

인포그래픽의 유형	데이터 시각화 분석특
<ul style="list-style-type: none"> ▶ 통계 기반 인포그래픽 ▶ 타임라인 기반 인포그래픽 ▶ 프로세스 기반 인포그래픽 ▶ 위치/지리 기반 인포그래픽 ▶ 비교형식 기반 인포그래픽 ▶ 스토리 기반 인포그래픽 	<ul style="list-style-type: none"> ▶ DataWangler : 웹기반의 데이터 정제서비스 ▶ R ▶ 구글 퓨전 테이블 ▶ D3.js ▶ OpenStreetMap

데이터 시각화에 필요한 환경설정 : XAMPP
 데이터 시각화 맛보기 : <http://d3js.org>

정부 3.0 공공데이터포털 <https://www.data.go.kr/> 경기데이터드림 /
 TourAPI 국내 유일의 다국어 관광정보 제공
 농림축산식품 공공데이터 포털
 네이버 데이터랩
 미국 공공데이터 포털 data.gov
 유럽연합 공공포털 data.gov.uk
 미국통계국 www.census.gov/library/
 호주통계청 : 통계가 어떻게 생산되는지 공부할 수 있도록 동영상으로 제공 www.abs.gov.au

데이터 정제는 데이터 내에 있는 노이즈를 제거하고 일관성이 결여되는 내용을 교정하는 것입니다.

엑셀

- ▶ 수억 건의 통계자료를 기계적으로 처리하기 어려움
- ▶ 데이터를 수집하기 어려움

R

- ▶ 데이터를 통계낸 후 해석하고 의사결정하는 데 효율적임
- ▶ 다른 빅데이터 분석 프로그램인 하둡, SAS, SPSS와 같은 전문적인 통계프로그램보다 배우기 쉬움
- ▶ 텍스트, 엑셀, DBMS 등 다양한 종류의 정형/비정형 데이터 이용이 가능
- ▶ 윈도우, 유닉스, 리눅스, 맥OS 등 다양한 플랫폼에서 수집/저장/분석/시각화가 가능하며 통계가 가능

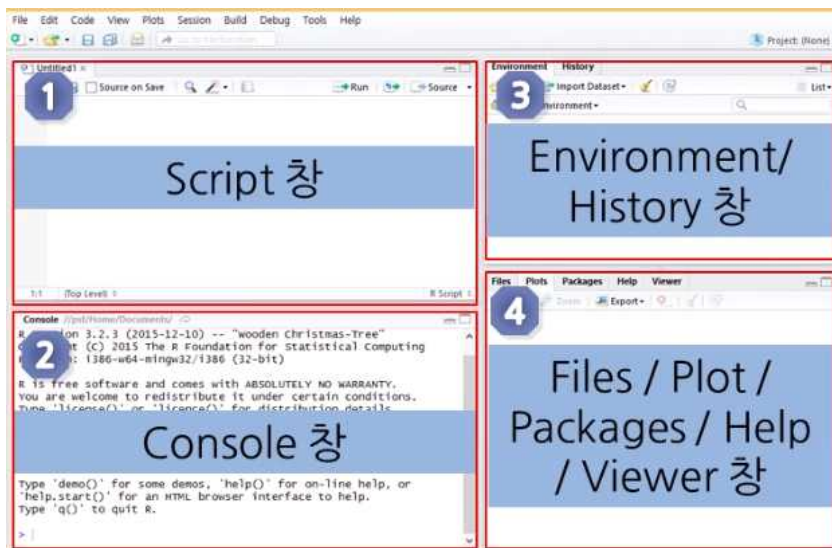
R 프로그램은 DOS 상태의 인터프리터 방식의 실행을 통해 프로그래밍 함

Rstudio : R 프로그램 환경에서 스크립트언어를 반복 사용할 수 있는 스크립트창, 히스토리창, 파일창, 환경설정창, 차트창 등을 제공하여 더욱 편리한 환경을 제공함.

자바 다운로드

www.oracle.com ▶ download ▶ JAVA SE ▶ JAVA(화면에서는 Java Platform(JDK) 8u111/8u112 ▶ Windows x83는 32bit용 / Windows x64는 64bit용 ▶ JDK설치 진행

▶ 시스템 환경변수 변경 : 시스템 ▶ 고급 시스템 설정(시작버튼 우클릭 - 고급 시스템 설정) ▶ 환경변수 : 시스템 변수(S)에서의 Path 더블클릭



Script창 :

기본 명령어를 사용하여 RScript를 실행 또는 저장할 수 있음

(실행방법)

ctrl + R / ctrl + Enter / 블록 설정 후 RStudio Script창 상단의 Run 단추

RScript언어 특징

1. 인터프리터 형태의 언어 : Console창에 직접 RScript문장을 입력하면 바로 결과를 반환받을 수 있음 하지만 작성한 스크립트를 재실행하려면 재입력해야하는 불편함이 있음. ◀ RStudio를 사용하는 경우 RScript창에서 실행을 원하는 코드로 이동하여 실행시키는 편의성을 누릴 수 있음.
2. '>'는 명령 프롬프트이고 명령을 실행하려면 명령 입력 후 Enter를 누름(console창에서는 enter만 눌러도 실행됨!)
- 3.
4. 대소문자 구분
5. 주석기호 : '#'

주요함수

data() : 데이터셋을 가져오기 위해서 data()함수를 호출. eg) data(mtcars)

head() : 앞 부분의 일부 데이터만 불러올 수 있음

```
library(googleVis)
```

```
demo(googleVis)
```

선차트 / 막대차트 / 열차트 / 지역차트 / 계단식 영역차트 / 콤보차트 / 구글 모션 차트

```
gvisTable
```

gvisMotionChart함수를 이용하여 쉽게 모션 차트를 그릴 수 있음

[data.frame](#)를 판독하고, 웹 페이지에 혹은 독립적인 페이지로 포함될 수 있으며 구글 시각화 API를 참조하여 모션 그래프를 생성함 ▶ 실제 차트는 플래시의 웹 브라우저에서 렌더링 됨

```
> gvisMotionChart(data, idvar = "id", timevar = "time",  
                  date.format = "%Y/%m/%d", option = list(), chartid)
```

	Fruit	Year	Location	Sales	Expenses	Profit	Date
1	Apples	2008	West	98	76	20	2008-12-31
2	Apples	2009	West	111	79	32	2009-12-31
3	Apples	2010	West	89	76	13	2010-12-31
4	Oranges	2008	East	96	81	15	2008-12-31
5	Bananas	2008	East	83	76	9	2008-12-31
6	Oranges	2009	East	93	80	13	2009-12-31
7	Bananas	2009	East	94	78	16	2009-12-31
8	Oranges	2010	East	98	91	7	2010-12-31
9	Bananas	2010	East	81	71	10	2010-12-31

```
M1 <- gvisMotionChart(Fruits, idvar = "Fruit", timevar =  
"Year")
```

```
M1
```

```
plot(M1) (년단위)
```

```
M2 <- gvisMotionChart(Fruits, idvar = "Fruit", timevar =  
"Date", date.format = "%Y%m%d")
```

```
plot(M2)
```

```
M2 <- gvisMotionChart(Fruits, idvar = "Fruit", timevar =  
"Date", date.format = "%Yw%w") (주단위)
```

년도에 따라 Fruits 데이터프레임(dataframe)의 Fruit항목들의 변하는 모습을 움직이는 모션 그래프로 나타남을 확인할 수 있음!! (항목명과 시간만 입력했는데 어떤 값을 기준으로 이동하는거지?)

미세먼지 모션 그래프 실습

www.data.go.kr에서 미세먼지 파일을 다운로드 받은 후

날짜자료의 속성을 -> 숫자 -> 일반의 형태로 차례대로 바꿔줌(각 시간대를 나타내는 숫자로 변형됨)

RStudio

```
setwd("c:/R")
```

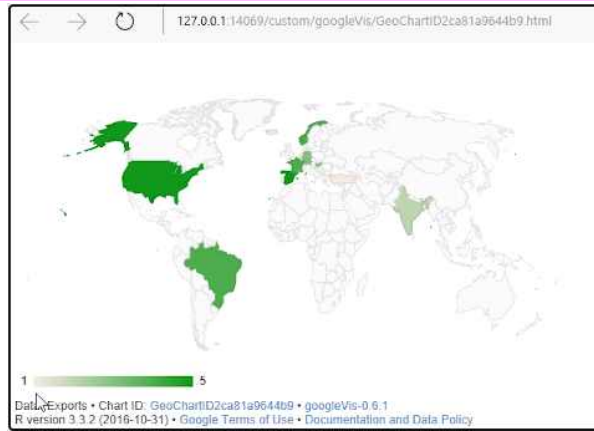
```
library(googleVis)
```

```
mise <- read.csv("미세먼지3.csv", header=T) 시간형식을 바꾸지 않았으면 The timevar has to be numeric  
or Date format 이라고 에러메세지가 발생함
```

```
mise$“일시”
```

```
mise2 <- gvisMotionChart(mise.idvar="지점", timevar="일시")
```

```
plot(mise2)
```



```
require(datasets)
states <- data.frame(state.name, state.x77)
GeoStates <- gvisGeoChart(states, "state.name",
  "Illiteracy", options=list(region="US",
    displayMode="regions",
    resolution="provinces", width=600, height=400)
```

그래픽 함수

plot 함수

`par(mfrow=c(1,3))` : 1행 3열의 구성으로 그래픽이 생성될 수 있도록 공간분할

`plot(1:10, type="p")` 점플롯(`type="p"`)

`plot(1:10, type="l")` 선플롯(`type="l"`)

`plot(1:10, type="b")` 점과 선으로 된 플롯(`type="b"`)

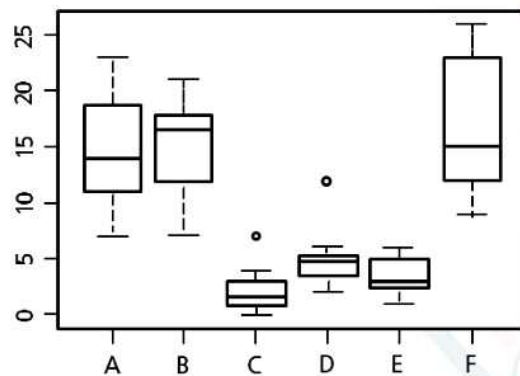
boxplot함수

상자모양의 차트를 그릴 수 있음.

데이터의 최소값, 1사분위수, 중위수(2사분위수), 평균, 3사분위수, 최대값을 구하여 표현함

The `boxplot()` function takes in any number of numeric vectors, drawing a boxplot for each vector. You can also pass in a list (or data frame) with numeric vectors as its components.

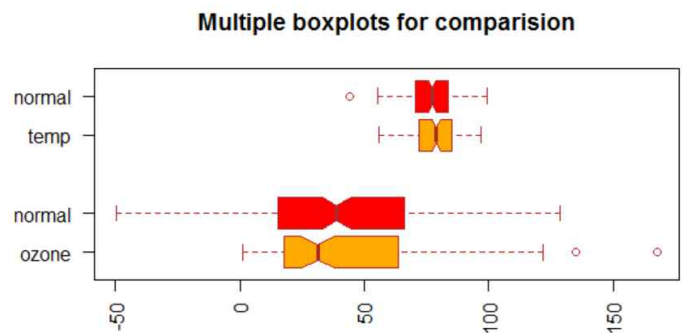
```
>InsectSprays  
>boxplot(count ~ spray,  
data = InsectSprays, col = "lightgray")
```



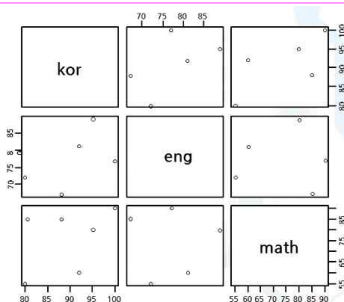
boxplot에 `count ~ spray`와 같은 형태의 데이터도 입력할 경우 `spray`별 `count`의 boxplot이 나타남.

We can draw multiple boxplots in a single plot, by passing in a list, data frame or multiple vectors.

```
boxplot(ozone, ozone_norm, temp, temp_norm,  
main = "Multiple boxplots for comparison",  
at = c(1,2,4,5),  
names = c("ozone", "normal", "temp", "normal"),  
las = 2,  
col = c("orange","red"),  
border = "brown",  
horizontal = TRUE,  
notch = TRUE  
)
```



hist함수 / qqnorm함수 / curve함수



데이터프레임을 `plot`함수에 넣어 산점도 그래프를 도출했을 때
(현재 kor, eng, math 세 벡터로 이루어진 데이터프레임이 표현됨)

저수준그래픽 함수

기존에 그려진 그래프에 추가적인 점, 선, 텍스트장식, 범례등을 추가로 그리는 함수

points	점을 추가	polygon	닫힌 다각형을 추가함	grid	격자	axis	축
lines	선을 추가	text	텍스트를 추가함	arrows	화살표	box	테두리
ablines	직선을 추가	mtext	텍스트를 추가함	rect	직사각형		
segments	선분을 추가	legend	범례를 추가함	title			

저수준그래픽 함수 사용예

arrow(7,7,8,8)	rect(1,6,3,8)
점 (7,7)과 (8,8)을 지나는 화살표 그리기	점(1,6)와 점(3,8) 사이에 사각형 그리기
segments(4,4,6,6)	text(2,7,"문자열")
점(1,6)와 점(3,8) 사이에 사각형 그리기	점(2,7) 위치에 "문자열" 부분 출력하기

고수준 그래픽 함수 옵션

main="메인 제목"	제목 설정	type="p"	점 모양의 그래프(기본값)
sub="서브 제목"	서브 제목 설정	type="l"	선 모양 그래프(꺼은선 그래프)
xlab="문자", ylab="문자"	x,y축에 사용할 문자열을 지정	type="b"	점과 선 모양 그래프
ann=F	x,y축 제목을 지정하지 않음	type="c"	"b"에서 점을 생략
tmag=2	제목 등에 사용되는 문자의 확대율 지정	type="o"	점과 선을 중첩해서 그린 그래프
axes=F	x,y축을 표시하지 않음	type="h"	각 점에서 x축까지의 수직선 그래프
axis	x,y축을 사용자의 지정값으로 표시	type="s"	왼쪽 값을 계단모양으로 연결한 그래프
		type="S"	오른쪽 값을 계단모양으로 연결한 그래프
		type="n"	축만 그리고 그래프를 그리지 않음

자료보고... 옮겨 적을 것...?::: 꽤 시간 소요될 듯.(20분?)

데이터 시각화 색상함수

색상함수 : RGB, 색관련 모수, R colors

동일한 그래프를 출력하는 네 가지 코드

`barplot(x, col="F6C2F4")`

`barplot(x, col=rgb(246,194,244, max=255))`

`barplot(x, col=hsv(0.83, 0.21, 0.96))`

`barplot(x, col="Plum1")`

RGB to HSV converter

색관련 모수 (Parameters of Color)

`col`

기호, 선, 문자등의 색상을 디폴트로 지정
(default plotting color)

`col.axis`

축의 색 지정
(Color for axis annotation)

`col.lab`

x축과 y축의 Label색 지정
(Color for x and y labels)

`col.main`

제목 색 지정
(Color for main title)

`col.sub`

부제목의 색 지정
(Color for sub titles)

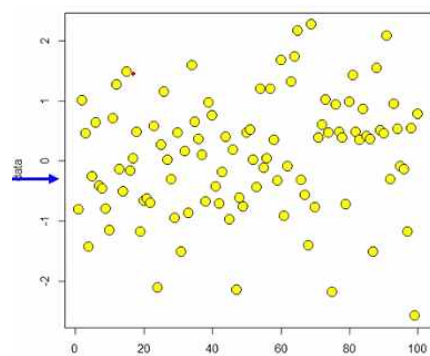
`fg`

그래프 전경 색 지정
(Color for foreground)

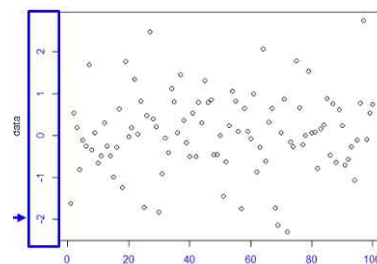
`bg`

그래프 배경 색 지정
(Color for background)

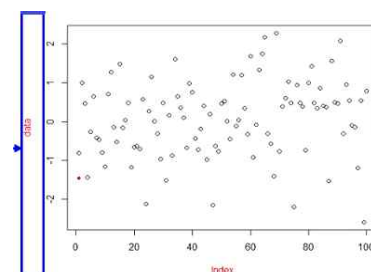
`plot(data, cex=1, pch=21, bg="yellow",
mian="점의 컬러 변경")`



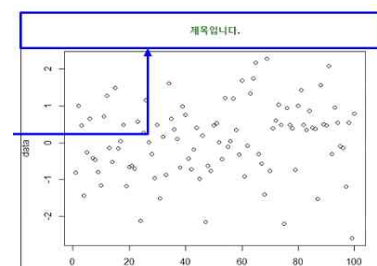
`plot(data, cex=1, pch=21, col.axis="blue", main="col.axis=blue")`



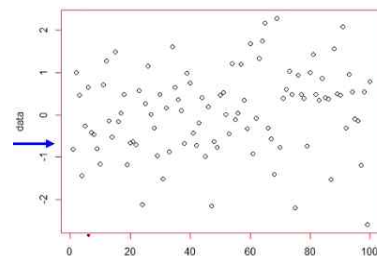
`plot(data, cex=1, pch=21, col.lab="red", main="col.lab=red")`



`plot(data, cex=1, pch=21, col.main="Darkgreen", main="제목입니다.")`



`plot(data, cex=1, pch=21, fg="red", main="그래프 전경색 바꿈")`



R colors

R에서 지원하는 색의 종류

657개

`colors()`함수

657개 전체 색 리스트가 화면에 뿌려짐

색과 숫자를 함께 정리한 사이트를 참조하면 편리함

Earl F.Glynn

RcolorBrewer함수

R이 제공하는 색상관련 패키지

RColorBrewer패키지 설치

```
install.packages("RColorBrewer")
```

```
library(RColorBrewer)
```

RColorBrewer패키지 안의 모든 색상 팔레트 보기

display.brewer.all() : 패키지 안에 있는 모든 색상의 팔레트를 보여줌

기타 : brewer.pal.info : 팔레트의 종류와 총 색상수와 카테고리 종류를 보여줌

분기적(Diverging)팔레트 / 정성적(Qualitative)팔레트 / 연속적(Sequential)팔레트

Par함수 : 차트를 mfrow에 맞게 행열을 나누어 보여줌

```
par(mfrow=c(1,3))
```


ggplot2란?

통계적 기법을 활용하여 고급 그래프 생성이 가능한 시각화 패키지

ggplot2의 장점

1. 하나의 문법으로 다양한 그래프를 그릴 수 있음
2. `install.packages(ggplot2)`, `library(ggplot2)`로 패키지를 쉽게 불러낼 수 있음
3. 데이터간의 이해관계를 살피는데 최적화 되어있음
4. 통계적 기법을 활용하여 고급 그래프 생성가능

ggplot2의 문법구조

- 데이터프레임(Data Frame)
- 색상, 크기 등의 외적요소(aes)
- 점, 선, 모양 같은 기하학적 요소(Geoms)
- 통계적 처리방법(Stats)
- aes에서 사용할 스케일(Scale)

산점도 함수

데이터 값의 (x,y)좌표에 기호를 표시하여 산점도를 만들 때 각 관측치에 대해 y축에 반응변수, x축에 예측 변수를 표시하기위해 필요한 함수

산점도

계량형 변수 쌍 간의 잠재적 관계를 조사하는데 사용할 수 있는 그래프

산점도 함수의 회귀직선 그리기

```
>kline <- lm(cars$dist~cars$speed, data=cars)
```

lm(linear mode)함수란? 단순선형회귀를 기준으로 $y=ax+b$ 라는 회귀함수(회귀곡선)를 구하여 독립변수(x)에 따른 종속 변수(y값)를 예측하는 모델링

안전지도 데이터 시각화 방법

ggmap 패키지의 활용

```
install.packages("ggmap")
```

```
library("ggmap")
```

```
geocode('korea', source='google')
```

/ geocode()함수 : 대상지역의 경도(lon.)와 위도(lat.)를 source로부터 가져오는 함수

geom_point()	지도에 점을 찍을 때 사용되는 함수	
geom_text()	점 주변에 글자를 넣을 때 사용되는 함수	
zoom	지도 표시 레벨 설정	
size	만들어진 점의 크기 설정	size 와 cex의 차이점?
alpha	만들어진 점의 투명도 설정	
label	글자를 입력할 때 사용하는 옵션	