
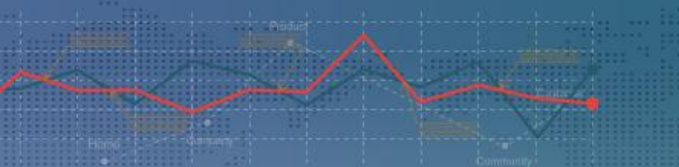




1강 데이터 시각화란 무엇인가?

고려대학교 통계학과 허명희 교수

1. 시각화는 왜?
 2. 시각화는 어떻게?
 3. 시각화의 도구 R
 4. R 시각화 사례
 5. R의 컬러 체계
- 



1. 시각화는 왜?

- 시각화의 목표
- 보기 1. 바라바시-앨버트 네트워크
- 보기 2. 런던의 콜레라 지도



1 시각화는 왜?

▶ 시각화의 목표 : 통계적 정보를 효과적인 그림으로 ...

- 데이터 : 가공되지 않은 원재료 같은 것
- 정보 : 데이터로부터 추출되는 가치 있는 것
- 시각화 : 정보의 인지

1 시각화는 왜?

▶ 보기 1. 바라바시-앨버트 네트워크

■ 사회네트워크의 한 종류

- 행렬 A 의 (i, j) 요소 a_{ij} 는 i 와 j 가 소통하면 1, 아니면 0
- $n = 400$ 인 경우 행렬 A 는 $400 * 400$, 비대각 칸의 수는 159,600

1 시각화는 왜?

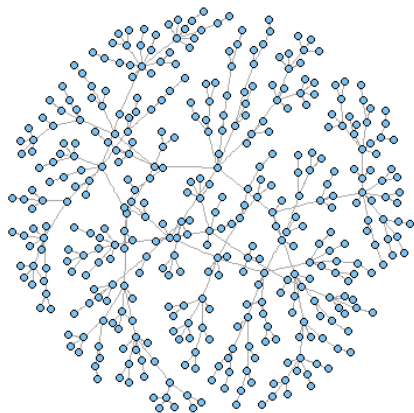
▶ 보기 1. 바라바시-앨버트 네트워크

- 바라바시 네트워크 : 새 노드는 어느 기존 노드와 결합하는가?
 - 새 노드는 이미 연결 노드가 많은 기존 노드에 붙을 가능성이 높음을 가정
 - 네트워크 1: 기존 연결선 수와 무관한 경우를 상정
 - 네트워크 2: 기존 연결선 수의 1.1 거듭 곱에 비례하는 경우
 - 네트워크1은 random 네트워크,
 - 네트워크2가 바라바시 네트워크
 - $n = 400$ 인 경우 두 네트워크는 어떻게 다른가?

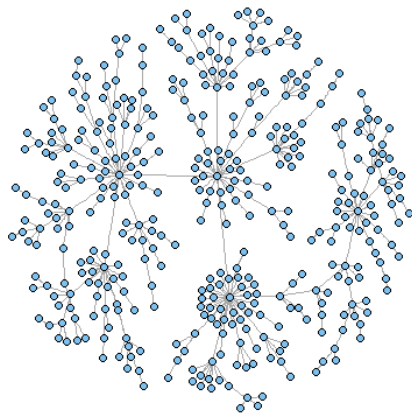
1 시각화는 왜?

▶ 보기 1. 바라바시-앨버트 네트워크

■ 네트워크 1



■ 네트워크 2



1 시각화는 왜?

▶ 보기 2. 런던의 콜레라 지도

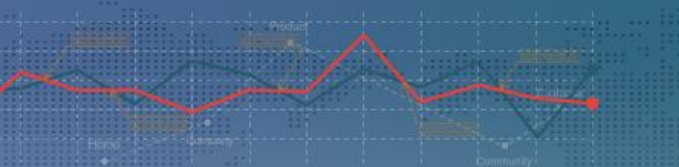
- 1854년 런던, 콜레라 발생 : 콜레라는 무엇으로 전염되는가?
 - 의학 연구자 John Snow
 - 지도 상에서 사망자 집 위치에 막대 선 하나 씩 넣음
 - 그 결과는?

1 시각화는 왜?

▶ 보기 2. 런던의 콜레라 지도

- Snow의 콜레라 지도 (Wikipedia, “1854 Broad Street cholera outbreak”)





2. 시각화는 어떻게?

- 단계 1. 데이터의 이해
- 단계 2. 목표설정
- 단계 3. 그래프 선정
- 단계 4. 소프트웨어의 활용
- 단계 5. 스토리텔링



2 시각화는 어떻게?

▶ 1단계: 데이터의 이해

- 개체의 수집 범위와 방법, 그리고 개체 수의 파악
- 변수의 측정(기록), 그리고 수준(명목형, 순서형, 수치형)
- 예 : 병원 A의 내원환자들에 대한 진료기록(수집범위, 방법),
000명(개체 수), 성별(명목형),
나이대(순서형), 질병의 지속일(수치형)

2 시각화는 어떻게?

▶ 2단계: 목표 설정

- 무엇을 알고 싶은가?
- 이 데이터가 답해 줄 수 있는가?
- 배경지식의 습득
- 예 : 질병의 지속일은 대체로 어느 정도인가?
성별로 차이가 있는가?

2 시각화는 어떻게?

▶ 3단계: 그래프 선정

- 어떤 그래프가 좋을까?

▶ 4단계: 소프트웨어의 활용 (R)

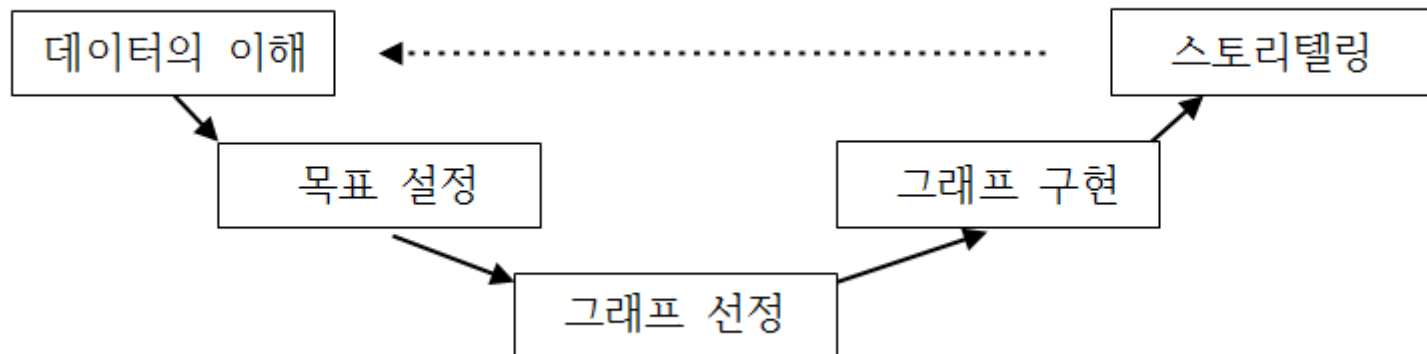
- 오픈소스의 데이터분석 도구, 시각화에 강점

▶ 5단계: 스토리텔링 (story telling)

- 핵심적 의미를 일반인에 전달


2 시각화는 어떻게?

▶ 데이터 시각화의 과정





3. 시각화의 도구

- R이란?
 - R의 기본 시스템 설치
 - R 스크립트의 실행
 - R 패키지의 설치
- 

3 시각화의 도구 R

▶ R이란?

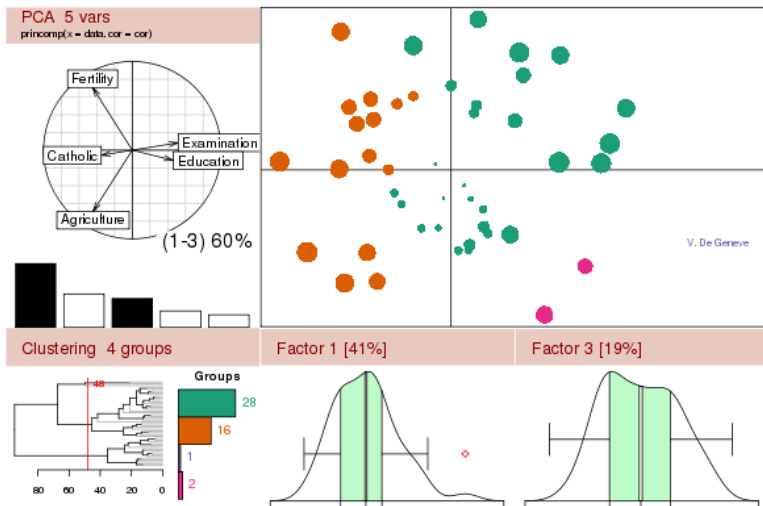
- 오픈소스, 사용자의 자발적 기여
- 개별적 needs에 맞추는 변형과 창작이 가능
- 그래프가 멋짐, R 그래픽스 패키지를 활용
 - `lattice`
 - `ggplot2`
 - `rgl`
 - `RColorBrewer`

3 시각화의 도구 R

▶ R의 기본 시스템 설치

■ R의 공식 홈페이지 <http://www.r-project.org> 접속

The R Project for Statistical Computing



3 시각화의 도구 R

▶ R의 기본 시스템 설치

■ cran 연결 → local mirror site 연결

Korea

- <http://cran.nexr.com/>

NexR Corporation, Seoul

- <http://healthstat.snu.ac.kr/CRAN/>

Graduate School of Public Health,
Seoul National University, Seoul

- <http://can.biodisk.org/>

The Genome Institute of UNIST
(Ulsan National Institute of Science
and Technology)

3 시각화의 도구 R

▶ R의 기본 시스템 설치

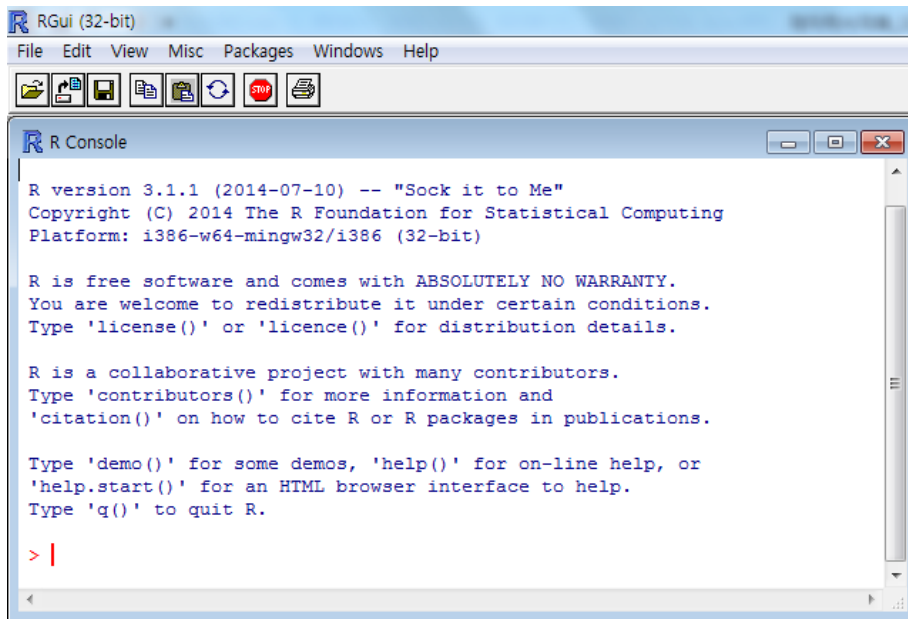
■ Download R for Windows 선택

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

3 시각화의 도구 R

▶ R의 기본 시스템 설치

■ R base를 설치, 실행



The screenshot shows the RGui (32-bit) application window. The menu bar includes File, Edit, View, Misc, Packages, Windows, and Help. Below the menu bar is a toolbar with icons for file operations and execution. The main window is the R Console, which displays the following text:

```
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

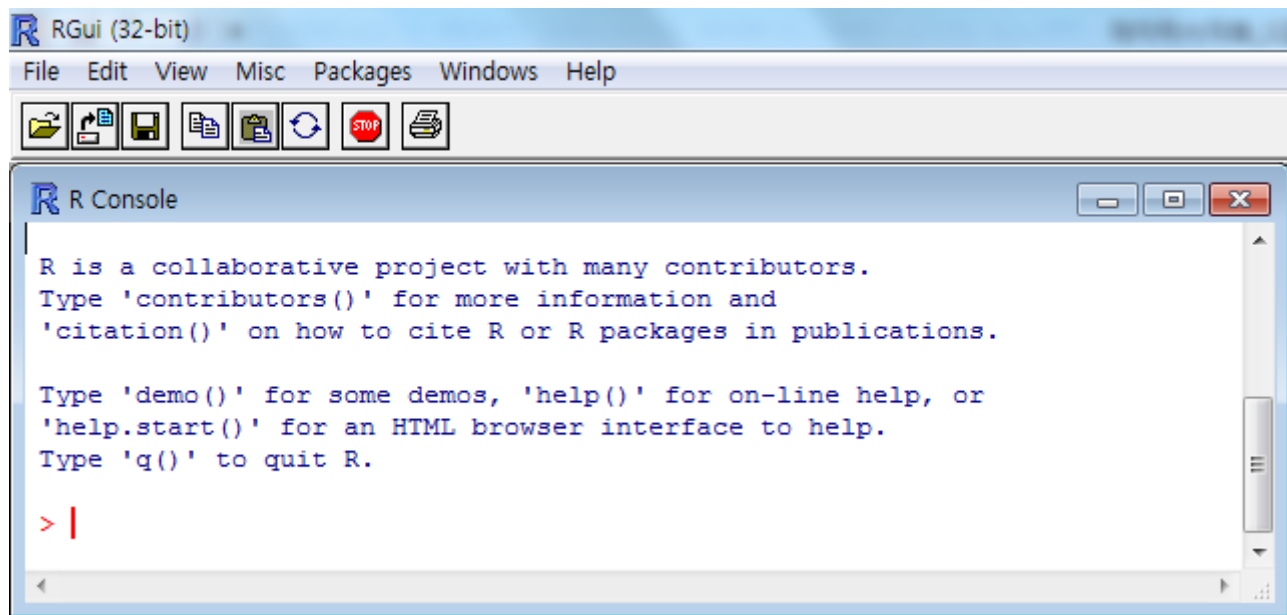
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

3 시각화의 도구 R

▶ R의 기본 시스템 설치

■ R base를 설치, 실행



3 시각화의 도구 R

▶ R 스크립트의 실행

- R Gui의 File 메뉴에서 Change directory
 - 교재 R 실습파일이 위치한 파일 폴더를 지정
- R Gui의 File 메뉴에서 Open script
 - 해당 R 스크립트를 오픈
- R 스크립트 창에서 실행한 부분을 선택하여
 - Run line or selection

3 시각화의 도구 R

▶ R 패키지의 설치

- R Gui의 Packages 메뉴에서 Install package(s)
 - CRAN mirror를 선택
 - R 패키지를 선택
- R Gui의 Packages 메뉴에서 Load package(s)
 - 설치 확인



4. 시각화 사례

- ## ■ 조선의 임금님들은 얼마나 재워하였나?



4 R 시각화 사례

▶ 조선의 임금님들은 몇 년이나 재위하였나?

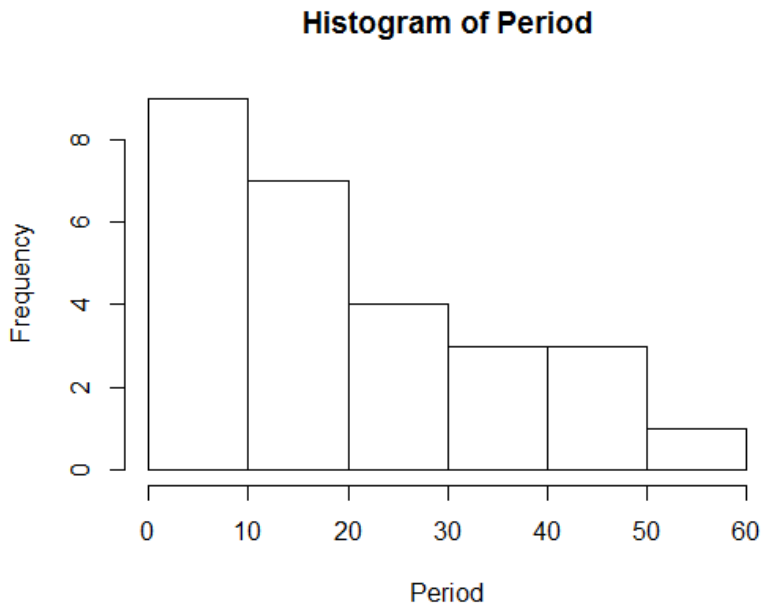
■ R 스크립트

```
kings <- read.table("chosun kings.txt", header=T)
str(kings)
attach(kings)
windows(width=5.5, height=4.5)
hist(Period)
```


4 R 시각화 사례

▶ 조선의 임금님들은 몇 년이나 재위하였나?

■ R 그래프 출력 : 기본형의 히스토그램



4 R 시각화 사례

▶ 조선의 임금님들은 몇 년이나 재위하였나?

■ R 스크립트

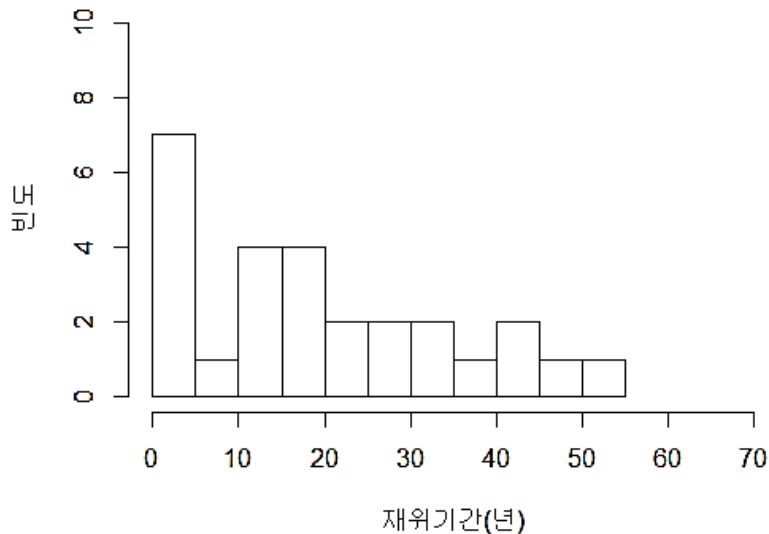
```
hist(Period, xlim=c(0,70), ylim=c(0,10),  
      nclass=14,  
      right=F,  
      main="조선 왕조",  
      xlab="재위기간(년)",  
      ylab="빈도")
```

4 R 시각화 사례

▶ 조선의 임금님들은 몇 년이나 재위하였나?

■ R 그래프 출력 : 향상된 히스토그램

조선 왕조





5. R의 컬러 체계

- RGB 시스템
- Chart of R Colors
- 각종 컬러 함수
- RColorBrewer의 순차형 컬러



▶ RGB 시스템

■ (red, green, blue)

■ 2자리의 헥사(16진법) 코드:

00, 01, ..., 09, 0A, 0B, 0C, 0D, 0E, 0F,

00, 01, ..., 09, 0A, 0B, 0C, 0D, 0E, 0F,

⋮

F0, F1, ..., F9, FA, FB, FC, FD, FE, FF.

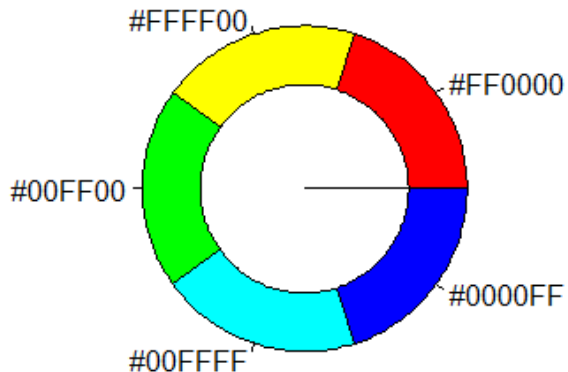
- “#FF0000” = 빨강(red)
“#00FF00” = 초록(green)
“#0000FF” = 파랑(blue)

5 R의 컬러 체계

▶ RGB 시스템

■ R 스크립트

```
color <- c("#FF0000", "#FFFF00", "#00FF00", "#00FFFF", "#0000FF", "#FF00FF")  
pie(rep(1,5), col=color, labels=color)  
par(new=T); pie(rep(1,1), col="white", radius=0.5, labels="")
```



▶ Chart of R Colors

1	white	#FFFFFF	255	255	255
2	aliceblue	#F0F8FF	240	248	255
3	antiquewhite	#FAEBD7	250	235	215
4	antiquewhite1	#FFEFDB	255	239	219
5	antiquewhite2	#EEDFCC	238	223	204
6	antiquewhite3	#CDC0B0	205	192	176
7	antiquewhite4	#8B8378	139	131	120

...

651	whitesmoke	#F5F5F5	245	245	245
652	yellow	#FFFF00	255	255	0
653	yellow1	#FFFF00	255	255	0
654	yellow2	#EEEE00	238	238	0
655	yellow3	#CDCD00	205	205	0
656	yellow4	#8B8B00	139	139	0
657	yellowgreen	#9ACD32	154	205	50

<http://research.stowers-institute.org/efg/R/Color/Chart/>

▶ grDevices 패키지가 제공하는 각종 컬러 함수

Rainbow colors



heat colors



terrain colors



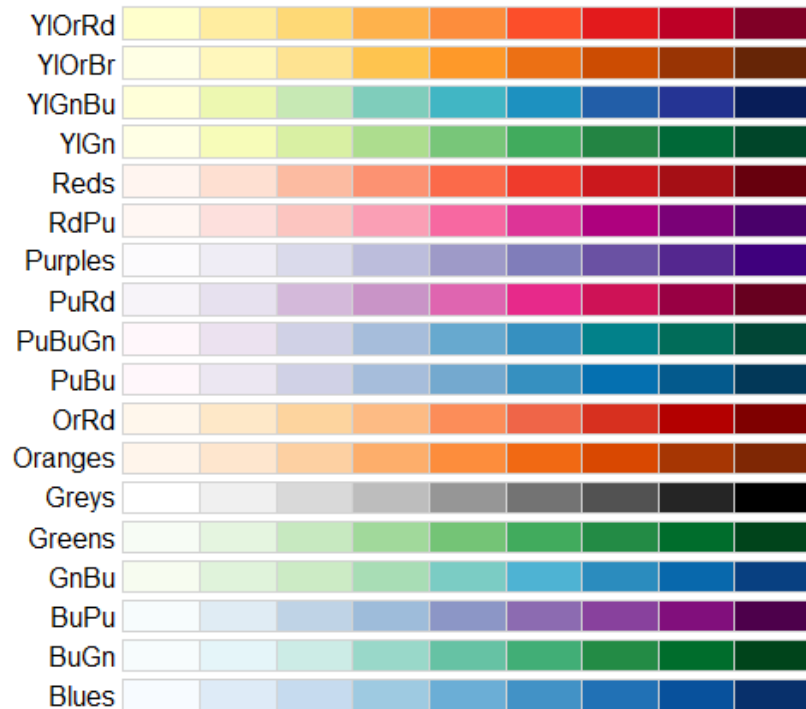
topo colors



Cyan-magenta colors



▶ RColorBrewer 패키지의 순차형 컬러





- 시각화 = 데이터에 담긴 통계적 정보를 그림의 형태로 드러내는 것
- 데이터 시각화의 과정
 - › 데이터의 이해 → 목표 설정 → 그래프 선정 → 그래프 구현 → 스토리텔링
- R은 오픈소스의 통계 소프트웨어, 시각화에 강점
 - › 스크립트 작성, 실행 [따라하기]
- R의 컬러체계: (R, G, B) 시스템, 2자리의 헥사 코드
 - › “#FF0000” = 빨강(red)