

파이썬도 있고 상업용 패키지도 많은데 왜 R로 공부하냐고?
엑셀이 있는데 왜 학교에서 수학 공부하냐?라고 묻는 것과 같다. ^^

데이터 불러들이기

```
raw_data <- read.csv("u.data", sep = "\t", header = F)
```

read.csv 함수는 텍스트 화일에도 적용이 가능함

도서관에서 필요한 패키지를 불러들이기

(통계분석을 위해서는 굳이 프로그래머가 될 필요가 없다
프로그래머가 만들어 놓은 패키지를 laibrary에서 불러와서
잘 사용하는 것이 경쟁력이다.)

```
install.packages("recommenderlab")  
library(car) 애플 부르면 여러가지가 편리하다  
some(raw_data) 무작위로 자료를 추출해서 볼때
```

R은 대문자와 소문자를 구분한다.

R을 설치할 때 "관리자 권한"으로 해야 한다. (메모리에 패키지를 불러 들이는 문제)

R에 사용되는 함수나 패키지 등에 대해 검색할 때는 검색 창에 "검색내용 in R"이라고 입력한다.

명령어를 익히는 가장 좋은 방법은 R console에서 ? 나 ?? 뒤에 명령어를 입력하고 검색 결과의 예시를 직접 실행해 보는 것이다.

확장자가 .R인 스크립트 화일은 R의 메뉴에서 "파일 => 스크립트 열기"로 불러들인다. 즉, 스크립트를 저장할 때는 .r을 붙이는 것이 좋다.

R의 패키지는 버전에 따라 작동을 잘 하는 경우가 있고 오류가 있는 경우가 있다.

Window 버전은 여러가지의 R버전을 같이 사용하여도 된다.

R이 "작업공간을 저장 할까?" 물어보면 저장을 하지 않는 것이 좋다.(메모리 문제)

스크립트만 주석을 잘 달아서(# 뒤에 입력) 저장해 놓으면 다시 시행하는데 문제가 없다.

엑셀의 명령어 자동 완성과 같이 코드의 일부를 치고 탭키를 누르면 자동 완성이 된다.

```
head(as(ratings_mattix[1:5, 1:5], "matrix")) 매트릭스 자료 미리보기  
image(ratings_matrix[1:5, 1:5]) 그림으로 그려서 보여주기  
View(Vector) 벡터로 보기  
sort(unique(ratings$rating), decreasing = F) 하나의 열만 정렬해서 보기
```

결과를 파일로 내보내기

```
write.csv(pred-matix, "pred-matrix.csv")  
write.csv(data, file="data.txt", append=F) append 옵션이 T면 기존의 파일에 추가하여 쓰고, F면 파일을 갈아 엮으라는 것
```

변수조건을 주어 데이터의 일부만 사용하기

```
data[data$age>30, ] data라는 데이터 화일의 age 변수의 값이 30을 초과하는 데이터만 보기
```

내장된 데이터 부르기

```
data(iris)
```

<< R 함수 / 명령어 리스트 >>

\$ 데이터에서 변수지정하기

..... 주석을 달때 사용

% % 나머지 구하기

% * % 행렬의 곱

% / % 정수 나누기

^ 승수 구하기

** 승수 구하기

& and 조건

| or 조건

() 대입결과 출력

? 함수명 설명

?? 함수에 대한 설명 문서 찾기

WWW 외부의 자료 읽을 때 디렉토리 표시

== 벡터비교

a o v 분산분석

abline 산점도에 회귀선 추가, 수직선 수평선 추가

acf 시계열의 자기상관 함수

acos 코사인의 역함수

adf.test 평균회귀성 검사

aggregate(sales~location, wine, sum) wine 화일에서 지역별로 판매량을 합을 구함

aggregate(sales~fruit+year, fruits, max) fruits에서 년도별, 과일별로 최대값 구하기

피봇에 준하는 자료정리

annotate plot 함수에서 각 선의 주석 옵션 선택 TRUE FALSE

anova(values~ind, data) 분산분석 ind 집단별 분산분석

aperm 배열의 전치

aplpack 다양한 그래프 package

append 벡터에 데이터 추가

append(vec, 15, after=5) vec의 5번째 뒤에 15를 추가

apply(df[, c(1,2)], 2, sum) 컬럼(열)단위 합계

apply(x, 1, max) 행 단위 최대값([1] 7 8 9)

apply(x, 1, min) 행 단위 최대값([1] 1 2 3)

apply(x, 2, mean) 열 단위 평균값([1] 2 5 8)

apply.daily 달력 주기 함수 적용

apriori 연관성 분석

apropos 오브젝트의 검색 apropos("^chi")

args(함수명) 함수명의 쓰임 방법

arima 아리마 모형 적합

array(c(1:12), c(3,2,2)) 3행2열 구조 2개

as.character() 문자형으로 변형

as.complex() 복소수형으로 변환

as.data.frame() 데이터 프레임으로 변환

as.Date 날짜 형식으로 변환 as.Date("2019-03-01")

as.Date("03-10-2018", format='%d-%m-%Y')

%d 일자를 숫자로 인식

%m 월을 숫자로 인식

%b 월을 영어 약자로 인식

%B 월을 전체 이름으로 인식

%y 연도를 숫자 두 자리로 인식

%Y 연도를 숫자 네 자리로 인식

as.double() 실수형으로 변환

as.factor() "범주형으로 변환, 특히 문자형을 plot하려는 경우 factor로 변환해야 가능"

as.integer() 정수로 변형

as.list() 리스트 형식으로 변환

as.logical() 논리형으로 변환

as.matrix 행렬 전체를 하나의 값으로

as.numeric(변수명) 자료 변형 숫자로

as.numeric 하면 자동으로 문자값을 숫자로 코드 변경 해 준다.

as.undirected 연결망 그래프 선 정리

as.vector() 벡터 형식으로 변환

asin 사인의 역함수

atan 탄젠트 역함수

attach data file을 불러서 사용, detach()로 해소함

attribute 오브젝트의 속성 확인 attribute(x)

available.packages() R이 가지고 있는 package목록

average 평균

barplot 막대 그래프

barplot(, col,)..... 막대그래프 색 지정

barplot2 막대그래프 신뢰구간 추가

biplot 주성분 분석 그래프

Box.test 자기상관 검정

boxplot 박스 플롯 그리기 boxplot(score ~ group)

by 집단별 함수 적용 by(traits, traits\$sex, summary)

c(1:20) combine value 함수, 1:20과 동일

cast 피벗팅 자료 변환

cat 여러 항목 출력 (print는 한 개 항목 출력)

cbind(x1,x2) 열 묶음, 추가로 matrix 객체 생성

ccf 두 시계열의 시차상관

ceiling 값의 올림

character "문자, 문자형 (plot을 그리기 위해서는 범주형으로 변환)"

chiq.test 두 변수의 독립성 검증

choose 조합의 수(nCr) 구하기

class 변수의 수준 확인

class "클래스(날짜,회계 등) 확인"

cmdscale 다차원 척도

coef 모형계수 점추정치

colnames 열 이름 부여

colSums 열의 합

combn 조합의 생성 nPr

confint 모형계수 신뢰구간

control + enter script 문에서 문장별 실행

coplot 요인별 산점도 그리기

coplot(horsepower ~ mpg | origin, data = Car93)

cor 상관계수

cor.test 상관계수의 유의성 검정

cor.test(x, y, method="spearman") 정규분포가 아닌 경우

Corpus 텍스트 마이닝 문서 읽기

cos 코사인 함수

cov 공분산

ctree 의사결정 나무

cumsum 누적합계

curve 함수를 그래프로 그리기 curve(sin, -3, 3)

cut , 데이터를 구간별로 나누기

cutree , 군집 추출

data() , R에 저장된 기본 data

data(iris) 내장된 데이터는 별도로 부르지 않고도 명령어가 수행이 된다. 다만 다른 화일명으로 바꿀때 이 함수를 사용

data.frame(x=c(1:5), y=seq(2,10,2), z=c('a','b','c','d','e')) , 데이터 프레임 만들기

data.frame(NO=no, Name=name, Pay=pay) " , 컬럼명 지정

dbimon , 이상분포 확률

dbscan , 밀도 기반 군집분석

ddply , 데이터 입출력 함수

decompose , 시계열 자료 분할

degree , 연결망 연결 방향 분석

demo() 현재 실행 패키지의 데모

demo(패키지명) 패키지의 데모

detach , 데이터 프레임 제거

det(A) 행렬의 값

dev.off() 출력할 파일 닫음

deviance , 잔차 제곱합

diag , 행렬의 대각선 값

Dictionary , 분석 대상 단어의 집합

diff , 구간 확률 계산

diff , 시계열 차이계산

dim , 행렬 설정

dim(available.packages()) 13046 17 -> cran site에서 제공하는 package정보

do.call , 여러 개의 행 조합

dwtest , 잔차의 자기상관

edit(df) , 데이터 편집기

example(함수명) 함수 사용 예

exp 지수 함수

expand.grid , 가능한 조합의 생성

faces , 얼굴 그래프

factanal , 요인분석

factor , 요인으로 변환

findAssocs , 같은 문장/문서에 나오는 연관성 있는 단어 찾기 findAssocs(tdm1, "hate", 0.5) 같이 나올 확률 (같은 문장에 나온 빈도 ÷ 전체 문장에 나온 빈도)

findFreqTerms , 일정 빈도 이상의 단어 찾기

first , 가장 오래된 자료

floor 값의 버림

format , 날짜를 문자열로 변환

function , 함수를 정의 하는 방법

gc 가비지 콜렉션의 실행

gdata , 워드 클라우드 작성

getwd(d:/rwork) , R의 기본 작업 디렉토리 확인

ggplot2 , 그래프 작성 package

glm , 로지스틱 회귀

googleVis , 움직이는 그래프 package

grep , get regular expression, 표현식이 있는 곳의 위치를 가르쳐 줌

grid , 그래프에 격자 추가

gsub , 모든 문자열 대체

hclust 군집분석
head(변수명) 데이터 앞의 6줄 보기
head(as(data, "matrix"), 10) matrix로 된 데이터 확인하기
help(함수명) 도움말
help(package = "recommenderlab") 패키지의 사용에 대한 자세한 내용 보기
hist(rnorm(20)) 난수에 대한 히스토그램 그리기
hist(데이터명) 히스토그램 그리기
history() 커맨드 히스토리
Hmisc 고급 데이터 분석 package

in.integer() 정수형 여부
influence.measures 영향력이 큰 관측값
inspect 데이터 내용 확인
install.packages('패키지명') 패키지 설치, 같은 컴퓨터에서는 한 번만 설치해도 됨
installed.packages() 현재 시스템에 설치된 전체 패키지 보임
interaction.plot 변수간의 상호작용
"intersect(x, y) " 벡터 x와 y에 공통으로 존재하는 값 - 교집합
IQR 사분위 범위
is.character() 문자형 여부
is.complex() 복소수형 여부
is.data.frame() 데이터 프레임 여부
is.double() 실수형 여부
is.factor() 범주형 여부
is.logical() 논리형 여부
is.na() 결측치 여부
is.nan() 비숫자 여부
is.numeric(변수명) 수치형 여부
ISOdate 년월일 벡터 처리

JRE 한글 텍스트 마이닝을 위한 환경
julian 율리우스력(70.01.01이후)

kmeans k means 군집분석
KoNLP 한글 텍스트 마이닝 package
kruskal.test 중앙값들간의 차이 검증
ks.test 표본의 동일집단 추출 확인

lag 시계열 시간변경
lapply 텍스트만 추출
lapply(c(list(1:5),list(6:10)), max) list 데이터 처리 함수 - list로 결과 반환
las label axis style 수평축 라벨(값 이름) 스타일 : 1 평행, 2 수직
last 가장 최근 자료
legend 그래프 범례 추가하기
legend(1,5, 2,4, as.character(levels(f)), pch=1:length(levels(f)))
length(member) 변수의 길이(벡터의 길이)
library('패키지명') 패키지 로딩, install한 패키지를 사용하기 위해 불러옴
limits 그래프의 범위지정 옵션
lines 여러 개의 데이터 셀 그래프
lines(density(x)) 히스토에 밀도 추가하기
list("lee", "이순신", 95, "hong", "홍길동", 85) 함수 내에서 여러 값을 하나의 키로 묶어서 반환할 경우 유용함
list(name="hong", age = 35, address="한양", gender="남자")
lm 변수값을 계산해서 회귀
lm(+0) 절편없는 선형회귀

lm(subset). 부분 집합에 대한 회귀모형
lm(poly). 다항식 회귀분석
lm(log). 변환자료 회귀분석
load RDdata 읽어 들이기
log 밑을 e로 하는 자연로그
log10 밑을 10으로 하는 로그함수
log2 밑을 2로 하는 로그 함수
log1p 1을 더한 결과의 자연로그
logic "참 거짓 TRUE, T, FALSE, F"
lpSolve 최적화 package
ls() 현재까지 저장한 변수 목록 rm 변수 삭제 objects 변수 확인
lty 그래프 선의 유형 lty=c("solid", "dashed", "dotted", "dotdash", "longdash", "twodash")
lubridate ... 날짜 데이터 처리를 위한 package
lwd 선의 두께

Map 데이터 프레임 변형
mapply 다수의 인자를 함수에 넘김
mat 행/열의 선택
match 특정값 찾기
matrix(10:19,2)10개 데이터를 2행으로 객체 생성
matrix(c(1,"hong",150,2, "lee", 250,3, "kim", 300) ,3 ,by=T) data frame 생성, 행우선, 3개 리스트 생성
matrix(c(1:10), nrow=2) 열 우선 2행2열 생성
matrix(c(1:10), nrow=2, byrow=T) by=T : 행 우선
matrix(c(1:5)) 열 기준으로 행렬 생성
matrix(c(1:9), nr=3, nc=3) 열우선 3행 3열 지정, nr nc로 축약 사용
mean(변수명) 기초통계량
median 중앙값
merge 데이터 통합(Vlookup) 시계열 넣기 시계열 병합
merge(live, dead, by="name")
transaction <- merge(buy, buygoods, by="buy_code")
mfrow 그래프 창 지정 par(mfrow = c(2, 2))
mode 변수의 모드 확인 mode(x)
mode(x) <- "integer" 모드의 변경
mode(변수명) 변수명의 데이터 유형
mosaic 모자이크 그래프
multiple axis 다축 그래프 그리기

NA not available 결측치, 데이터가 있지만 활용이 불가능한 것
names 벡터의 원소에 이름을 지정
na.omit NA값 포함 행 삭제
NaN 비 숫자
na.strings data file에 있는 값을 R의 NA로 인식 na.strings=c("NA")
nchr 문자열의 길이
ncol 열의 갯수
nrow 행의 갯수
newCol 데이터에 열 추가하기
newRow 행에 자료 추가하기
null 데이터가 없는 것
numeric "정수, 실수"

objects() 생성한 변수들을 확인
oneway.test 일원 분산분석
opinion 감성분석 오피니언 마이닝

options(width , 가로 넓혀서 출력
order , 데이터 정렬기준
outer , 문자열의 조합
outlier.test , 이상치 탐색
ovid , 텍스트 마이닝 데이터 조회

pacf , 부분 자기상관 함수
pairwise.test , 집단의 모평균 쌍비교
par , 그래프 매개변수 변경
par(ask=TRUE) , 여러 개의 그래프 일시중지
par(mfcol , 한 페이지에 여러 개 그래프
par(mflow , 한 페이지에 여러 개 그래프
party , 의사결정 나무 package
paste , 문자열 합치기
pbinom , 이산 밀도함수
pdf("'d:/Rwork/batch.pdf'") , 지정된 경로의 파일에 결과 출력
pexp , 지수분포 확률 계산
plot(,pch , 각 쌍별 산점도 생성
plot,col="" "") , 선그래프 선의 색상
plot(,lty"" "") , 선그래프 선의 종류
plot(,lwd="" "") , 선그래프 선의 두께
plot(type=""h'') , 이산 히스토 그래프
plot,type=""l'') , 선그래프
plot(quantileFunction , 분위수 그래프
plot(TurkeyHSD , 집단간 평균의 차이
plot(변수명) , 그래프 그리기
plot(x, ann = F) , 주석 금지
plotcluster , 숫자로 대체한 군집 그래프
plyr , "데이터 분리, 적용, 조합 package"
pmax , 쌍에서 최대값
pmin , 쌍에서 최소값
pnorm , 정규분포 확률 계산
points , 낮은 수준의 산점도
POSIXlt , 날짜의 일부 추출
prcomp , 주성분 분석
pred[1:5, ,] 매트릭스 자료 보는 법, 화일명 ..
predict , 각 모형을 기반으로 예측값 구하기
predict(,interval) , 회귀모형 예측 구간
princomp , 주성분 분석
print , 출력
print(dframe) , 데이터프레임의 기본통계량
prop.table(분할표, 1) 행비율
prop.table(분할표, 2) 열비율
prop.table(분할표) 전체 합비율
prop.test , 두 집단의 동일비율 검정, 모비율 검정
prop.test(11, 20, 0.5, alternative="greater", conf.level=0.95) 20경기에서 11번 승리한 경우 승률이 50%보다 크다고 할 수 있는가?

q() , 작업종료
qqline , 정규분포 확인
qqnorm , 정규분포 확인
quantile , "1,3분위수"
quantile(vec, .05) quantile(vec, c(.05, .95))

randomForest "의사결정 나무, 대용량 데이터"
rbind(x1,x2) 행 묶음, 추가로 matrix 객체 생성
rbinom 이항분포 난수
rcorr 변수값의 순위를 고려한 상관계수
read.csv("http://www.data.co.kr/student.csv", header=TRUE) 웹에서 파일 불러오기
read.csv(file="student4.txt", sep=";", na.strings="-") csv 화일 읽기
read.csv('화일명') 기본 디렉토리에 저장된 csv화일 불러오기
read.table 텍스트 파일 읽기, text file의 마지막은 enter로 입력을 끝내야 에러가 나지 않는다
read.table(file="student1.txt", header=TRUE) 컬럼명이 있는 파일 불러오기
read.table(file="student2.txt", sep=";", header=TRUE) "구분자(세미콜론, 탭)이 있는 경우의 파일 불러오기"
read.table(file="student3.txt", sep=" ", header=TRUE, na.strings="-") 특정 문자 NA처리
readline() 한 줄 단위로 입력 받기
readLines 복잡한 파일 읽기
read.spss spss화일 읽기(패키지 foreign)
rebind 자료와 합을 같이 출력
remove.packages('패키지명') 패키지 삭제
removeSparseTerms 희소성의 기준으로 자료 제거
reorder 정렬
rep(1,5) 11111
rep(1:3, 3) 반복 1 2 3 1 2 3 1 2 3
rep(1:3, each=3) 1 1 1 2 2 2 3 3 3
rep(c(6,9),3) 6 9 6 9 6 9
reshape 데이터 구조 전환
reshape2 데이터 구조 가공 도구
resid 모형의 잔차
resid 시계열 추세 없애기
rm 변수삭제
rm(list=ls()) 모든 변수를 삭제
rnorm 정규분포 난수 생성 rnorm(100) 100개 생성
rollapply 롤링 함수
rollmean 이동평균 계산
round 값의 반올림
round(dist 맨하탄 거리
round(dist 유클리드 거리
rownames 행 이름 부여
rowSums 행의 합
rpart "의사결정 나무, 예측오차 최소화"
runs.test 이진 수열의 랜덤성 검증 (package : tseries)
runif 난수생성(균등분포)

sample 표본 추출 sample(data\$year, 10, replace=T) 복원추출
sampleBy 표본의 추출 (package : doBy)
sapply(c(list(1:5),list(6:10)), max) list 데이터 처리 함수, Vector로 결과 반환
save(data, file="화일명")..... RData 파일로 저장
save.image 작업공간 저장
savePlot("급여현황.png", type="png") 그래프 파일에 저장
scale 자료의 정규화
특정값의 정규화 (x1 - mean(x)) / sd(x)
scan() 화면에서 숫자 입력 받기
scan(what="") 화면에서 문자 입력 받기
scan(what=character(0)) 문자 입력(위와 같음)
sd 표준편차
search() 현재 로딩된 패키지 목록

searchTwitter 트위터 관련 메시지 가져오기
seq(1,10,2) 1부터 10까지 2씩 증가
sequence(10) 1부터 정수로 10까지
seq(1:10) 1부터 10까지
set.seed 매번 같은 난수 생성
setdiff(x, y) 벡터 x에는 있는데 y에는 없는 값 - 차집합
setwd('주소') "R의 기본 작업 디렉토리를 ""주소""로 변경"
shapiro.test 정규성 검증, 귀무가설 : ~에 따른다, $p > 0.05$ 정규분포를 따른다.
shiny 샤이니 그래프 package
shortest.paths 연결망 최단거리 파악
sin 사인 함수
slope, intercept 단순 직교회귀
SMA 시계열 분석 package
smooth 스무스 그래프
solve(A) 역행렬
split 행렬에서 열 분리
sqldf sql 문장 사용 package
stack 집단 합치기
stars 별 그래프
stem, leaf 줄기잎 그래프 stem(data, [group==1])
step(backward 후진 단계별 회귀모형
step(both 단계별 회귀모형
step(forward 전진 단계별 회귀모형
storage.mode 오브젝트의 스토리지 모드 확인
str 구조와 내용 확인 str(x)
str_extract("abcd12aaa33", "[1-9]{2}") 연속된 숫자2개가 첫번째 발견 항목
str_extract() 문자열의 위치(index) 리턴
str_extract_all("abcd12aaa33", "[1-9]{2}") 모두
str_extract_all('데이터', '[조건]{추출할 글자 개수}') 데이터 중에 조건에 만족하는 데이터를 {개수}만큼 추출
str_join 문자열 연결
str_length () 문자열 길이 리턴
str_locate() 문자열에서 특정 문자열 패턴의 첫번째 위치 찾기
str_locate_all() 문자열에서 특정 문자열 패턴의 전체 위치 찾기
str_replace(d, "김길동", "홍길동") 문자열 교체
str_split() 기준문자를 중심으로 부분 문자열 리스트 출력
str_sub("abcd12aaa33", 3,6) 서브스트링 생성
str_sub() 범위에 해당하는 부분 문자열 출력
stringr str_ 의 함수를 사용하는 패키지
stringsAsFactors 문자로 입력되는 열은 기본적으로 요인으로 인식 이때 FALSE로 옵션을 부여하면 문자로 인식한다
sub 첫번째 문자열 대체
subset "행/열의 일부를 선택, 부분집합"
subset(df, x>=2 & y<=6) 2개 조건이 참인 레코드 대상 서브셋 생성
subset(df, x>=3) x가 3이상인 레코드 대상 서브셋 생성
subset(df, y<=8) y가 8이하인 레코드 대상 서브셋 생성
substr 문자열 추출 substr("아라리요" , 1, 2)
sum(변수명) 합
summary 통계정보 요약
summary(lm(..... 회귀 모형의 통계량
summaryBy 조건에 따라 그룹별로 작업 (package : doBy)
summary(table(initial, outcome)) 두 변수의 독립성 검증, 카이스퀘어
sweep 여러개의 데이터에 동일한 기준 적용
Sys.Date() 오늘의 날짜
Sys.time() 시스템 날짜

system.time , 시스템 소요 시간

t.test , t 검정, 모평균 추정, 두 집단의 평균비교 t.test(x, mu = 80)

t.test , 모평균의 신뢰구간

table , "도수분포표, 분할표, 교차표"

table(x) 변수가 하나인 경우의 빈도표

table(data\$x,data\$y,deparse.level=2) 변수가 2개인 분할표

xtabs(~x+y, data) 변수 2개

addmargins(xtabs(~x+y,data)) 레이블 추가

xtabs(freq~., data) 빈도와 x y 값이 주어지는 경우

xtabs(cbind(합,불)~., data) 각각에 두 가지가 주어진 테이블

tail , 데이터 뒤의 일부보기

tan 탄젠트 함수

tapply , 군집별 평균

tapply , 집단별 데이터 처리

TermDocumentMatrix , 텍스트 마이닝 행렬 전환

title , 그래프에 라벨 추가

tm , 텍스트 마이닝 패키지

tm_map , XML문서를 text로 전환

tolower , 대문자 소문자 변경

trim , 공백 문자 제거

ts(kings) , 시계열 데이터 선언

plot.ts(king,ts) 그래프 확인

AR 자기회귀모형 : p 시점 이전의 영향을 받음

ACF 자기상관함수의 결과가 빠르게 감소하고 PACF 부분자기상관함수의 결과가 n 시점에서 절단점(점선 밖)을 갓으면 AR(n-1) 모형

MA 이동평균모형 : 유한한 갯수의 백색잡음의 결합

PACF가 빠르게 감소하고 ACF가 절단점을 가짐

ARIMA(p,d,q) 자기회귀누적이동평균

t.test(x,y) 평균차이 검정

TTR , 시계열 분석 package

twitter , 트위터 분석 패키지

twitter , 트위터 자료 받기 (감성분석)

typeof 오브젝트 유형의 확인 typeof(x)

unlist(data1)리스트 형식을 해제하고 Vector 형태 출력

unstack , "열 분리, 프레임 변경"

var , 분산

vcd , 모자이크 플롯 그래프 package

vignette(package="grid") , 패키지의 간이 매뉴얼

which , 두 데이터 필드간의 값 비교

which,max , 최대값 찾기

which,min , 최소값 찾기

wilcox.test 비모수적 비교실험 분석

wilcox.test 중앙값 median 의 신뢰구간

wilcox.test(x, conf.int=T, conf.level=0.95)

win.graph() , 그래픽스 창 추가로 열기

with 열의 이름 꺼내기

with(iris, plot(x, y, pch = as.integer(group))) 그룹별로 다른 모양으로 그래프 그리기

wordcloud , 워드클라우드 package

wordle , 단어 빈도 시각화

write.csv , 파일 저장 write(data, "data.csv")

writeCorpus 텍스트 마이닝 결과 저장

writeXLS 화일을 엑셀 화일로 저장하기 위한 패키지 (윈도우에는 perl이 설치 되어 있어야 가능)

zoo, xts 시계열분석