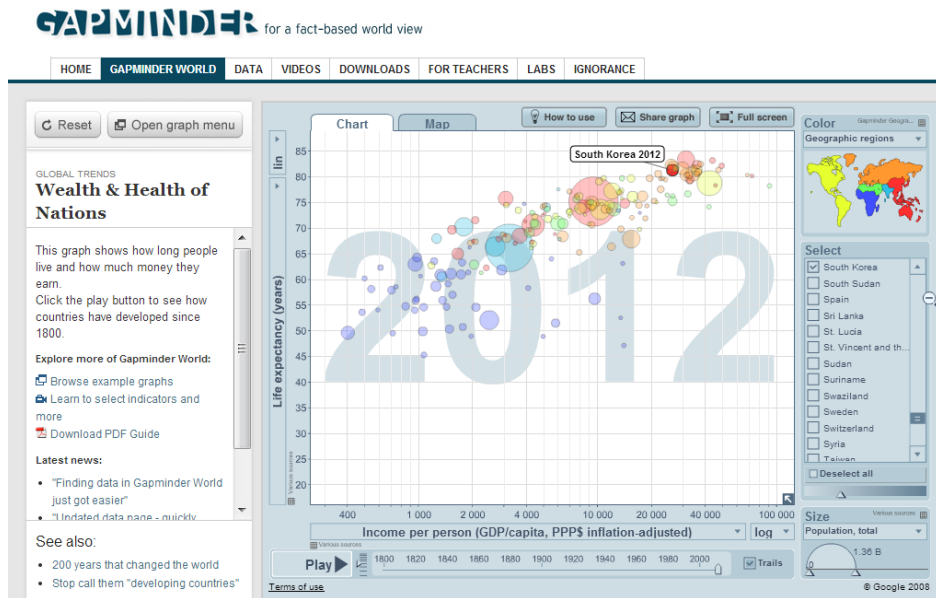
2. 웹을 이용한 동적 · 대화형 시각화의 사례

- 인구피라미드
- 모션 차트
- 경기순환시계
- 세계교역과 복잡성
- 지도를 이용한 그래프
- 국제 이민 흐름의 시간적 변화

▶ 인구 피라미드



모션 차트

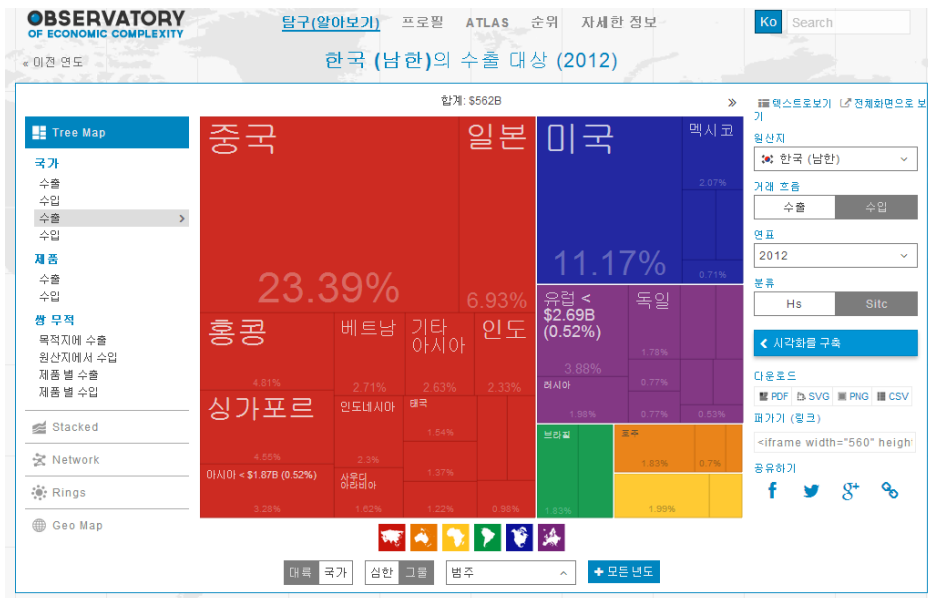


▶ 경기순환시계



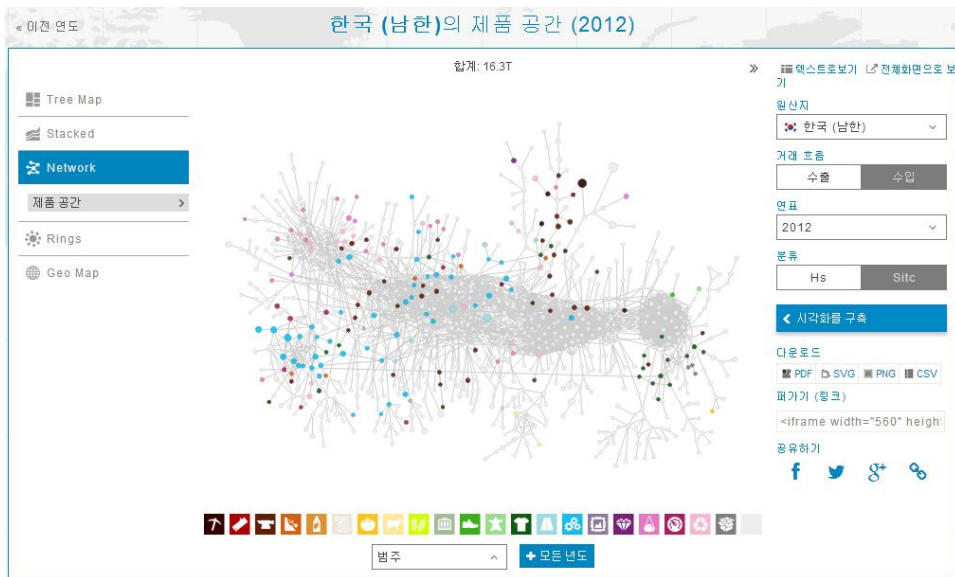
출처 : <http://kosis.kr/bcc/main.html>

▶ 세계교역과 복잡성(수출 국가의 구성)



출처 : http://atlas.media.mit.edu/explore/tree_map/hs/export/kor/show/all/2012/

▶ 세계교역과 복잡성(상품 관련성)



출처 : <http://atlas.media.mit.edu/explore/network/hs/export/kor/all/show/2012/>

▶ 지도를 이용한 그래프(세계은행의 GDP 대비 수출 비중)

Exports of goods and services (% of GDP)

DATAWAVE DOWNLOAD DATA SHARE

Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.

World Bank national accounts data, and OECD National Accounts data files.

Catalog Sources World Development Indicators

TABLE MAP GRAPH

Search all indicators

Go

Featured indicators

Economic Policy & External

Agriculture, value added (% of GDP)

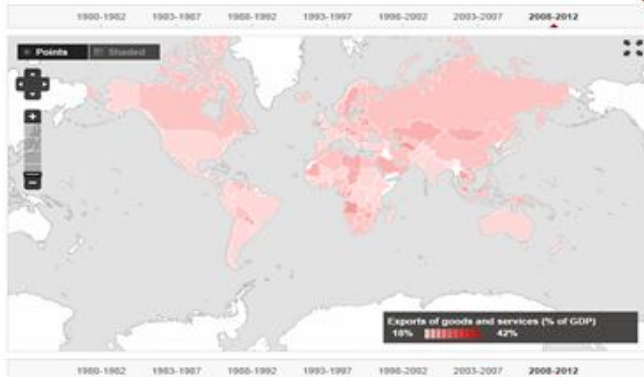
Cash surplus/deficit (% of GDP)

Central government debt, total (% of GDP)

Charges for the use of intellectual property, payments (BoP, current US\$)

Charges for the use of intellectual property, receipts (BoP, current US\$)

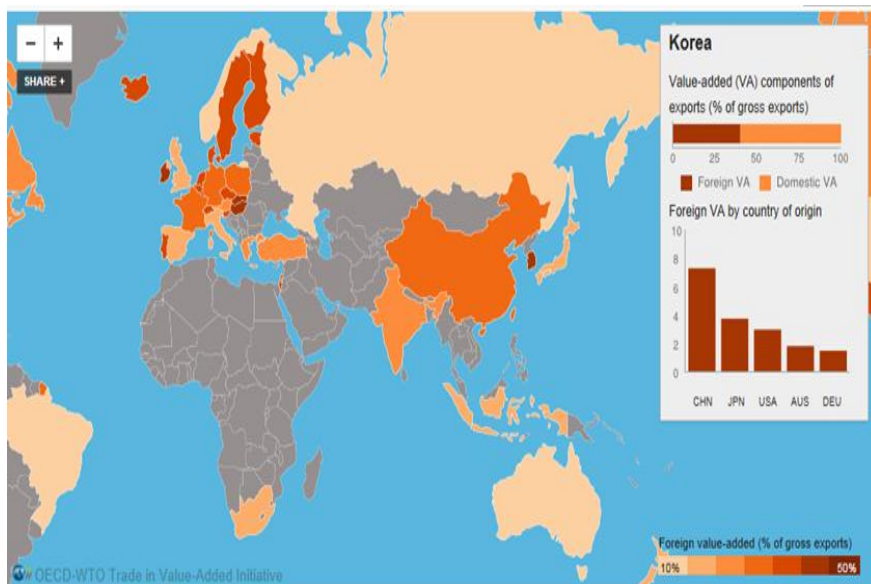
Current account balance (BoP, current US\$)



The maps displayed on the World Bank web site are for reference only and do not imply any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries.

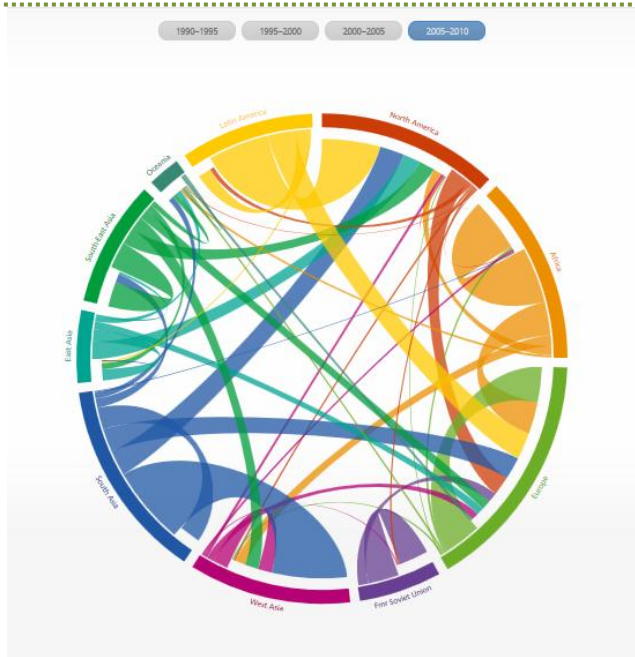
출처 : <http://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?display=map>

▶ 지도를 이용한 그래프(OECD회원국 GDP대비 수출비중)



출처 : <http://www.oecd.org/statistics/datalab/trade-in-value-added.htm>

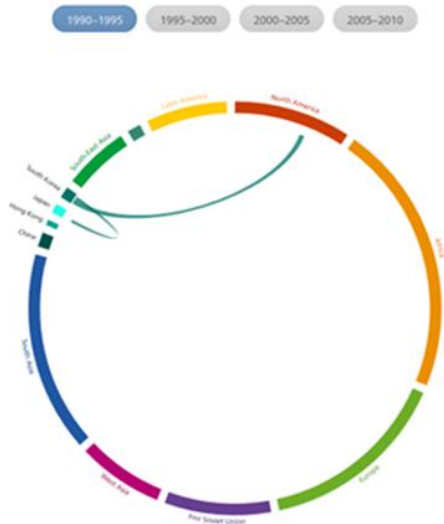
▶ 국제 이민 흐름의 시간적 변화



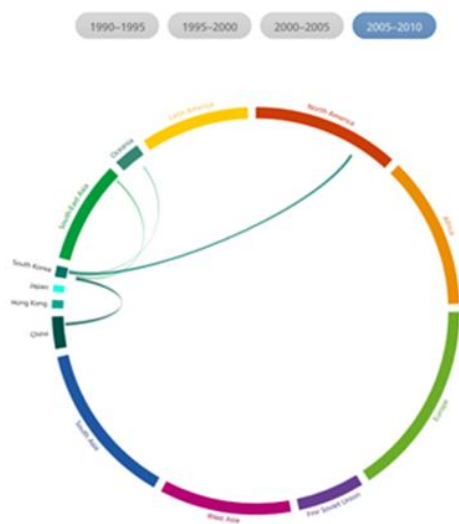
출처 : <http://www.global-migration.info/>

▶ 국제 이민 흐름의 시간적 변화[우리나라]

(1) 1990년 ~ 1995년



(2) 2005년 ~ 2010년



3. ggvis 패키지를 이용한 데이터 시각화

- ggvis 패키지의 특징
- 국내총생산 그래프 작성
- 종합주가지수 변동률의 확률밀도함수 추정



▶ ggvis 패키지

- ggplot2 패키지와 비슷한 아이디어로 만들어진 대화형 데이터 시각화 패키지
 - Vega 시각화 문법 이용하여 웹에 표현
- ggvis의 특징
 - 데이터처리 %> (pipeline) 이용
 - geom → layer

▶ 국내총생산 선 그래프 작성

```
library(zoo)
library(ggvis)
# 데이터 읽기
gdp_1 <- read.csv("i:/R/datavis/gdp.csv", header=TRUE)
연도   <- seq(as.Date("1970-01-01"), as.Date("2013-10-01"),
              "3 months")
gdp_kr <- cbind(연도, gdp_1)
```



▶ 국내총생산 선 그래프 작성

그래프 그리기

```
gdp_kr %>% ggvis(~연도) %>%  
  layer_lines(y=~gdp/1000, opacity := input_slider(0, 1,  
    label="Original series")) %>%  
  layer_lines(y=~gdpsa/1000, stroke := input_select(  
    c("Red" = "red", "Blue" = "blue", "Green" = "green"),  
    label = "SA Color")) %>%  
  add_axis("x", title = "") %>%  
  add_axis("y", title = "GDP")
```

▶ 국내총생산의 추이

국내총생산의 추이

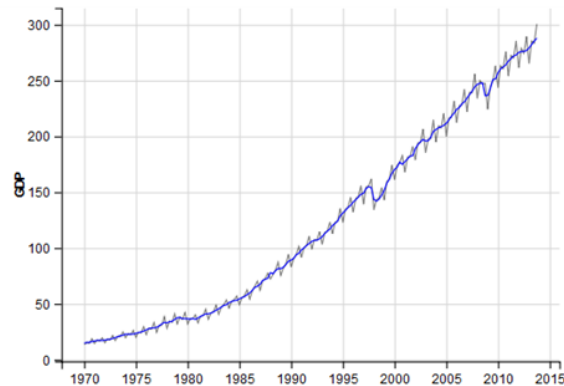


Original series

SA Color

0.5

Red



Original series

SA Color

0.5

Blue



▶ 종합주가지수 변동률의 확률밀도함수 추정

```
library(ggvis)
library(quantmod)
# Yahoo! Finance에서 종합주가지수 불러오기
kospil <- getSymbols("^KS11", auto.assign = FALSE)[, 4]
kospil$ma <- runMean(kospil, n = 200)
kospil$rate <- (kospil$KS11.Close -
lag(kospil$KS11.Close,1)) / lag(kospil$KS11.Close,1)
*100
kospil = as.data.frame(kospil)
```

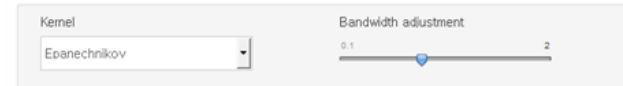
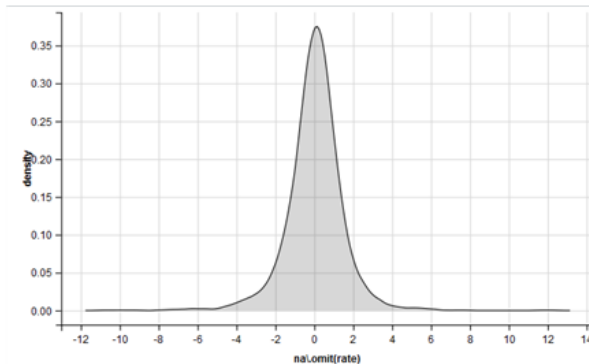
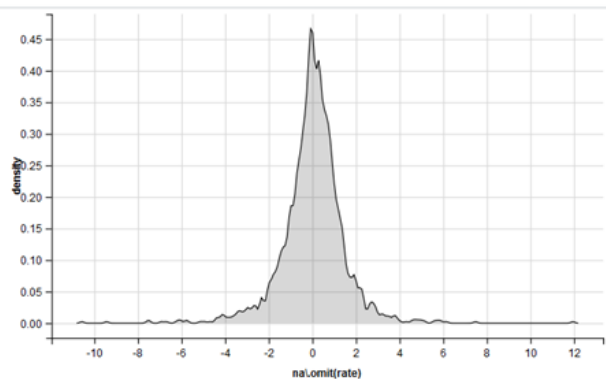
▶ 종합주가지수 변동률의 확률밀도함수 추정

종합주가지수 변동률의 확률밀도함수 추정

```
kospi %>% ggvis(~na.omit(rate)) %>%  
  layer_densities(  
    adjust = input_slider(.1, 5, value = 1, step = .1, label =  
      "Bandwidth adjustment"),  
    kernel = input_select( c("Gaussian" = "gaussian",  
      "Epanechnikov"="epanechnikov", "Rectangular" = "rectangular",  
      "Triangular" = "triangular", "Biweight" = "biweight",  
      "Cosine" = "cosine", "Optcosine" = "optcosine"),  
    label = "Kernel"))
```

▶ 종합주가지수 변동률의 확률밀도함수 추정

종합주가지수 변동률의 확률밀도함수 추정



4. shiny 패키지

- 붓꽃 데이터 군집화의 사용자인터페이스 스크립트 : ui.R
- 붓꽃 데이터 군집화의 서버 스크립트 : server.R
- 붓꽃 데이터의 분류

▶ shiny 패키지

- 대화형 웹 애플리케이션을 만들 수 있는 R 패키지
 - 사용자 인터페이스 스크립트 : ui.R
 - 서버 스크립트 : server.R
- shiny의 작동
 - 두 프로그램을 같은 폴더에 둠
 - runApp()

▶ 붓꽃 데이터 군집화의 ui.R

```
shinyUI(pageWithSidebar(  
  headerPanel('붓꽃 k-means 군집화'),  
  sidebarPanel(  
    selectInput('xcol', 'X 변수', names(iris)),  
    selectInput('ycol', 'Y 변수', names(iris),  
               selected=names(iris)[[2]]),  
    numericInput('clusters', '군집 수', 3,  
                 min = 1, max = 9)),  
  mainPanel(  
    plotOutput('plot1')  
  )  
))
```

▶ 붓꽃 데이터 군집화의 server.R

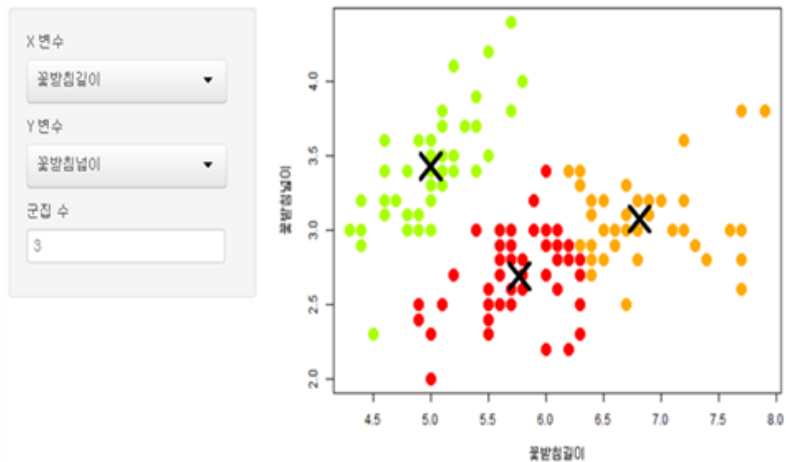
```
palette(rainbow(9))
shinyServer(function(input, output, session) {
  selectedData <- reactive({iris[, c(input$xcol,
input$ycol)]})
  clusters <- reactive({kmeans(selectedData(),
input$clusters)})
  output$plot1 <- renderPlot({
    par(mar = c(5.1, 4.1, 0, 1))
    plot(selectedData(),
          col = clusters()$cluster, pch = 20, cex = 3)
    points(clusters()$centers, pch = 4, cex = 4, lwd = 4)
  })
})
```

4 shiny 패키지

▶ 붓꽃 데이터의 분류

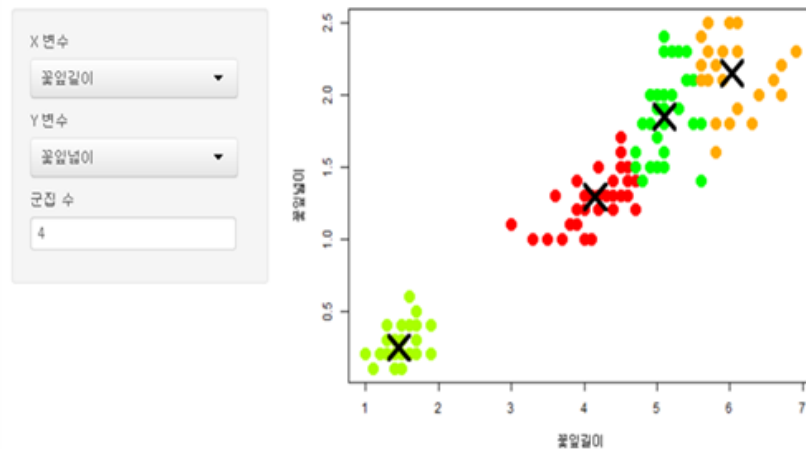
(1) 군집 수 3

붓꽃 k-means 군집화



(2) 군집 수 4

붓꽃 k-means 군집화





다음시간안내

웹을 이용한 동적 · 대화형 데이터 시각화 (2)