

목차

- 첫째마당

- 안녕, R?

- 데이터 분석을 위한 연장 챙기기

안녕, R?

R이 뭔가요? - R 이해하기

R ?

- . 오克兰드 대학의 로스 이하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)이 1992년 처음 공개
- . 전 세계에 걸쳐 데이터전문가들이 애용하는 통계 프로그래밍 언어이자, 계산 및 시각화를 지원하는 데이터 분석 프로그램

R의 특징

- 프로그래밍 언어
 - R은 컴퓨터에 어떤 일을 시키려고 컴퓨터와 의사소통을 할 수 있게 만든 프로그래밍 언어
 - 프로그래밍 언어는 문법과 체계가 있으며, 코드를 작성하고 실행하면 결과가 나오는 프로그램을 의미
 - R 역시 문법에 맞추어 코드를 작성하고 실행하면 결과가 나옴

R의 특징

- 프로그래밍 언어
 - R은 인터프리터 언어(interpretive language)라고도 함
 - 인터프리터 언어란 사용자가 코드를 입력하면 컴퓨터가 대기하고 있음
 - 코드 내용을 번역해서 사용자가 결과값을 바로 확인 할 수 있게 출력하는 프로그래밍 언어
 - 사용자는 코드를 작성하여 실행 버튼만 누르면 바로 실행 결과를 확인할 수 있음

R의 특징

- 데이터 처리와 그래프 그리기, 통계 분석에 특화
 - R은 통계학자들이 개발함
 - 통계학자의 아이디어와 개념이 R에 고스란히 녹아 있음
 - 이것은 R을 활용한다면 데이터 분석을 위한 기초 체계와 이론을 자연스럽게 습득할 수 있다는 의미

R의 장점

- 최신 분석 기법
- 비정형 데이터 분석
- 무료 ⇒ 많은 사용자 ⇒ 공부하기 좋음.

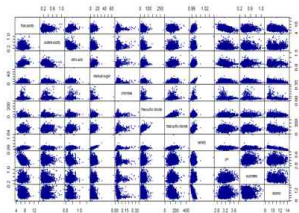
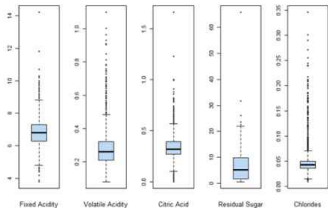
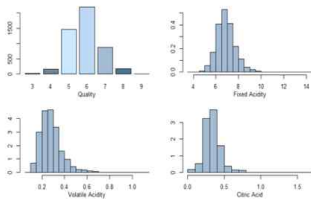
R의 단점

- . No Manual, No A/S
- . 영어의 압박

R 어디에 사용하나?

EDA (Exploratory Data Analysis)

- 데이터를 본격적으로 처리해서 결과를 얻는 작업을 하기 전에 어떤 의미 있는 데이터가 있을지 확인하는 탐색적 데이터 분석(Exploratory Data Analysis, EDA)을 진행하는데, R은 이런 작업을 하기에 편리한 환경을 제공함



	Minimum	Q1	Median	Q3	Maximum	Range	IQR	MAD	Mean	StDev	SErr
fixed.acidity	3.00	6.20	6.00	7.20	14.20	10.40	1.00	0.74	6.85	0.94	0.012
volatile.acidity	0.08	0.21	0.26	0.32	1.10	1.02	0.11	0.09	0.28	0.10	0.001
citric.acid	0.00	0.27	0.32	0.39	1.66	1.66	0.12	0.09	0.33	0.12	0.002
residual.sugar	0.60	1.70	3.20	9.90	65.00	65.20	8.20	5.54	6.39	5.07	0.072
chlorides	0.01	0.04	0.04	0.05	0.55	0.54	0.01	0.01	0.05	0.01	0.000
free.sulfur.dioxide	2.00	25.00	34.00	46.00	209.00	207.00	21.00	16.31	35.51	17.01	0.245
total.sulfur.dioxide	9.00	100.00	134.00	167.00	440.00	431.00	59.00	43.00	130.36	42.50	0.607
density	0.99	0.99	0.99	1.00	1.04	0.05	0.00	0.00	0.99	0.00	0.000
pH	2.72	3.09	3.18	3.28	3.82	1.10	0.19	0.15	3.19	0.15	0.002
sulphates	0.22	0.41	0.47	0.55	1.08	0.86	0.14	0.10	0.49	0.11	0.002
alcohol	0.00	9.50	10.40	11.40	14.20	6.20	1.90	1.48	10.51	1.22	0.010

머신 러닝

머신러닝

정답의 유무에
따른 구분

지도학습

비지도학습

학습 목적에
따른 구분

분류

회귀

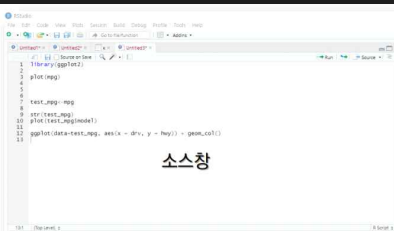
군집화

연관규칙
분석

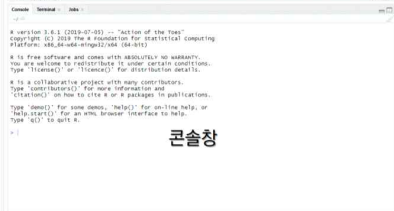
텍스트 마이닝



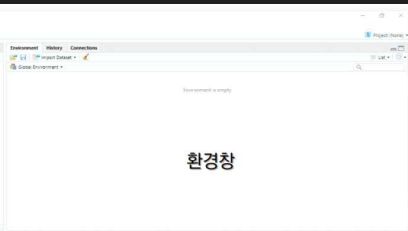
R과 R 스튜디오 설치하기



소스창



콘솔창



환경창



파일/플롯/패키지/헬프/뷰어

R 스튜디오 기초 사용법

R Studio 환경설정

- 워킹 디렉토리 : R에서 파일을 불러오거나 저장할 때 사용하는 폴더
- 프로젝트 만들기 : 소스코드, 이미지, 문서, 외부프로그램에서 생성된 자료 등 수많은 파일을 활용하게 되는데, 파일들을 프로젝트 폴더별로 관리하면 편리함

R Studio 추가설정 확인 사항

Options

 General Code Appearance Pane Layout Packages R Markdown Sweave Spelling Git/SVN Publishing Terminal

RStudio theme:

Modern ▼

Zoom:

125% ▼

Editor font:

Consolas ▼

Editor font size:

14 ▼

7

8

9

10

11

12

13

14

16

18

24

36

Interbivore

Merbivore Soft

Mono Industrial

Monokai

Pastel On Dark ▼

Add...

Remove

```
# plotting of R objects
plot <- function(x, y, ...)
{
  if (is.function(x) &&
      is.null(attr(x, "class")))
  {
    if (missing(y))
      y <- NULL

    # check for ylab argument
    hasylab <- function(...)
      !all(is.na(
        pmatch(names(list(...)),
                  "ylab"))))

    if (hasylab(...))
      plot.function(x, y, ...)
  }
  else
    plot.function(
      x, y,
```

OK

Cancel

Apply

Options



General



Code



Appearance



Pane Layout



Packages



R Markdown



Sweave



Spelling



Git/SVN



Publishing



Terminal

RStudio theme:

Modern

Zoom:

125%

Editor font:

Consolas

Editor font size:

14

Editor theme:

Mono Industrial
Monokai
Pastel On Dark
Solarized Dark
Solarized Light
Textmate (default)
Tomorrow
Tomorrow Night
Tomorrow Night 80s
Tomorrow Night Blue
Tomorrow Night Bright
Twilight
Vibrant Ink
Xcode

Add...

Remove

```
# plotting of R objects
plot <- function (x, y, ...)
{
  if (is.function(x) &&
      is.null(attr(x, "class")))
  {
    if (missing(y))
      y <- NULL

    # check for ylab argument
    hasylab <- function(...)
      !all(is.na(
        pmatch(names(list(...)),
                  "ylab"))))

    if (hasylab(...))
      plot.function(x, y, ...)
    else
      plot.function(
        x, y,
```

OK

Cancel

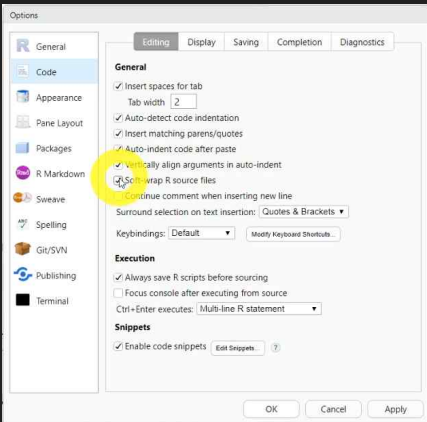
Apply

Soft-wrap 자동 줄바꾸기 옵션 설정

명령어가 길어지면

화면 밖으로 넘어가지 않고

자동으로 줄을 바꿈



Options



General



Code



Appearance



Pane Layout



Packages



R Markdown



Sweave



Spelling



Git/SVN



Publishing



Terminal

Editing

Display

Saving

Completion

Diagnostics

General

☒ Highlight selected word☒ Highlight selected line☒ Show line numbers☐ Show marginMargin column ☐ Show whitespace characters☐ Show indent guides☒ Blinking cursor☐ Show syntax highlighting in console input☐ Allow scroll past end of document☒ Allow drag and drop of text☐ Highlight R function callsFold Style:

Console

Limit length of lines displayed in console to: ANSI Escape Codes:

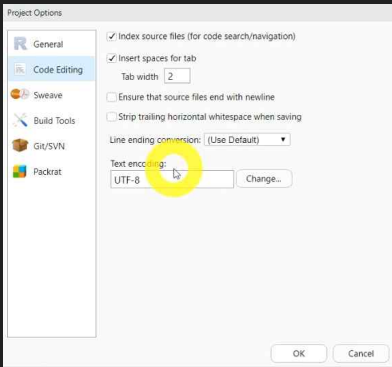
OK

Cancel

Apply

인코딩 방식

글자(한글)가 안깨지게 인코딩을 변경



안녕, R? - 친숙해지기

내용 의미 무관하게 일단 실행해 보기

패키지 설치

```
install.packages("dplyr")  
install.packages("ggplot2")
```

패키지 로드

```
library(dplyr)  
library(ggplot2)
```

데이터 검토

```
head(mpg)
```

```
> # 데이터 검토
```

```
> head(mpg)
```

```
# A tibble: 6 x 11
```

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
	<chr>	<chr>	<dbl>	<int>	<int>	<chr>	<chr>	<int>	<int>	<chr>	<chr>
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact

```
>
```

dim()

몇 행, 몇 열로 구성되어 있는지

```
> dim(mpg)
[1] 234  11
> |
```

str()

데이터에 들어 있는 변수들의 속성을 보여줌

```
> str(mpg)
tibble [234 x 11] (s3: tbl_df/tbl/data.frame)
 $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
 $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
 $ displ      : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
 $ year       : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
 $ cyl        : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
 $ trans      : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
 $ drv        : chr [1:234] "f" "f" "f" "f" ...
 $ cty        : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
 $ hwy        : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
 $ fl         : chr [1:234] "p" "p" "p" "p" ...
 $ class      : chr [1:234] "compact" "compact" "compact" "compact" ...
```

```
> |
```

mpg 데이터 각 컬럼이 갖는 의미

컬럼이름	설명
manufacturer	제조회사
model	자동차 모델명
displ	배기량
year	생산년도
cyl	실린더 개수
trans	변속기 종류
drv	구동방식
cty	도시연비
hwy	고속도로 연비
fl	연료 종류
class	자동차 종류

summary()

- 변수의 값을 요약한 ‘요약 통계량’을 산출
- 변수의 특성을 파악하는 데 도움이 됨

> summary(mpg)

manufacturer	model	displ	year	cyl	trans
Length:234	Length:234	Min. :1.600	Min. :1999	Min. :4.000	Length:234
Class :character	Class :character	1st Qu.:2.400	1st Qu.:1999	1st Qu.:4.000	Class :character
Mode :character	Mode :character	Median :3.300	Median :2004	Median :6.000	Mode :character
		Mean :3.472	Mean :2004	Mean :5.889	
		3rd Qu.:4.600	3rd Qu.:2008	3rd Qu.:8.000	
		Max. :7.000	Max. :2008	Max. :8.000	

drv	cty	hwy	fl	class
Length:234	Min. : 9.00	Min. :12.00	Length:234	Length:234
Class :character	1st Qu.:14.00	1st Qu.:18.00	Class :character	Class :character
Mode :character	Median :17.00	Median :24.00	Mode :character	Mode :character
	Mean :16.86	Mean :23.44		
	3rd Qu.:19.00	3rd Qu.:27.00		
	Max. :35.00	Max. :44.00		

> summary(mpg)

manufacturer	model	displ	year	cyl	trans	drv	cty
Length:234	Length:234	Min. :1.600	Min. :1999	Min. :4.000	Length:234	Length:234	Min. : 9.00
Class :character	Class :character	1st Qu.:2.400	1st Qu.:1999	1st Qu.:4.000	Class :character	Class :character	1st Qu.:14.00
Mode :character	Mode :character	Median :3.300	Median :2004	Median :6.000	Mode :character	Mode :character	Median :17.00
		Mean :3.472	Mean :2004	Mean :5.889			Mean :16.86
		3rd Qu.:4.600	3rd Qu.:2008	3rd Qu.:8.000			3rd Qu.:19.00
		Max. :7.000	Max. :2008	Max. :8.000			Max. :35.00

hwy	fl	class
Min. :12.00	Length:234	Length:234
1st Qu.:18.00	Class :character	Class :character
Median :24.00	Mode :character	Mode :character
Mean :23.44		
3rd Qu.:27.00		
Max. :44.00		

> |

View()

· ‘뷰어 창’에 원자료를 직접 보여 주는 기능

· 수많은 데이터를 직접봐서는 알 수 있는 게 없음.

· 많이 쓰이지는 않으나 데이터 전체를 참고로 확인.

· 맨 앞의 V는 대문자로 입력해야 함

· 예: View(mpg)

	manufacturer	model	displ	year	cyl	trans	drv	city	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	28	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2.0	2006	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2.0	2006	4	auto(lv)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2006	6	auto(lv)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2.0	2006	4	manual(m6)	4	20	28	p	compact
11	audi	a4 quattro	2.0	2006	4	auto(s6)	4	19	27	p	compact
12	audi	a4 quattro	2.8	1999	6	auto(l5)	4	15	25	p	compact
13	audi	a4 quattro	2.8	1999	6	manual(m5)	4	17	25	p	compact
14	audi	a4 quattro	3.1	2006	6	auto(s6)	4	17	25	p	compact
15	audi	a4 quattro	3.1	2006	6	manual(m6)	4	15	25	p	compact
16	audi	a6 quattro	2.8	1999	6	auto(l5)	4	15	24	p	midsize
17	audi	a6 quattro	3.1	2006	6	auto(s6)	4	17	25	p	midsize
18	audi	a6 quattro	4.2	2006	8	auto(s6)	4	16	23	p	midsize
19	chevrolet	c1500 suburban 2wd	5.3	2006	8	auto(l4)	r	14	20	r	suv
20	chevrolet	c1500 suburban 2wd	5.3	2006	8	auto(l4)	r	11	15	e	suv
21	chevrolet	c1500 suburban 2wd	5.3	2006	8	auto(l4)	r	14	20	r	suv
22	chevrolet	c1500 suburban 2wd	5.7	1999	8	auto(l4)	r	13	17	r	suv
23	chevrolet	c1500 suburban 2wd	6.0	2006	8	auto(l4)	r	12	17	r	suv
24	chevrolet	corvette	5.7	1999	8	manual(m6)	r	16	26	p	2seater
25	chevrolet	corvette	5.7	1999	8	auto(l4)	r	15	23	p	2seater
26	chevrolet	corvette	6.2	2006	8	manual(m6)	r	16	26	p	2seater
27	chevrolet	corvette	6.2	2006	8	auto(s6)	r	15	25	p	2seater
28	chevrolet	corvette	7.0	2006	8	manual(m6)	r	15	24	p	2seater
29	chevrolet	k1500 tahoe 4wd	5.3	2006	8	auto(l4)	4	14	19	r	suv
30	chevrolet	k1500 tahoe 4wd	5.3	2006	8	auto(l4)	4	11	14	e	suv
31	chevrolet	k1500 tahoe 4wd	5.7	1999	8	auto(l4)	4	11	15	r	suv
32	chevrolet	k1500 tahoe 4wd	6.5	1999	8	auto(l4)	4	14	17	d	suv
33	chevrolet	malibu	2.4	1999	4	auto(l4)	f	19	27	r	midsize
34	chevrolet	malibu	2.4	2006	4	auto(l4)	f	22	30	r	midsize

Showing 1 to 36 of 234 entries, 11 total columns

감사합니다.