

# 웹에서 하는 R 통계

문건웅

*2016-05-12 16:11:36*

국내외를 막론하고 의료인들이 의학논문을 쓰는 데 가장 많이 쓰였던 통계 패키지는 SPSS, SAS 등이었으나 최근에는 free software인 R이 각광받고 있다. R은 free software이면서 운영체제에 관계없이 사용할 수 있으며 뛰어난 기능과 확장성, 미려한 그래픽 등을 제공하므로 의학논문을 쓰는데 아주 유용한 도구이나 명령어를 키보드로 입력해야 하고 많은 명령어들을 기억해야 하므로 진입장벽이 상대적으로 높게 느껴지는 것이 사실이다. 이러한 진입장벽을 낮추기 위해 저자는 의학논문에서 많이 쓰이는 여러 통계방법들을 자동화하고 클릭만으로 R의 기능을 이용할 수 있는 사용자 환경을 개발하였다. 또한 자신의 컴퓨터에 R프로그램을 다운로드 받고 설치하는 번거로운 과정을 피하기 위해 web-r.org 서버를 구축하여 서비스를 시작하였다. 자신의 컴퓨터의 운영체제가 무엇이든(윈도우즈, 매킨토시, 리눅스 등) 또는 아이패드와 같은 태블릿, 심지어는 스마트폰으로도 접속하여 쉽게 통계 분석을 하고 바로 논문 출판 가능한 고화질의 그래프를 클릭만으로 그릴 수 있게 하였다. 단 인터넷 표준을 지원하는 웹 브라우저를 사용해야 한다. (낮은 버전의 explorer는 웹표준을 지원하지 않으므로 구글 크롬(chrome), 사파리(safari), 파이어폭스(firefox)등의 웹 브라우저를 권장한다.)

# 제 1 장

## 클릭만으로 표 만들기


### 제 1 절 web-r.org 접속하기

제일 먼저 웹에서 하는 R통계 홈페이지인 web-r.org에 접속하자. 저자가 만든 웹에서 하는 R통계는 R설치없이 웹에 있는 서버를 이용하여 통계분석을 하고 보다 R을 쉽게 사용하기 위한 패키지 개발 및 Shiny app 공동개발을 목표로 하고 있다. 웹에서 하는 R통계는 프로그램 설치가 필요 없다. 인터넷이 되는 어떤 컴퓨터도 가능하다. 또한 컴퓨터가 아니더라도 아이패드 등 태블릿 또는 스마트폰으로도 접속하여 통계를 돌리고 그림을 그릴 수 있다. 웹에서 하는 R통계는 웹표준을 지킨다. 하지만 우리나라에서 널리 사용되는 MS windows의 explorer 하위버전은 웹표준을 지키지 않으므로 구글chrome이나 firefox, 또는 safari등의 웹표준 인터넷 브라우저를 사용할 것을 권장한다. 다음 화면은 구글 크롬으로 web-r.org에 처음 접속한 화면이다. 처음 접속한 경우 회원 가입을 하고 로그인을 해야 서비스를 이용할 수 있다. 회원가입은 무료이다. 회원가입만 하면 준회원으로 일주일간 모든 서비스를 이용할 수 있으며 일주일 이상 사용할 경우 서버운영을 위해 일정금액을 후원금으로 입금하면 정회원이 될 것을 권장한다. 자세한 사항은 공지사항을 읽어보자.

Home
web-r.org
Keon-Woong

## 웹에서 하는 R 통계

Home
웹에서 하는 R 통계
R 패키지
책 게시판
게시판
Servers
Issues



**웹에서 클릭만으로 하는 R 통계분석**


### Web-R.org 에 오신 것을 환영합니다.

"웹에서 하는 R 통계"는 통계에 관심이 있으나 R을 어려워하는 여러 연구자들을 위한 프로젝트입니다. R 설치없이 클릭만으로 웹에 있는 서버를 이용하여 통계분석을 하고 보다 R을 쉽게 사용하기 위한 패키지 개발 및 Shiny app 공동개발을 목표로 하고 있습니다. R을 사랑하는 여러분의 많은 참여를 기다립니다.

현재는 회원가입만 하시면 준비완료로 일주일간 모든 서비스를 이용하실 수 있으나 추후 서비스가 제한될 수 있습니다.

서버운영을 위해 일정금액을 후원금으로 보내주시면 정회원으로 등급시켜드립니다. 후원금은 연의상 5만원/1년으로 정했으며 학생은(대학생까지) 1만원/1년으로 해드립니다. 공지사항의 정회원이 되는 방법을 참조하시기 바랍니다.

새로 가입하신 분들은 가입인사를 남겨주세요.



*R Statistics and Graphs for Medical Papers*

의학논문 작성을 위한 R 통계와 그래프

로그인 유지

**현재 접속중**

로그인하면 볼 수 있습니다.

**접속통계**

오늘: 53  
어제: 249  
전체: 41,486

**글작성**

1. cardiomoan <sup>9</sup>
2. ozenhealth <sup>4</sup>
3. nipponica <sup>1</sup>
4. SumiHan <sup>1</sup>
5. 김원정 <sup>1</sup>
6. Jue <sup>1</sup>
7. 피스나클 <sup>1</sup>
8. 가비아 <sup>1</sup>

2016-05-06 ~ 2016-05-12

**댓글작성**

1. cardiomoan <sup>33</sup>
2. 김보영 <sup>6</sup>
3. SumiHan <sup>1</sup>

2016-05-06 ~ 2016-05-12

**이번주 조회수**

1. cardiomoan <sup>2,025</sup>
2. ozenhealth <sup>22</sup>
3. Jue <sup>7</sup>
4. SumiHan <sup>3</sup>
5. 김원정 <sup>3</sup>
6. 팔랑팔랑 <sup>3</sup>
7. Jhs <sup>1</sup>
8. nipponica <sup>1</sup>

2016-05-06 ~ 2016-05-12

**공지사항/문고답하기** more

문고 답하기 생분본스에 대해서 문의드립니다. 1

문고 답하기 [비교통계] 결과가 나오도록 하는 방법 3

문고 답하기 2 X 2 table에서 categorical variable을 비교한 P 값 1

문고 답하기 Survival 분석에서 cumulative proportion과 95% CI 1

문고 답하기 예제 데이터 열 제목 편집방법 2

문고 답하기 csv파일 업로드 안되는 문제 2 2

문고 답하기 회귀분석 결과를 통해 HR 표기하기 2 2

문고 답하기 데이터 업로드 후 작업이 안됩니다. T.T 1

문고 답하기 정규성 검정에 대해서 여쭙습니다. 2 2

문고 답하기 데이터분석에서 설명변수와 조절변수 차이 문의 2

**웹 R통계/패키지/책 게시판** more

Rose plot with ggplot2 1

Pie plot과 Donut plot의 결함 1

책에 있는 R code 안 추송함 R 파일입니다. 2

웹에서 하는 R통계 v2.0 베타버전을 공개합니다. 1

구조방정식 없이 lavaan.org에 소개되었습니다. 2 2

Kormaps 패키지를 이용한 단계구분도 쉽게 그리기... 5

Kormaps 패키지를 이용한 단계구분도 쉽게 그리기... 5

웹에서 하는 R구조방정식 암 경시 공개 2

다중회귀도형에서 상호작용의 시각화 1

다중회귀도형에서 설명변수들의 상대적 중요성 1

**자유게시판/운영진게시판** more

윤문정

신물석

namuhanayevo

namuhanayevo

신물석

신물석

arranger

유로정

R하기

딩크형

**가입인사/방명록/등업게시판** more

cardiomoan

cardiomoan

cardiomoan

cardiomoan

cardiomoan

cardiomoan

cardiomoan

cardiomoan

cardiomoan

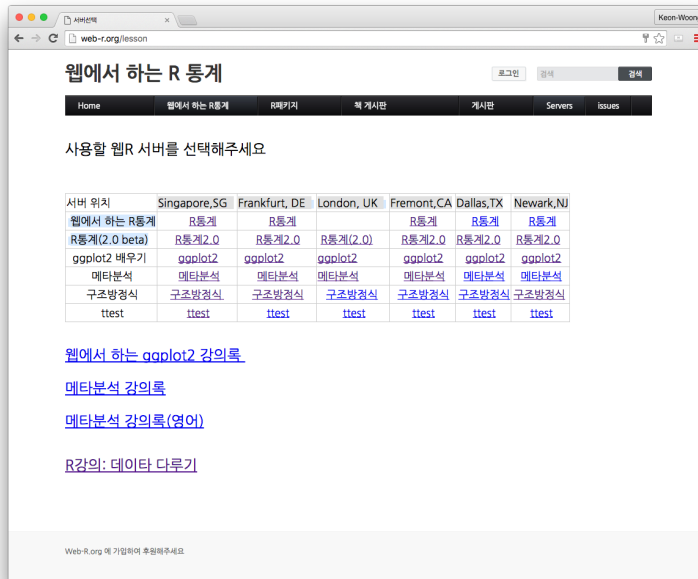
4

## 제 2 절 웹에서 하는 R통계 초기 화면

화면 메뉴중 Servers 메뉴 중 서버선택을 누른다.



다음과 같이 화면이 바뀐다. 2016년 5월 현재 웹R에서 서비스되는 프로그램(어플리케이션 또는 앱)은 모두 6개이다. 모두 6개의 서버를 통해 서비스되고 있으므로 이중 하나의 서버를 골라 R통계 2.0을 선택한다.



## 제 3 절 클릭만으로 표 만들기

잠시 기다리면 다음과 같은 화면을 볼 수 있다. 처음 접속시 준비과정이 필요하므로 몇 초 정도 기다리면 아랫부분에 표가 보이는데 표가 보이면 모든 준비가 끝난 것이다. 표가 나타날 때까지 잠시만 서버에게 시간을 주자. 처음 시작시 영어로 되어 있으며 한국어를 원할 경우 한국어를 선택 할 수 있다(1). 화면 중간 부분에서(화살표) 데이터를 선택하거나(2) 데이터 이름을 입력(3)하면 화면 아래부분에 데이터의 일부를 보여주는 표가 나타난다.

Web-R v2.0 beta

With this app, you can perform analysis **without** R in your computer. You can **analyze** data, make **tables** and **plots** and download the report with **html** or **PDF** format. You can also download the high-quality plots with desired size and resolution. You can download the results as a **powerpoint** file.

This app is a **deveLoping** beta version of web-r v2.0. It may cause an error. If you find any bug, please e-mail me.

Select Language  
English  한국어(Korean) **1**

Web-R.org

Data Table Descriptive Explore Plots Compare Categorical Multiple Group Regression survival  
Tree Analysis Diagnosis ROC curve R command Report/Plot Options Pubmed Wordcloud Propensity Score Matching  
Principal component citation information

Please **select** one of sample data or **upload** your own. You can upload data as a **csv**, **xlsx** (Microsoft Excel), **dbf** (dbase 3+), **sav** (SPSS), **dta** (STATA), **sas7bdat** (SAS) format. If you have any error, please upload as a **csv** format.

Upload data(\*.csv preferred)  
Choose File No File chosen

Select data  
**2**  acs  
 radial  
 colon  
 GBSG2  
 Iris  
 lalonde  
 decathlon2  
 CO2  
 breslow.dat  
 UScereal  
**4**  show help for data

do Preprocessing  **Reset preprocessing**

You can preprocess the data by entering the R command(s) here. Please uncheck the checkbox before enter/change the R command(s) and recheck the checkbox.

Enter the data name  
acs **3**

Select Data Preprocessing  
None

**Undo preprocessing** **Reset preprocessing**

show  
 all data  completed data  missed data

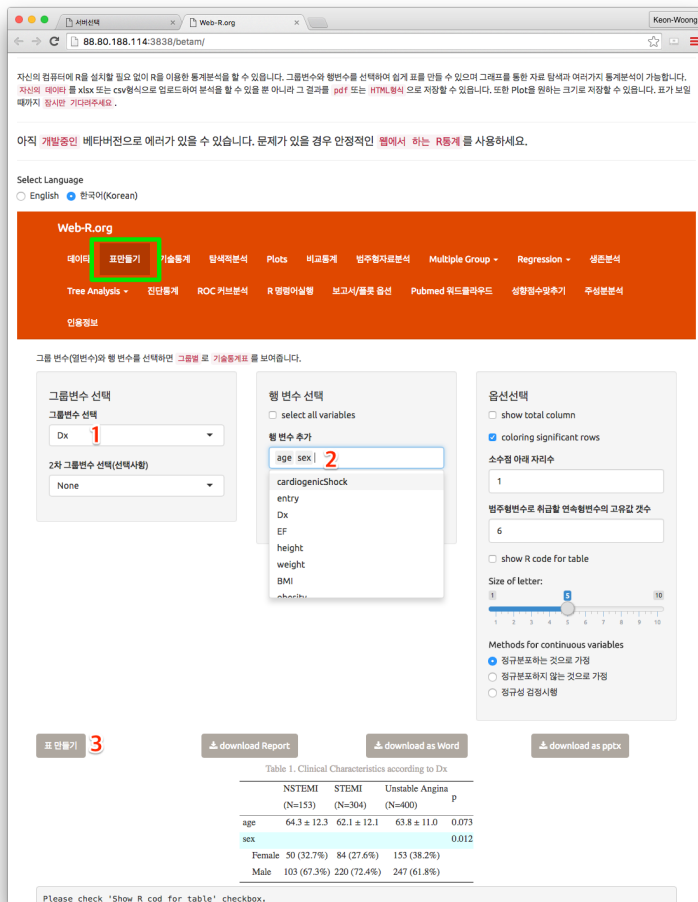
total: 857 , columns: 17 , missed: 180 **Update**

	age	sex	cardiogenicShock	entry	Dx	EF	height	weight	BMI	obesity	TC	LDLC	HDLc	TG
1	62	Male	No	Femoral	STEMI	18.00	168.00	72.00	25.51	Yes	215.00	154	35	155
2	78	Female	No	Femoral	STEMI	18.40	148.00	48.00	21.91	No				166
3	76	Female	Yes	Femoral	STEMI	20.00				No				

처음 실행되면 예제 데이터로 acs 가 선택되어 있다. acs는 저자가 만든 moonBook 패키지에 포함되어 있는 급성관동맥증후군(acute coronary syndrome) 환자의 데이터로 모두 857명의 자료이다. 데이터에 대한 도움말을 원할 경우 show help for data 체크박스(4)를 선택하여 도움말을 확인해본다.

### 3.1 표만들기

표가 나타나면 화면 윗부분의 표만들기 메뉴를 선택한다(녹색 사각형부분). 그룹변수로 Dx를 선택(1)하고 행변수추가부분에 age,sex를 선택한 후(2) 표만들기 버튼(3)을 클릭한다. 아랫부분에 Dx를 그룹변수로 하고 age, sex를 행변수로 하는 표가 나타난다. 환자의 나이(age)는 연속변수이므로 평균과 표준편차로 표시되고 성별은(sex) 범주형 자료이므로 관찰치(백분율)로 표시된다. 또한 해당하는 통계를 돌려 p값을 나타내주고 p 값이 0.05 이하인 경우 하늘색으로 표시된다.(디폴트 값으로 옵션선택 중 coloring significant rows가 선택되어 있으며 원하지 않는 경우 체크박스를 해제하면 된다.)



이때 p 값을 구하는 검정방법은 그룹변수가 이항변수일 경우(즉, 남/여, 치료군/대조군 등 두 군으로 나눌 경우)에는 행변수가 연속형변수로 정규분포하는 경우에는 t-test를 실시하고 정규분포하지 않는 경우에는 Wilcoxon Rank-Sum 검정을 실시한다. 그룹변수가 세군이상인 경우(Dx의 경우 세개의 군으로 나누어진다.) 정규분포하는 연속형변수는 anova, 정규분포하지 않는 경우에는 Kruskal-Wallis Rank-Sum 검정을 실시한다.(자세한 내용을 모르더라도 표 만드는데 지장은 없지만 논문 쓸때 무슨 검정을 했다고 써주어야 한다.) 범주형 자료는 카이제곱 검정과 Fisher's 검정을 실시한다.

### 3.2 모든 변수 선택

일일이 행 변수를 클릭하기 번거로울때는 select all variables(1)를 선택한 후 표만들기 버튼을 누른다. 그러면 다음 페이지에 있는 표가 그려진다.

Web-R.org

데이터 | **표만들기** | 기술통계 | 탐색적분석 | Plots | 비교통계 | 범주형자료분석 | Multiple Group ▾ | Regression ▾ | 생존분석

Tree Analysis ▾ | 진단통계 | ROC 커브분석 | R 명령어실행 | 보고서/플롯 옵션 | Pubmed 워드클라우드 | 성향점수맞추기 | 주성분분석

인용정보

그룹 변수(열변수)와 행 변수를 선택하면 그룹별로 기술통계표를 보여줍니다.

**그룹변수 선택**  
 그룹변수 선택: Dx **1**

2차 그룹변수 선택(선택사항): None

**행 변수 선택**  
 2  select all variables

행 변수 추가:

행 변수 삭제:

**옵션선택**  
 show total column  
 coloring significant rows

소수점 아래 자리수:

범주형변수로 취급할 연속형변수의 고유값 갯수:

그러면 데이터에 있는 모든 변수를 비교한 표가 그려지는데 별로 의미없는 자료도 포함되어 표가 너무 길어질수 있으므로 그때에는 행 변수 삭제에 빼고 싶은 변수를 선택하면 그 변수를 제외한 표가 그려진다.



Table 1. Clinical Characteristics according to Dx

	NSTEMI (N=153)	STEMI (N=304)	Unstable Angina (N=400)	P
age	64.3 ± 12.3	62.1 ± 12.1	63.8 ± 11.0	0.073
sex				0.012
Female	50 (32.7%)	84 (27.6%)	153 (38.2%)	
Male	103 (67.3%)	220 (72.4%)	247 (61.8%)	
cardiogenicShock				0.000
No	149 (97.4%)	256 (84.2%)	400 (100.0%)	
Yes	4 ( 2.6%)	48 (15.8%)	0 ( 0.0%)	
entry				0.001
Femoral	58 (37.9%)	133 (43.8%)	121 (30.2%)	
Radial	95 (62.1%)	171 (56.2%)	279 (69.8%)	
EF	55.0 ± 9.3	52.4 ± 9.5	59.2 ± 8.7	0.000
height	163.3 ± 8.2	165.1 ± 8.2	161.7 ± 9.7	0.000
weight	64.3 ± 10.2	65.7 ± 11.6	64.5 ± 11.6	0.361
BMI	24.1 ± 3.2	24.0 ± 3.3	24.6 ± 3.4	0.064
obesity				0.186
No	106 (69.3%)	209 (68.8%)	252 (63.0%)	
Yes	47 (30.7%)	95 (31.2%)	148 (37.0%)	
TC	193.7 ± 53.6	183.2 ± 43.4	183.5 ± 48.3	0.057
LDLC	126.1 ± 44.7	116.7 ± 39.5	112.9 ± 40.4	0.004
HDLC	38.9 ± 11.9	38.5 ± 11.0	37.8 ± 10.9	0.501
TG	130.1 ± 88.5	106.5 ± 72.0	137.4 ± 101.6	0.000
DM				0.209
No	96 (62.7%)	208 (68.4%)	249 (62.2%)	
Yes	57 (37.3%)	96 (31.6%)	151 (37.8%)	
HBP				0.002
No	62 (40.5%)	150 (49.3%)	144 (36.0%)	
Yes	91 (59.5%)	154 (50.7%)	256 (64.0%)	
smoking				0.000
Ex-smoker	42 (27.5%)	66 (21.7%)	96 (24.0%)	
Never	50 (32.7%)	97 (31.9%)	185 (46.2%)	
Smoker	61 (39.9%)	141 (46.4%)	119 (29.8%)	

### 3.3 전체자료의 변수값 표시하기

두 군 또는 여러군 비교와 함께 전체자료를 보고 싶을 경우 화면 오른쪽의 show total column을 체크하고(1) 왼쪽의 표만들기를 누르면(2) 전체 자료의 변수 값이 표시된다(3).

The screenshot shows the Web-R.org interface with the following settings:

- 그룹변수 선택:** Dx
- 2차 그룹변수 선택(선택사항):** None
- 행 변수 선택:** select all variables
- 옵션선택:**
  - show total column
  - coloring significant rows
  - 소수점 아래 자리수: 1
  - 범주형변수로 취급할 연속형변수의 고유값 갯수: 6
  - show R code for table
  - Size of letter: 5
  - Methods for continuous variables:
    - 정규분포하는 것으로 가정
    - 정규분포하지 않는 것으로 가정
    - 정규성 검증시행

The resulting table is:

	NSTEMI (N=153)	STEMI (N=304)	Unstable Angin (N=400)	Total (N=857)	
age	64.3 ± 12.3	62.1 ± 12.1	63.8 ± 11.0	63.3 ± 11.7	.073
sex					.012
Female	50 (32.7%)	84 (27.6%)	153 (38.2%)	287 (33.5%)	
Male	103 (67.3%)	220 (72.4%)	247 (61.8%)	570 (66.5%)	
cardiogenicShock					.000
No	149 (97.4%)	256 (84.2%)	400 (100.0%)	805 (93.9%)	
Yes	4 (2.6%)	48 (15.8%)	0 (0.0%)	52 ( 6.1%)	
entry					.001
Femoral	58 (37.9%)	133 (43.8%)	121 (30.2%)	312 (36.4%)	
Radial	95 (62.1%)	171 (56.2%)	279 (69.8%)	545 (63.6%)	
EF	55.0 ± 9.3	52.4 ± 9.5	59.2 ± 8.7	55.8 ± 9.6	.000
height	163.3 ± 8.2	165.1 ± 8.2	161.7 ± 9.7	163.2 ± 9.1	.000
weight	64.3 ± 10.2	65.7 ± 11.6	64.5 ± 11.6	64.8 ± 11.4	.361
BMI	24.1 ± 3.2	24.0 ± 3.3	24.6 ± 3.4	24.3 ± 3.3	.064
obesity					.186

### 3.4 2차 그룹변수 선택

그룹변수를 두 개 선택할 수도 있다. 예를 들어 환자를 먼저 진단명에 따라 세군으로 나누고 각각의 군을 다시 남녀 성별로 나누어 비교하고자 할 때에는 그룹변수(1)와 함께 2차그룹변수(2)를 선택하면 그림과 같은 표를 그릴 수 있다.

그룹 변수(열변수)와 행 변수를 선택하면 **그룹별** 로 기술통계표 를 보여줍니다.

그룹변수 선택  
 그룹변수 선택  
 Dx 1  
 2차 그룹변수 선택(선택사항)  
 sex 2

행 변수 선택  
 3 select all variables  
 행 변수 추가  
 행 변수 삭제

열변수 선택  
 show total column  
 coloring significant rows  
 소수점 아래 자리수  
 1  
 범주형변수로 취급할 연속형변수의 고유값 갯수  
 6  
 show R code for table  
 Size of letter:  
 1 5 10  
 Methods for continuous variables  
 정규분포하는 것으로 가정  
 정규분포하지 않는 것으로 가정  
 정규성 검증시행

표 만들기 4    download Report    download as Word    download as pptx

Descriptive Statistics Stratified by 'Dx' and 'SEX'

	STEMI				NSTEMI				Unstable Angina			
	Female (N=84)	Male (N=220)	Total (N=304)	P	Female (N=50)	Male (N=103)	Total (N=153)	P	Female (N=153)	Male (N=247)	Total (N=400)	P
age	69.1 ± 10.4	59.4 ± 11.7	62.1 ± 12.1	0.000	70.9 ± 11.4	61.1 ± 11.6	64.3 ± 12.3	0.000	67.7 ± 10.7	61.4 ± 10.6	63.8 ± 11.0	0.000
cardiogenicShock	0.535				1.000							
No	73 (86.9%)	183 (83.2%)	256 (84.2%)		49 (98.0%)	100 (97.1%)	149 (97.4%)		153 (100.0%)	247 (100.0%)	400 (100.0%)	
Yes	11 (13.1%)	37 (16.8%)	48 (15.8%)		1 (2.0%)	3 (2.9%)	4 (2.6%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	
entry	0.045				0.366				0.243			
Femoral	45 (53.6%)	88 (40.0%)	133 (43.8%)		22 (44.0%)	36 (35.0%)	58 (37.9%)		52 (34.0%)	69 (27.9%)	121 (30.2%)	
Radial	39 (46.4%)	132 (60.0%)	171 (56.2%)		28 (56.0%)	67 (65.0%)	95 (62.1%)		101 (66.0%)	178 (72.1%)	279 (69.8%)	
EF	52.3 ± 10.9	52.4 ± 8.9	52.4 ± 9.5	0.970	54.8 ± 9.1	55.1 ± 9.4	55.0 ± 9.3	0.891	59.4 ± 8.8	59.1 ± 8.7	59.2 ± 8.7	0.798
height	155.7 ± 5.4	168.7 ± 6.0	165.1 ± 8.2	0.000	154.2 ± 5.1	167.5 ± 5.7	163.3 ± 8.2	0.000	152.6 ± 6.7	167.3 ± 6.4	161.7 ± 9.7	0.000
weight	57.4 ± 9.0	68.8 ± 10.9	65.7 ± 11.6	0.000	57.2 ± 10.3	67.5 ± 8.4	64.3 ± 10.2	0.000	57.1 ± 9.1	69.0 ± 10.6	64.5 ± 11.6	0.000
BMI	23.6 ± 3.2	24.1 ± 3.4	24.0 ± 3.3	0.282	24.1 ± 4.3	24.1 ± 2.6	24.1 ± 3.2	0.967	24.5 ± 3.5	24.6 ± 3.4	24.6 ± 3.4	0.816
obesity	0.628				1.000				0.653			
No	60 (71.4%)	149 (67.7%)	209 (68.8%)		35 (70.0%)	71 (68.9%)	106 (69.3%)		99 (64.7%)	153 (61.9%)	252 (63.0%)	
Yes	24 (28.6%)	71 (32.3%)	95 (31.2%)		15 (30.0%)	32 (31.1%)	47 (30.7%)		54 (35.3%)	94 (38.1%)	148 (37.0%)	

### 3.5 연속형 변수의 통계방법 선택

연속형 변수의 통계방법은 기본적으로 정규분포하는 것으로 가정하고 통계처리를 한다. 이때에는 평균±표준편차로 표시된다.

Methods for continuous variables

- 정규분포하는 것으로 가정
- 정규분포하지 않는 것으로 가정
- 정규성 검정시행

표 만들기   download Report   download as Word   download as pptx

Descriptive Statistics Stratified by 'DX' and 'SEX'

	STEMI				NSTEMI				Unstable Angina			
	Female	Male	Total	P	Female	Male	Total	P	Female	Male	Total	P
age	69.1 ± 10.4	59.4 ± 11.7	62.1 ± 12.1	0.000	70.9 ± 11.4	61.1 ± 11.6	64.3 ± 12.3	0.000	67.7 ± 10.7	61.4 ± 10.6	63.8 ± 11.0	0.000
cardiogenicShock	0.535											
No	73 (86.9%)	183 (83.2%)	256 (84.2%)		49 (98.0%)	100 (97.1%)	149 (97.4%)		153 (100.0%)	247 (100.0%)	400 (100.0%)	
Yes	11 (13.1%)	37 (16.8%)	48 (15.8%)		1 (2.0%)	3 (2.9%)	4 (2.6%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	

정규분포하지 않는 것으로 가정하고 통계처리할 경우 중앙값(median)과 사분위수(interquartile range)로 표시된다.

Methods for continuous variables

- 정규분포하는 것으로 가정
- 정규분포하지 않는 것으로 가정
- 정규성 검정시행

표 만들기   download Report   download as Word   download as pptx

Descriptive Statistics Stratified by 'DX' and 'SEX'

	STEMI				NSTEMI				Unstable Angina			
	Female	Male	Total	P	Female	Male	Total	P	Female	Male	Total	P
age	70.0 [63.0;77.0]	59.5 [52.0;67.5]	62.0 [53.0;71.0]	0.000	74.5 [64.0;79.0]	59.0 [53.0;70.5]	65.0 [55.0;75.0]	0.000	70.0 [61.0;75.0]	61.0 [54.5;68.0]	65.0 [56.0;72.0]	0.000
cardiogenicShock	0.535											
No	73 (86.9%)	183 (83.2%)	256 (84.2%)		49 (98.0%)	100 (97.1%)	149 (97.4%)		153 (100.0%)	247 (100.0%)	400 (100.0%)	
Yes	11 (13.1%)	37 (16.8%)	48 (15.8%)		1 (2.0%)	3 (2.9%)	4 (2.6%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	
entry	0.045				0.366				0.243			

정규성 검정시행을 선택할 경우 정규성 검정을 시행하여 정규분포하는 것은 평균과 표준편차로 표시하고 정규분포하지 않는 것은 중앙값과 사분위수로 표시한다.

Methods for continuous variables

- 정규분포하는 것으로 가정
- 정규분포하지 않는 것으로 가정
- 정규성 검정시행

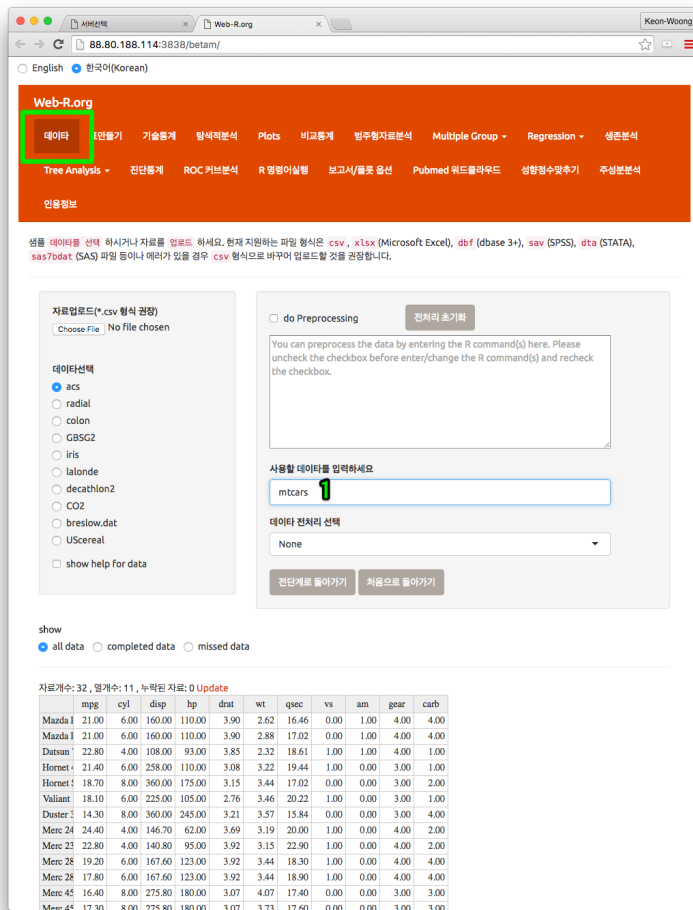
표 만들기   download Report   download as Word   download as pptx

Descriptive Statistics Stratified by 'DX' and 'SEX'

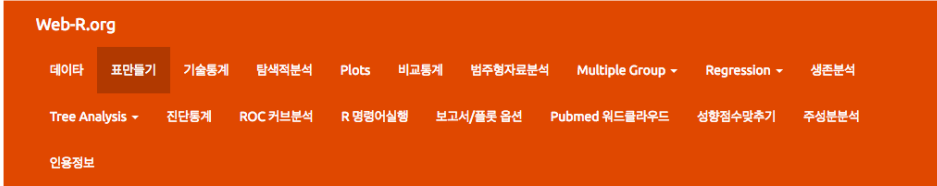
	STEMI				NSTEMI				Unstable Angina			
	Female	Male	Total	P	Female	Male	Total	P	Female	Male	Total	P
age	69.1 ± 10.4	59.4 ± 11.7	62.1 ± 12.1	0.000	70.9 ± 11.4	61.1 ± 11.6	64.3 ± 12.3	0.000	70.0 [61.0;75.0]	61.0 [54.5;68.0]	65.0 [56.0;72.0]	0.000
cardiogenicShock	1.000											
No	73 (86.9%)	183 (83.2%)	256 (84.2%)		49 (98.0%)	100 (97.1%)	149 (97.4%)		153 (100.0%)	247 (100.0%)	400 (100.0%)	
Yes	11 (13.1%)	37 (16.8%)	48 (15.8%)		1 (2.0%)	3 (2.9%)	4 (2.6%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	

### 3.6 범주형변수로 취급할 연속형변수의 고유값 개수

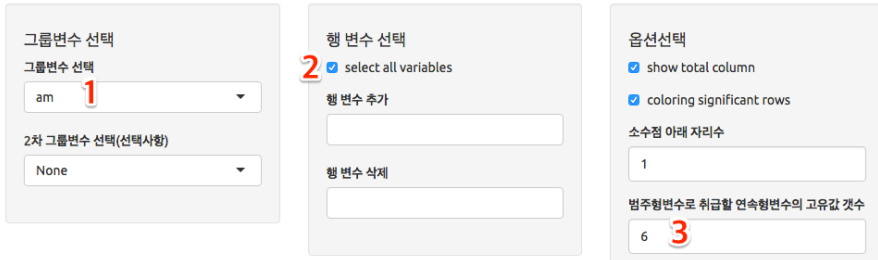
자료를 정리할 때 범주형 변수를 연속형변수처럼 입력하는 경우가 많다. 예를 들어 남, 여를 “남”, “여” 또는 “Male”, “Female” 등으로 입력하지 않고 1, 2 등 숫자를 사용하여 입력하는 경우 자료를 처리할때 남자가 몇 %, 여자가 몇 %등으로 처리해야 하지만 이것을 숫자로 처리하게 되면 평균 1.6이고 표준편차가 0.2라는 등의 결과가 나올수 있다. 예를 들어 1974년 Motor Trend라는 잡지에 나온 32개의 자동차의 데이터를 사용해보자. 데이터 (녹색사각형) 메뉴로 돌아가서 사용할 데이터를 입력하세요 부분에 mtcars(1) 를 입력한 후 아래와 같은 표가 나오는지 확인한다.



표만들기 메뉴로 돌아가 그룹변수에 am을 선택하고(1), 행변수로 select all variables를 선택해보자(2). 디폴트 값으로 범주형변수로 취급할 연속형 변수의 고유값 개수가 6으로 되어 있다(3).



그룹 변수(열변수)와 행 변수를 선택하면 그룹별로 기술통계표를 보여줍니다.



표만들기 버튼을 누른 후 만들어진 표를 보면 cyl(실린더갯수)(1), gear(전진기어단 수)(2), carb(기화기갯수)(3)가 모두 범주형변수처럼 처리되어 있다.

	0 (N=19)	1 (N=13)	Total (N=32)	P
mpg	17.1 ± 3.8	24.4 ± 6.2	20.1 ± 6.0	0.000
cyl				0.013
4	3 (15.8%)	8 (61.5%)	11 (34.4%)	
6	4 (21.1%)	3 (23.1%)	7 (21.9%)	
8	12 (63.2%)	2 (15.4%)	14 (43.8%)	
disp	290.4 ± 110.2	143.5 ± 87.2	230.7 ± 123.9	0.000
hp	175.0 [116.5;192.5]	109.0 [66.0;113.0]	123.0 [96.0;180.0]	0.044
drat	3.1 [3.1; 3.7]	4.1 [3.9; 4.2]	3.7 [3.1; 3.9]	0.000
wt	3.5 [3.4; 3.8]	2.3 [1.9; 2.8]	3.3 [2.5; 3.6]	0.000
qsec	18.2 ± 1.8	17.4 ± 1.8	17.8 ± 1.8	0.206
vs				0.556
0	12 (63.2%)	6 (46.2%)	18 (56.2%)	
1	7 (36.8%)	7 (53.8%)	14 (43.8%)	
gear				0.000
3	15 (78.9%)	0 (0.0%)	15 (46.9%)	
4	4 (21.1%)	8 (61.5%)	12 (37.5%)	
5	0 (0.0%)	5 (38.5%)	5 (15.6%)	
carb				0.284
1	3 (15.8%)	4 (30.8%)	7 (21.9%)	
2	6 (31.6%)	4 (30.8%)	10 (31.2%)	
3	3 (15.8%)	0 (0.0%)	3 (9.4%)	
4	7 (36.8%)	3 (23.1%)	10 (31.2%)	
6	0 (0.0%)	1 (7.7%)	1 (3.1%)	
8	0 (0.0%)	1 (7.7%)	1 (3.1%)	

범주형변수로 취급할 연속형 변수의 고유값 개수를 2로 바꾸고 표만들기 버튼을 누르면 표가 다시 그려지는데 cyl(실린더갯수), gear(전진기어단수), carb(기화기갯수)가 모두 연속형 변수로 처리되어 평균±표준편차로 나타나 있다.

The screenshot shows a web browser window with the URL 88.80.188.114:3838/betam/. The interface includes a control panel on the right with the following elements:

- A dropdown menu for "범주형변수로 취급할 연속형변수의 고유값 갯수" (Number of unique values for continuous variables to be treated as categorical) set to 2 (labeled 1).
- A checkbox for "show R code for table" which is unchecked.
- A "Size of letter:" slider set to 5 (labeled 2).
- Radio buttons for "Methods for continuous variables":
  - 정규분포하는 것으로 가정 (Assume normal distribution)
  - 정규분포하지 않는 것으로 가정 (Assume non-normal distribution)
  - 정규성 검증시행 (Perform normality test)

Below the control panel are four buttons: "표 만들기" (labeled 2), "download Report", "download as Word" (labeled 3), and "download as pptx" (labeled 4). The main content area displays "Table 1. Clinical Characteristics according to am" with the following data:

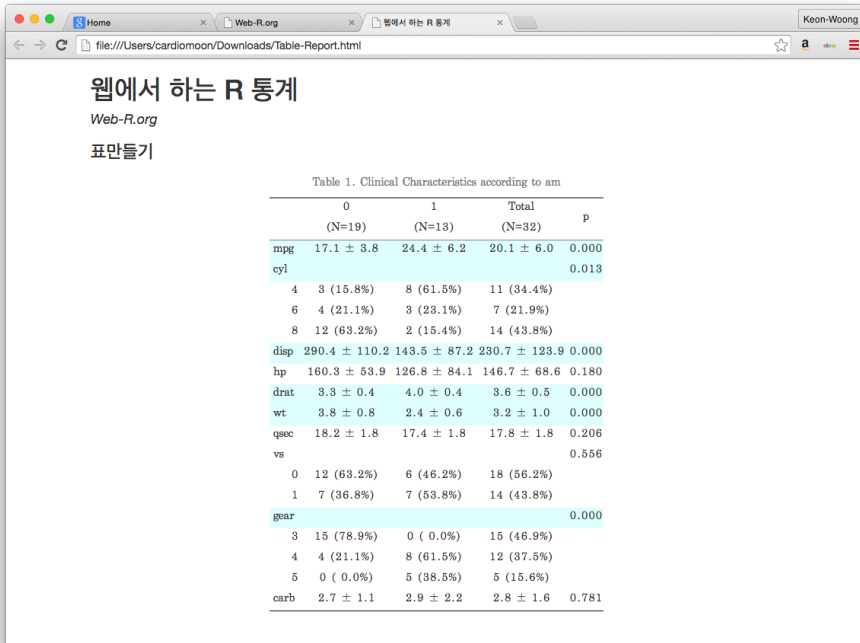
	0 (N=19)	1 (N=13)	Total (N=32)	P
mpg	17.1 ± 3.8	24.4 ± 6.2	20.1 ± 6.0	0.000
cyl	8.0 [ 6.0; 8.0]	4.0 [ 4.0; 6.0]	6.0 [ 4.0; 8.0]	0.004
disp	290.4 ± 110.2	143.5 ± 87.2	230.7 ± 123.9	0.000
hp	175.0 [116.5;192.5]	109.0 [66.0;113.0]	123.0 [96.0;180.0]	0.044
drat	3.1 [ 3.1; 3.7]	4.1 [ 3.9; 4.2]	3.7 [ 3.1; 3.9]	0.000
wt	3.5 [ 3.4; 3.8]	2.3 [ 1.9; 2.8]	3.3 [ 2.5; 3.6]	0.000
qsec	18.2 ± 1.8	17.4 ± 1.8	17.8 ± 1.8	0.206
vs				0.556
0	12 (63.2%)	6 (46.2%)	18 (56.2%)	
1	7 (36.8%)	7 (53.8%)	14 (43.8%)	
gear	3.0 [ 3.0; 3.0]	4.0 [ 4.0; 5.0]	4.0 [ 3.0; 4.0]	0.000
carb	3.0 [ 2.0; 4.0]	2.0 [ 1.0; 4.0]	2.0 [ 2.0; 4.0]	0.720

At the bottom of the interface, there is a message: "Please check 'Show R cod for table' checkbox."

만들어진 테이블을 자신의 컴퓨터로 다운로드 받을 수 있다. 마이크로소프트 워드 문서로 다운로드하려면 download as Word버튼(3)을 누르고 파워포인트 파일로 다운로드 받으려면 download as pptx버튼을 클릭하면 된다(4).

### 3.7 테이블 다운로드 받기

download Report 버튼을 클릭하면 표를 다운로드 받을 수 있다. 기본 값으로는 html 형식으로 다운로드 되는데 Table-Report.html 이름으로 다운로드 된다. 다운로드되는 위치(1)는 자신의 컴퓨터에 따라 달라질 수 있다.



웹에서 하는 R 통계  
Web-R.org  
표만들기

Table 1. Clinical Characteristics according to am

	0 (N=19)	1 (N=13)	Total (N=32)	p
mpg	17.1 ± 3.8	24.4 ± 6.2	20.1 ± 6.0	0.000
cyl				0.013
4	3 (15.8%)	8 (61.5%)	11 (34.4%)	
6	4 (21.1%)	3 (23.1%)	7 (21.9%)	
8	12 (63.2%)	2 (15.4%)	14 (43.8%)	
disp	290.4 ± 110.2	143.5 ± 87.2	230.7 ± 123.9	0.000
hp	160.3 ± 53.9	126.8 ± 84.1	146.7 ± 68.6	0.180
drat	3.3 ± 0.4	4.0 ± 0.4	3.6 ± 0.5	0.000
wt	3.8 ± 0.8	2.4 ± 0.6	3.2 ± 1.0	0.000
qsec	18.2 ± 1.8	17.4 ± 1.8	17.8 ± 1.8	0.206
vs				0.556
0	12 (63.2%)	6 (46.2%)	18 (56.2%)	
1	7 (36.8%)	7 (53.8%)	14 (43.8%)	
gear				0.000
3	15 (78.9%)	0 ( 0.0%)	15 (46.9%)	
4	4 (21.1%)	8 (61.5%)	12 (37.5%)	
5	0 ( 0.0%)	5 (38.5%)	5 (15.6%)	
carb	2.7 ± 1.1	2.9 ± 2.2	2.8 ± 1.6	0.781



## 제 4 절 보고서/Plot 옵션 바꾸기

웹R에서 다운로드 되는 모든 보고서는 html 또는 pdf 형식으로 다운로드 받을 수 있다. 그림은 기본적으로 png형식으로 다운로드 된다. 다운로드 되는 보고서와 Plot의 옵션은 보고서/Plot옵션에서 바꿀 수 있다(1).

The screenshot shows the Web-R.org website interface. The top navigation bar includes links for '데이터', '표만들기', '기술통계', '탐색적분석', 'Plots', '비교통계', '범주형자료분석', 'Multiple Group', 'Regression', and '생존분석'. The '보고서/플롯 옵션' (Report/Plot Options) link is highlighted. Below the navigation bar, there is a section for 'Download Report' and 'Powerpoint Format'. The 'Download Report' section has several options, with red numbers 1 through 4 indicating specific settings: 1 for 'Table' (checked), 2 for 'HTML' (selected), 3 for 'normal' (selected), and 4 for the 'Download' button. The 'Powerpoint Format' section has 'normal(4:3)' selected, with a red number 5. The 'Plot Options' section has 'Plot width' set to 7, 'Plot height' to 5, and 'Resolution(ppi)' to 300, with a red number 6. The 'units' dropdown is set to 'in', and the 'Plot theme' dropdown is set to 'few', with a red number 7. Below the options is a 'Sample Figure' which is a scatter plot of Sepal.Length vs Sepal.Width for three species: setosa (red), versicolour (green), and virginica (blue).

Web-R.org

통계 결과를 pdf 또는 HTML 형식의 보고서로 만듭니다.

**Download Report**

Report includes:

- Table
- Descriptive
- Plots
- Correlation
- Regression
- Logistic regression
- Survival
- Decision Tree
- ROC curve
- R command

download report as:

- PDF
- HTML

LaTeX table option:

- normal
- longtable
- sidewaysstable

**Powerpoint Format**

normal(4:3)  wide screen

**Plot Options**

통계결과 plot을 png형식으로 내려받을 때 크기를 정합니다. 그림은 하나의 zip 파일로 압축되어 다운로드됩니다.

Plot width: 7, Plot height: 5, Resolution(ppi): 300

units: in

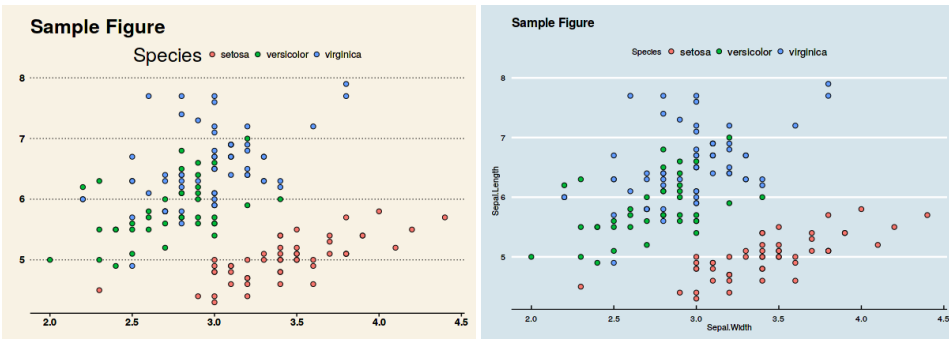
Plot theme: few

**Sample Figure**

Sepal.Length vs Sepal.Width

Species: setosa, versicolour, virginica

보고서/Plot은 하나의 메뉴에서 각각 출력할 수도 있고 이 곳에서 한꺼번에 출력할 수도 있다. 하나의 데이터를 이용하여 여러 분석을 한 이후에 보고서에 포함할 내용(1)에서 선택후 Download버튼(4)을 누르면 여러 분석 결과가 한꺼번에 다운로드 된다. 단, 여러 개의 데이터 파일을 이용하는 경우 데이터 파일이 바뀌면 보고서를 한꺼번에 출력할 수 없다. pdf형식의 보고서를 출력하기를 원할 경우 보고서 형식 선택(2)에서 pdf를 선택하면 된다. 여기서 선택한 형식으로 웹에서 하는 R통계의 모든 보고서가 다운로드 된다. LaTeX 표 선택사항(3)은 보고서 형식을 LaTeX로 선택할 경우에만 해당되는데 표만들기에서 다운로드 되는 표가 길 경우(한 페이지를 넘는 경우) longtable을 선택하면 되고 폭이 넓은 경우에는 sidewaysstable을 선택하면 옆으로 길게 인쇄된다. 오른쪽의 Powerpoint format(5)에서는 파워포인트 파일의 형식을 선택할 수 있으며 Plot옵션(6)에서는 플롯의 폭, 높이, 해상도, 단위(inch, cm, mm)를 정할 수 있다. 또한 웹R 에서 그려지는 대부분의 그래프는 ggplot2라는 R패키지를 이용하여 그려지는데 Plot의 테마(7)를 정할 경우 Plot의 모양이 달라진다. 기본 테마는 few테마로 지정되어 있으나 원하는 테마로 변경 가능하다.



위의 그림은 같은 그림을 wsj(wall street journal에서 쓰는 테마)(왼쪽)와 economist 테마(이코노미스트 잡지에서 쓰는 테마)(오른쪽)를 적용했을 때의 모양이다. 일반적으로 그림이 너무 화려하면 그림만 눈에 들어오고 그림이 주는 메시지는 기억이 나지 않는 경향이 있으므로 너무 화려한 테마는 피하는 것이 좋다. 또한 논문에 제출할 그림은 칼라로 인쇄할 경우 별도의 인쇄비를 받는 경우가 있으므로 주의하여야 한다.

## 제 5 절 인용정보

논문을 쓸 때 R을 사용한 경우 다음과 같은 형식으로 인용하여야 한다.

```
citation()
```

To cite R in publications use:

```
R Core Team (2015). R: A language and environment for
statistical computing. R Foundation for Statistical Computing,
Vienna, Austria. URL https://www.R-project.org/.
```

A BibTeX entry for LaTeX users is

```
@Manual{,
  title = {R: A Language and Environment for Statistical Computing},
  author = {{R Core Team}},
  organization = {R Foundation for Statistical Computing},
  address = {Vienna, Austria},
  year = {2015},
  url = {https://www.R-project.org/},
}
```

```
We have invested a lot of time and effort in creating R, please
cite it when using it for data analysis. See also
'citation("pkgname")' for citing R packages.
```

특정 패키지를 사용한 경우 그 패키지에 대한 인용정보는 다음과 같이 확인할 수 있다. 예를 들어 ggplot2 패키지로 플롯을 그렸을 때 인용정보는 다음과 같다.

```
citation("ggplot2")
```

To cite ggplot2 in publications, please use:

```
H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
Springer-Verlag New York, 2009.
```

A BibTeX entry for LaTeX users is

```
@Book{,
  author = {Hadley Wickham},
  title = {ggplot2: Elegant Graphics for Data Analysis},
  publisher = {Springer-Verlag New York},
  year = {2009},
```

```
isbn = {978-0-387-98140-6},  
url = {http://had.co.nz/ggplot2/book},  
}
```

저자가 만든 moonBook패키지의 인용정보는 다음과 같다.

```
citation("moonBook")
```

To cite moonBook in publications, please use:

```
Keon-Woong Moon. R statistics and graphs for medical papers.  
Hannaare Seoul, 2015.
```

A BibTeX entry for LaTeX users is

```
@Book{,  
  author = {Keon_Woong Moon},  
  title = {R statistics and graphs for medical papers},  
  publisher = {Hannarae Seoul},  
  year = {2015},  
  isbn = {978-89-5566-179-8},  
}
```

하지만 대부분의 경우 통계분석을 하는데 사용한 모든 R패키지를 다 인용하지는 않는다. 일반적으로 사용되는 통계방법이 아닌 새로운 통계방법이나 통계방법을 특별히 언급해야 할 필요가 있는 경우에만 인용하는 것이 일반적이다.