

# 의학논문 작성을 위한 R 통계

문건웅

*2015-07-15 09:04:41*



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 스마트 폰으로도 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 된다. 자세한 사항은 공지사항을 읽어보자.

Home
web-r.org
Keon-Woong

## 웹에서 하는 R 통계

Home
웹에서 하는 R 통계
R 패키지
책: R 통계와 그래프
게시판

### Web-R.org 에 오신 것을 환영합니다.

"웹에서 하는 R 통계"는 통계에는 관심이 있으나 R에 대해 거부감이 있는 여러 연구자들을 위한 프로젝트입니다. R 설치없이 웹에 있는 서버를 이용하여 통계분석을 하고 보다 R을 쉽게 사용하기 위한 패키지 개발 및 Shiny app 공동 개발을 목표로 하고 있습니다. R을 사랑하는 여러분의 많은 참여를 기다립니다.

현재는 회원가입만 하시면 조회원으로 일주일간 모든 서비스를 이용하실 수 있으나 추후 서비스가 제한될수 있습니다. 서버운영을 위해 일정금액을 후원금으로 보내주시면 정회원으로 등급시켜드립니다.(학생 1만원/1년, 일반인 3만원/1년)

공지사항의 조회원이 되는 방법을 참조하시기 바랍니다.

로그인 유지

회원가입  
 ID/PW 3:21

**현재 접속중**

로그인하면 볼 수 있습니다.

**접속 통계**

오늘: 12  
 어제: 32  
 전체: 6,312

**글작성**

1. [cardimoon](#) <sup>1</sup>  
 2015-06-25 ~ 2015-07-01

**댓글작성**

1. [cardimoon](#) <sup>1</sup>  
 2. [sasRspss](#) <sup>1</sup>  
 2015-06-25 ~ 2015-07-01

**이번호 조회수**

1. [cardimoon](#) <sup>4</sup>  
 2015-06-25 ~ 2015-07-01

**공지사항/방명록/문고답하기** more

문고 답하기 [print.ztable](#) <sup>1</sup>

문고 답하기 [Plot inset 만들기](#)

문고 답하기 [웹에서 하는 R 통계 - R코드](#) <sup>2</sup>

문고 답하기 [R서버 설치 문의드립니다.](#) <sup>4</sup>

문고 답하기 [MS word서지 저장관련](#) <sup>1</sup>

문고 답하기 [표만들기 관련](#) <sup>2</sup>

공지사항 [\[행사\] R User Conference in Korea 2015](#)

방명록 [음 새로 만들어진 메뉴요. 첫번째로 글을 남깁니다...](#) <sup>1</sup>

문고 답하기 [문고답하기 게시판을 운영합니다](#)

공지사항 [문고답하기 게시판을 운영합니다.](#)

**웹 R 통계/ 패키지/ 책 게시판** more

웹에서 하는 R 통계 게시판 [ggplot2 multitolot with error bars](#)

웹에서 하는 R 통계 게시판 [6월 15일 개신사합니다.](#) <sup>1</sup>

웹에서 하는 R 통계 게시판 [연속할변수의 생존분석 및 생존분석의 자동화](#) <sup>5</sup>

책:R통계와 그래프 [2015년 대한민국학술원 우수학술도서에 선정](#) <sup>7</sup>

웹에서 하는 R 통계 게시판 [5월 2일 추경분부서를 추가하였습니다.](#) <sup>1</sup>

웹에서 하는 R 통계 게시판 [4월 21일 역셀화일.xlsx 형식 지원](#)

웹에서 하는 R 통계 게시판 [4월 20일 Propensity-Score Matching 추가](#) <sup>1</sup>

웹에서 하는 R 통계 게시판 [4월 6일 : PubMed WordCloud 기능 추가](#) <sup>2</sup>

웹에서 하는 R 통계 게시판 [4월 5일 개신 사합니다.](#) <sup>1</sup>

웹에서 하는 R 통계 게시판 [한글레이아웃 샘플 문서 test.rmd화일 및 header.tex화일](#) <sup>1</sup>

**자유게시판/운영진게시판** more

자유게시판 [Interactive ggplot2](#) <sup>1</sup>

자유게시판 [나무 분석 질문입니다.](#) <sup>1</sup>

자유게시판 [표 작성 질문](#) <sup>1</sup>

자유게시판 [오류 질문 - PDF 변환 안되는 file입니다.](#) <sup>1</sup>

자유게시판 [오류 질문 - 표작성시](#) <sup>1</sup>

운영진게시판 [변수 선택의 어려움](#) <sup>1</sup>

운영진게시판 [shiny server pro Health check 권로](#) <sup>2</sup>

**등업게시판** more

[cardimoon](#) [등업신청합니다.](#) <sup>2</sup>

[미소전사](#) [등업신청합니다.](#) <sup>1</sup>

[simstim](#) [등업신청합니다.](#) <sup>3</sup>

[미소전사](#) [등업신청합니다.](#) <sup>1</sup>

[미소전사](#) [계정신청합니다.](#) <sup>1</sup>

[hysh](#) [등업신청합니다.](#) <sup>1</sup>

[91라](#) [Rstudio server 계정 요청드립니다.](#) <sup>1</sup>

2015-06-24 15:09 [동계호브R문왕호보](#)

2015-06-21 23:36 [vision](#)

2015-06-19 10:36 [sue](#)

2015-06-18 16:05 [ygun21](#)

2015-06-17 07:07 [BellaLuna](#)

2015-06-15 21:52 [simstim](#)

2015-06-15 20:32 [김원희](#)

## 제 2 절 웹에서 하는 R통계 초기 화면

회원가입을 하고 이메일 인증을 거쳐 로그인을 하면 로그인 전에는 보이지 않던 메뉴가 보인다. 아래 그림을 참조하자. 메뉴 중 웹에서 하는 R통계를 누른다.

The screenshot shows the website's navigation bar. The '웹에서 하는 R 통계' menu item is highlighted with a red box. Below it, a list of sub-items is visible: '웹에서 하는 R통계', '웹에서 하는 ggplot2', '웹에서 배우는 ztable', '웹에서 하는 RStudio', 'Web-based R analysis', and '웹에서 하는 R통계 게시판'.

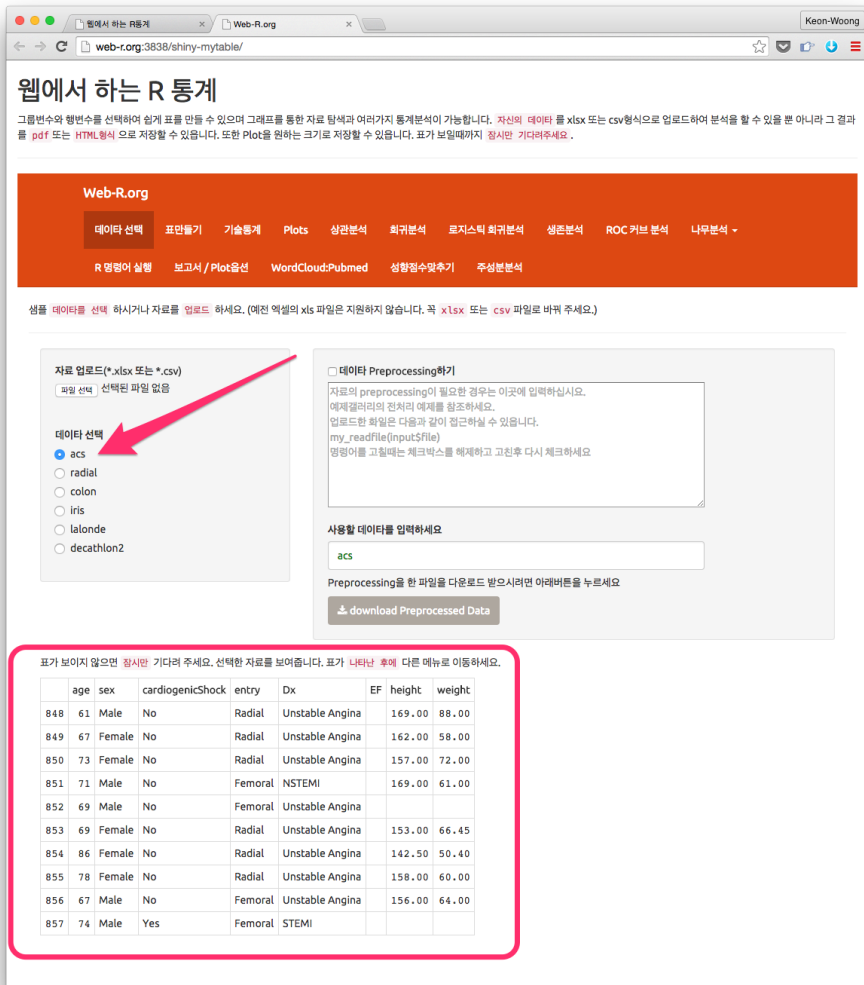
이제는 해당 메뉴를 누르면 바로 해당 메뉴로 이동할 수 있습니다. (학생 1만원/1년, 일반인 3만원/1년)  
공지사항의 정회원이 되는 방법을 참조하시기 바랍니다.

cardiomoon 로그인	현재 접속중	접속통계	글작성	댓글작성	이번호 조회수
최근 로그인: 2015-07-01	cardiomoon	오늘: 12 어제: 32 전체: 6,312	1. cardiomoon <sup>1</sup> 2015-06-25 ~ 2015-07-01	1. cardiomoon <sup>1</sup> 2. sasRspss <sup>1</sup> 2015-06-25 ~ 2015-07-01	1. cardiomoon <sup>4</sup> 2015-06-25 ~ 2015-07-01

공지사항/방명록/문고답하기	more	웹 R통계/패키지/책 게시판	more
<p>문고 답하기   print ztable 1</p> <p>문고 답하기   Plot inset 만들기</p> <p>문고 답하기   웹에서 하는 R통계 - R코드 2</p> <p>문고 답하기   R서버 설치 문외도입니다. 4</p> <p>문고 답하기   MS word문서 저장관련 1</p> <p>문고 답하기   표만들기 관련 2</p> <p>공지사항   [행사] R User Conference in Korea 2015</p> <p>방명록   을 새로 만들어진 메뉴군요. 첫번째로 글을 남깁니다... 1</p> <p>문고 답하기   문고답하기 게시판을 운영합니다</p> <p>공지사항   문고답하기 게시판을 운영합니다.</p>		<p>웹에서 하는 R통계 게시판   ggplot2 multiplot with error bars</p> <p>웹에서 하는 R통계 게시판   6월 15일 개선사항입니다. 1</p> <p>웹에서 하는 R통계 게시판   연속형변수의 생존분석 및 생존분석의 자동화 5</p> <p>책:R통계와 그래프   2015년 대한민국학술원 우수학술도서 선정 7</p> <p>웹에서 하는 R통계 게시판   5월 2일 주성분분석을 추가하였습니다. 1</p> <p>웹에서 하는 R통계 게시판   4월 21일 엑셀파일.xlsx 형식 지원</p> <p>웹에서 하는 R통계 게시판   4월 20일 Propensity-Score Matching 추가 1</p> <p>웹에서 하는 R통계 게시판   4월 6일 : PubMed WordCloud 기능 추가 2</p> <p>웹에서 하는 R통계 게시판   4월 5일 개선 사항입니다. 1</p> <p>웹에서 하는 R통계 게시판   한글레이아웃 샘플 문서 test.rmd파일 및 header.tex와...</p>	

## 제 3 절 클릭만으로 표 만들기

잠시 기다리면 다음과 같은 화면이 새로 나타난다. 화면 중간 부분에서(화살표) 데이터라 선택할 수 있으며 화면 아래부분에 데이터의 일부를 보여주는 표가 나타난다. 처음 접속시 준비과정이 필요하므로 몇 초 정도 기다리면 아랫부분에 표가 보이는데 표가 보이면 모든 준비가 끝난 것이다. 표가 나타날 때까지 잠시만 서버에게 시간을 주자.



웹에서 하는 R 통계

그룹번호와 행번호를 선택하여 쉽게 표를 만들 수 있으며 그래프를 통한 자료 탐색과 여러가지 통계분석이 가능합니다. 자신의 데이터 를 xlsx 또는 csv형식으로 업로드하여 분석을 할 수 있을 뿐 아니라 그 결과를 pdf 또는 HTML 형식으로 저장할 수 있습니다. 또한 Plot을 원하는 크기로 저장할 수 있습니다. 표가 보이기까지 잠시만 기다려주세요.

Web-R.org

데이터 선택 | 표만들기 | 기술통계 | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석

R 명령어 실행 | 보고서 / Plot 옵션 | WordCloud:Pubmed | 성장점수찾주기 | 주성분분석

샘플 데이터를 선택 하시거나 자료를 업로드 하세요. (예전 역설의 xls 파일은 지원하지 않습니다. 꼭 xlsx 또는 csv 파일로 바꿔 주세요.)

자료 업로드(\*.xlsx 또는 \*.csv)  
파일 선택 | 선택된 파일 없음

데이터 선택  
 acs  
 radial  
 colon  
 iris  
 lalonde  
 decathlon2

데이터 Preprocessing 하기  
자료의 preprocessing이 필요한 경우는 이곳에 입력하십시오.  
예제:컬러의 진치라 예제를 참조하십시오.  
업로드한 파일은 다음과 같이 접근하실 수 있습니다.  
my\_readfile(input\$file)  
명령어를 고칠때는 체크박스를 해제하고 고친후 다시 체크하십시오

사용할 데이터를 입력하세요  
acs

Preprocessing을 한 파일을 다운로드 받으시려면 아래버튼을 누르세요

download Preprocessed Data

표가 보이지 않으면 잠시만 기다려 주세요. 선택한 자료를 보여줍니다. 표가 나타난 후에 다른 메뉴로 이동하세요.

	age	sex	cardiogenicShock	entry	Dx	EF	height	weight
848	61	Male	No	Radial	Unstable Angina	169.00	88.00	
849	67	Female	No	Radial	Unstable Angina	162.00	58.00	
850	73	Female	No	Radial	Unstable Angina	157.00	72.00	
851	71	Male	No	Femoral	NSTEMI	169.00	61.00	
852	69	Male	No	Femoral	Unstable Angina			
853	69	Female	No	Radial	Unstable Angina	153.00	66.45	
854	86	Female	No	Radial	Unstable Angina	142.50	50.40	
855	78	Female	No	Radial	Unstable Angina	158.00	60.00	
856	67	Male	No	Femoral	Unstable Angina	156.00	64.00	
857	74	Male	Yes	Femoral	STEMI			

처음 실행되면 예제 데이터로 acs 가 선택되어 있다. acs는 저자가 만든 moonBook패키지에 포함되어 있는 급성관동맥증후군(acute coronary syndrome) 환자의 데이터로 모두 857명의 자료이다.



### 3.1 표만들기

표가 나타나면 화면 윗부분의 표만들기 메뉴를 선택한다(녹색 사각형부분). 그룹변수로 Dx를 선택(1)하고 행변수추가부분에 age,sex를 선택해보자(2). 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 유의한 행 색으로 구분이 선택되어 있으며 원하지 않는 경우 체크박스를 해제하면 된다.)

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%)	

이때 p 값을 구하는 검정방법은 그룹변수가 이항변수일 경우(즉, 남/여, 치료군/대

조군 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 모든 변수 선택을 클릭하면 된다. 그러면 다음 페이지에 있는 표가 그려진다.

**웹에서 하는 R 통계**

그룹변수와 행변수를 선택하여 쉽게 표를 만들 수 있으며 그래프를 통한 자료 탐색과 여러가지 통계분석이 가능합니다. 자신의 데이터를 xls는 csv형식으로 업로드하여 분석을 할 수 있을 뿐 아니라 그 결과를 pdf 또는 HTML형식으로 저장할 수 있습니다. 또한 Plot을 원하는 크기로 저장할 수 있습니다. 표가 보일때까지 잠시만 기다려주세요.

Web-R.org

데이터 선택 | **표만들기** | 기술통계 | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석

R 명령어 실행 | 보고서 / Plot 옵션 | WordCloud:Pubmed | 성향점수 맞추기 | 주성분분석

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

**그룹 변수(열변수) 선택**

그룹 변수 선택: Dx 1

2차 그룹변수선택(선택사항): None

**행 변수 선택**

모든 변수 선택 2

행 변수 추가:

행 변수 제외:

**옵션 선택**

선택후 **표만들기** 버튼을 눌러주세요

전체자료의 변수값 표시하기

유의한 행 색으로 구분

소수점 아래 자리수: 1

범주형변수로 취급할 연속형변수의 고유값 개수: 6

표에 사용한 R 코드 보여주기

그러면 데이터에 있는 모든 변수를 비교한 표가 그려지는데 별로 의미없는 자료도 포함되어 표가 너무 길어질수 있으므로 그때에는 행 변수 제외에 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 전체자료의 변수값 표시하기를 체크하고(1) 왼쪽의 표만들기를 누르면(2) 전체 자료의 변수 값이 표시된다(3).

Web-R.org

데이터 선택 **표만들기** 기술통계 Plots 상관분석 회귀분석 로지스틱 회귀분석 생존분석 ROC 커브 분석 나무분석

R 명령어 실행 보고서 / Plot 옵션 WordCloud: Pubmed 성향점수찾기 주성분분석

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

**그룹 변수(열변수) 선택**

그룹 변수 선택:  
Dx

2차 그룹변수선택(선택사항):  
None

**행 변수 선택**

모든 변수 선택

행 변수 추가:

행 변수 제외:

**옵션 선택**

선택후 **표만들기** 버튼을 눌러주세요

전체자료의 변수값 표시하기

유의한 행 색으로 구분

소수점 아래 자리수  
1

범주형변수로 취급할 연속형변수의 고유값 개수  
6

표에 사용한 R 코드 보여주기

글씨 크기 선택:  
1 5 10

연속형 변수의 통계방법 선택

정규분포하는 것으로 가정

정규분포하지 않는 것으로 가정

정규성 검증시행

**2** **표만들기**

**3**

Table 1. Clinical Characteristics according to Dx

	NSTEMI (N=153)	STEMI (N=304)	MI (N=406)	Total (N=857)	P
age	64.3 ± 12.3	62.1 ± 12.1	63.8 ± 11.0	63.3 ± 11.7	0.073
sex					0.012
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					0.000
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					0.001
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%)	

### 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 그룹변수와 함께 2차그룹변수를 선택하면 그림과 같은 표를 그릴 수 있다.

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

- 그룹 변수 (일변수) 선택:** Dx
- 2차 그룹변수 선택 (선택사항):** sex
- 행 변수 선택:** 모든 변수 선택
- 열 변수 추가:** (empty)
- 행 변수 제외:** (empty)
- 출력 선택:**
  - 전체자료의 변수값 표시하기
  - 유일한 행 색으로 구분
- 소수점 아래 자리수:** 1
- 범주형 변수로 취급할 연속형 변수의 고유값 개수:** 6
- 표에 사용한 워드 코드 보여주기:**
- 글씨 크기 선택:** 5
- 연속형 변수의 통계방법 선택:**
  - 정규분포하는 것으로 가정
  - 정규분포하지 않는 것으로 가정
  - 정규성 검사시행

The resulting table is titled "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	69.4 ± 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	0.538				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							
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

### 3.5 연속형 변수의 통계방법 선택

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

표 만들기
download Table
download as CSV

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											
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%)	

연속형 변수의 통계방법 선택  
 정규분포하는 것으로 가정  
 정규분포하지 않는 것으로 가정  
 정규성 검정시행

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

표 만들기 2
download Table
download as CSV

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	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%)	

연속형 변수의 통계방법 선택  
 정규분포하는 것으로 가정  
 정규분포하지 않는 것으로 가정  
 정규성 검정시행

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

표 만들기 2
download Table
download as CSV

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	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%)	

연속형 변수의 통계방법 선택  
 정규분포하는 것으로 가정  
 정규분포하지 않는 것으로 가정  
 정규성 검정시행

### 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. 데이터선택(녹색사각형) 메뉴로 돌아가서 사용할 데이터를 입력하세요 부분에 mtcars 를 입력한 후 아래와 같은 표가 나오는지 확인한다.

웹에서 하는 R 통계

그룹변수와 행변수를 선택하여 쉽게 표를 만들 수 있으며 그래프를 통한 자료 탐색과 여러가지 통계분석이 가능합니다. 자신의 데이터를 xlsx 또는 csv형식으로 업로드하여 분석을 할 수 있을 뿐 아니라 그 결과를 pdf 또는 HTML형식으로 저장할 수 있습니다. 또한 Plot을 원하는 크기로 저장할 수 있습니다. 표가 보일때까지 잠시만 기다려주세요.

Web-R.org

데이터 선택 | 표만달기 | 기술통계 | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석

R 명령어 실행 | 보고서 / Plot옵션 | WordCloud:Pubmed | 성향점수맞추기 | 주성분분석

샘플 데이터를 선택 하시거나 자료를 업로드 하세요. (예전 엑셀의 xls 파일은 지원하지 않습니다. 꼭 xlsx 또는 csv 파일로 바꿔 주세요.)

자료 업로드 (\*.xlsx 또는 \*.csv)  
 파일 선택 | 선택된 파일 없음

데이터 선택  
 acs  
 radial  
 colon  
 iris  
 lalonde  
 decathlon2

데이터 Preprocessing하기  
 자료의 preprocessing이 필요한 경우는 이곳에 입력하십시오.  
 예제 갤러리의 전처리 예제를 참조하세요.  
 업로드한 파일은 다음과 같이 접근하실 수 있습니다.  

```
my_readfile(input$file)
```

 명령어를 고칠때는 체크박스를 해제하고 고친후 다시 체크하세요

**사용할 데이터를 입력하세요**

Preprocessing을 한 파일을 다운로드 받으시려면 아래버튼을 누르세요

[download Preprocessed Data](#)

표가 보이지 않으면 잠시만 기다려 주세요. 선택한 자료를 보여줍니다. 표가 나타난 후에 다른 메뉴로 이동하세요.

	mpg	cyl	disp	hp	drat	wt	qsec	vs
AMC Javelin	15.20	8.00	304.00	150.00	3.15	3.44	17.30	0.00
Camaro Z28	13.30	8.00	350.00	245.00	3.73	3.84	15.41	0.00
Pontiac Firebird	19.20	8.00	400.00	175.00	3.08	3.85	17.05	0.00
Fiat X1-9	27.30	4.00	79.00	66.00	4.08	1.94	18.90	1.00
Porsche 914-2	26.00	4.00	120.30	91.00	4.43	2.14	16.70	0.00
Lotus Europa	30.40	4.00	95.10	113.00	3.77	1.51	16.90	1.00
Ford Pantera L	15.80	8.00	351.00	264.00	4.22	3.17	14.50	0.00
Ferrari Dino	19.70	6.00	145.00	175.00	3.62	2.77	15.50	0.00
Maserati Bora	15.00	8.00	301.00	335.00	3.54	3.57	14.60	0.00
Volvo 142E	21.40	4.00	121.00	109.00	4.11	2.78	18.60	1.00

표만들기 메뉴로 돌아가 그룹변수에 am을 선택하고(1), 행변수로 모든 변수 선택을 체크해보자(2). 디폴트 값으로 범주형변수로 취급할 연속형 변수의 고유값 개수가 6으로 되어 있다(3).

Web-R.org

데이터 선택 | **표만들기** | 기술통계 | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석 >

R 명령어 실행 | 보고서 / Plot을선 | WordCloud:Pubmed | 심황점수맞추기 | 주성분분석

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

**그룹 변수(열변수) 선택**  
 그룹 변수 선택: am (1)  
 2차 그룹변수선택(선택사항): None

**행 변수 선택**  
 모든 변수 선택 (2)  
 행 변수 추가:   
 행 변수 제외:

**옵션 선택**  
 선택후 표만들기 버튼을 눌러주세요  
 전체자료의 변수값 표시하기  
 유의한 행 색으로 구분  
 소수점 아래 자리수: 1  
 범주형변수로 취급할 연속형변수의 고유값 개수: 6 (3)

표를 보면 cyl(실린더갯수)(1), gear(전진기어단수)(2), carb(기화기갯수)(3)가 모두 범주형변수처럼 처리되어 있다.

www.web-r.org/shiny-mytable/

표만들기 | download Table | download as CSV

Table 1. Clinical Characteristics according to am

	0 (N=19)	1 (N=13)	p
mpg	17.1 ± 3.8	24.4 ± 6.2	0.000
cyl			0.013
	4 3 (15.8%)	8 (61.5%)	
	6 4 (21.1%)	3 (23.1%)	
	8 12 (63.2%)	2 (15.4%)	
diap	290.4 ± 110.2	143.5 ± 87.2	0.000
hp	160.3 ± 53.9	126.8 ± 84.1	0.180
drat	3.3 ± 0.4	4.0 ± 0.4	0.000
wt	3.8 ± 0.8	2.4 ± 0.6	0.000
qsec	18.2 ± 1.8	17.4 ± 1.8	0.206
vs			0.556
	0 12 (63.2%)	6 (46.2%)	
	1 7 (36.8%)	7 (53.8%)	
gear			0.000
	3 15 (78.9%)	0 ( 0.0%)	
	4 4 (21.1%)	8 (61.5%)	
	5 0 ( 0.0%)	5 (38.5%)	
carb			0.284
	1 3 (15.8%)	4 (30.8%)	
	2 6 (31.6%)	4 (30.8%)	
	3 3 (15.8%)	0 ( 0.0%)	
	4 7 (36.8%)	3 (23.1%)	
	6 0 ( 0.0%)	1 ( 7.7%)	
	8 0 ( 0.0%)	1 ( 7.7%)	

'표에 사용한 R 코드를 보여주기'를 체크하세요.



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.

범주형변수로 취급할 연속형변수의 고유값 개수

2

표에 사용한 R 코드 보여주기

글씨 크기 선택:

1 30

연속형 변수의 통계방법 선택

정규분포하는 것으로 가정

정규분포하지 않는 것으로 가정

정규성 검증시행

download Table

download as CSV

Table 1. Clinical Characteristics according to am

	0 (N=19)	1 (N=13)	P
mpg	17.1 ± 3.8	24.4 ± 6.2	0.000
cyl	6.9 ± 1.5	5.1 ± 1.6	0.002
disp	290.4 ± 110.2	143.5 ± 87.2	0.000
hp	160.3 ± 53.9	126.8 ± 84.1	0.180
drat	3.3 ± 0.4	4.0 ± 0.4	0.000
wt	3.8 ± 0.8	2.4 ± 0.6	0.000
qsec	18.2 ± 1.8	17.4 ± 1.8	0.206
vs			0.556
	0 12 (63.2%)	6 (46.2%)	
	1 7 (36.8%)	7 (53.8%)	
gear	3.2 ± 0.4	4.4 ± 0.5	0.000
carb	2.7 ± 1.1	2.9 ± 2.2	0.781

'표에 사용한 R 코드를 보여주기'를 체크하세요.



## 제 2 장

# 기술통계 보기

“기술통계” 메뉴의 화면은 모두 다섯 개의 탭으로 구성 되어 있다. 자료구조, 자료요약, 결측치 확인, 이상치 등의 탭을 하나씩 눌러보면서 자료의 기술통계를 확인할 수 있으며 dataTable 탭에서는 모든 데이터를 확인할 수 있다.

## 제 1 절 기술통계 보기

Web-R.org

데이터 선택 | 표만 보기 | **기술통계** | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석

R 명령어 실행 | 보고서 / Plot 옵션 | WordCloud: Pubmed | 성향점수 맞추기 | 주성분분석

여러 가지 기술통계를 보여줍니다. 아래 탭에서 하나를 선택해 보세요.

download Report

자료구조 | 자료요약 | 결측치 확인 | 이상치 | dataTable

```
data.frame': 857 obs. of 17 variables:
 $ age      : int  62 78 76 89 56 73 58 62 59 71 ...
 $ sex      : chr  "Male" "Female" "Female" "Female" ...
 $ cardiogenicShock: chr "No" "No" "Yes" "No" ...
 $ entry    : chr  "Femoral" "Femoral" "Femoral" "Femoral" ...
 $ Dx       : chr  "STEMI" "STEMI" "STEMI" "STEMI" ...
 $ EF       : num  18 18.4 20 21.8 21.8 22 24.7 26.6 28.5 31.1 ...
 $ height   : num  168 148 NA 165 162 153 167 160 152 168 ...
 $ weight   : num  72 48 NA 50 64 59 78 50 67 60 ...
 $ BMI      : num  25.5 21.9 NA 18.4 24.4 ...
 $ obesity  : chr  "Yes" "No" "No" "No" ...
 $ TC       : num  215 NA NA 121 195 184 161 136 239 169 ...
 $ LDLC     : int  154 NA NA 73 151 112 91 88 161 88 ...
 $ HDLC     : int  35 NA NA 20 36 38 34 33 34 54 ...
 $ TG       : int  155 166 NA 89 63 137 196 30 118 141 ...
 $ DM       : chr  "Yes" "No" "No" "No" ...
 $ HBP      : chr  "No" "Yes" "Yes" "No" ...
 $ smoking  : chr  "Smoker" "Never" "Never" "Never" ...
```

화면 윗부분의 “기술통계”를 클릭하면(녹색사각형부분) 데이터에 대한 기술통계를 볼 수 있다. 기술 통계 부분은 모두 다섯 개의 탭으로 구성되어 있는데 이 중 첫 번째인 “자료구조”를 보면(주홍색 화살표) 이 자료는 모두 17개의 변수에 대한 857개의 관측치로 구성되어 되어 있다는 것을 알 수 있으며 각각의 변수의 구조를 알 수 있다. 먼저 age는 정수형이고(int, integer) sex는 문자형(chr, character), EF는 숫자형(num, numeric)인 것을 알 수 있다. R에서는 str() 명령어로 자료구조를 볼 수 있다.

## 제 2 절 자료요약

```

자료구조  자료요약  결측치 확인  이상치  dataTable
age          sex          cardiogenicShock      entry
Min.   :28.00 Length:857      Length:857           Length:857
1st Qu.:55.00 Class :character  Class :character     Class :character
Median :64.00 Mode  :character  Mode  :character     Mode  :character
Mean   :63.31
3rd Qu.:72.00
Max.   :91.00

Dx          EF          height          weight
Length:857  Min.   :18.00  Min.   :130.0  Min.   : 30.00
Class :character  1st Qu.:50.45  1st Qu.:158.0  1st Qu.: 58.00
Mode  :character  Median :58.10  Median :165.0  Median : 65.00
Mean   :55.83  Mean   :163.2  Mean   : 64.84
3rd Qu.:62.35  3rd Qu.:170.0  3rd Qu.: 72.00
Max.   :79.00  Max.   :185.0  Max.   :112.00
NA's   :134    NA's   :93    NA's   :91

BMI          obesity      TC          LDLC
Min.   :15.62 Length:857      Min.   : 25.0  Min.   : 15.0
1st Qu.:22.13 Class :character  1st Qu.:154.0  1st Qu.: 88.0
Median :24.16 Mode  :character  Median :183.0  Median :114.0
Mean   :24.28  Mean   :185.2  Mean   :116.6
3rd Qu.:26.17  3rd Qu.:213.0  3rd Qu.:141.0
Max.   :41.42  Max.   :493.0  Max.   :366.0
NA's   :93     NA's   :23    NA's   :24

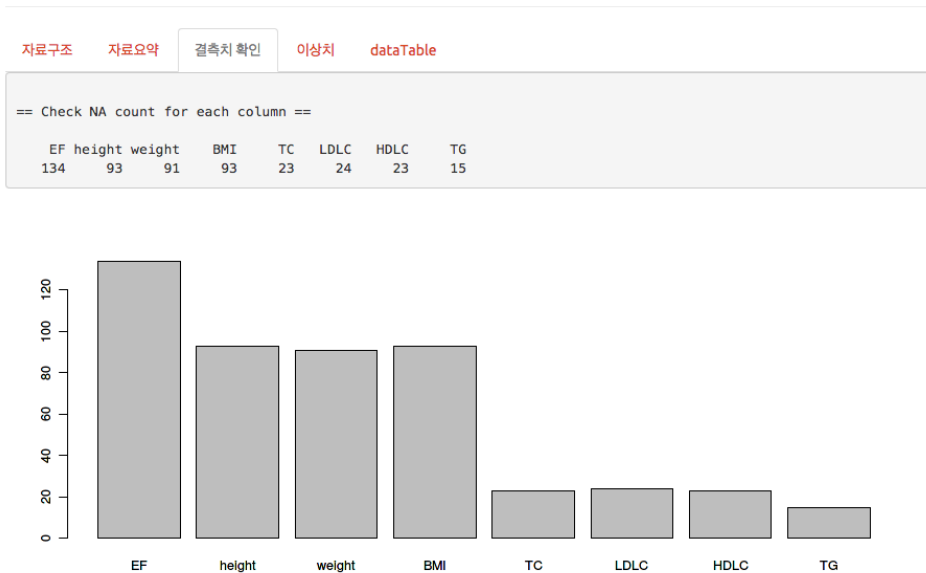
HDLC          TG          DM          HBP
Min.   : 4.00  Min.   : 11.0  Length:857      Length:857
1st Qu.:32.00  1st Qu.: 68.0  Class :character  Class :character
Median :38.00  Median :105.5  Mode  :character  Mode  :character
Mean   :38.24  Mean   :125.2  Mean   :125.2
3rd Qu.:45.00  3rd Qu.:154.0  3rd Qu.:154.0
Max.   :89.00  Max.   :877.0  Max.   :877.0
NA's   :23     NA's   :15

smoking
Length:857
Class :character
Mode  :character

```

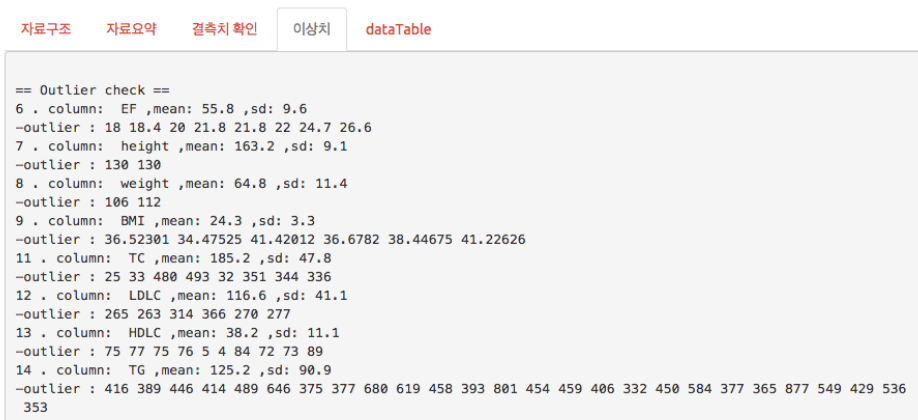
두번째 탭인 자료요약을 눌러보면 다음 그림과 같은 화면을 볼 수 있다. 정수나 숫자 등 연속형 변수인 경우에는 최소값과 사분위수 중앙값, 평균 등이 나오고 문자형 변수인 경우에는 갯수가 나온다. 각각의 변수에 결측치가 몇개인지도 확인할 수 있는데 결측치는 다음 탭인 “결측치확인”에서 보다 일목요연하게 볼 수 있다. 자료의 요약은 R에서는 summary() 명령어로 볼 수 있다.

## 제 3 절 결측치확인



“결측치 확인” 탭에서는 먼저 각 변수별로 결측치가 몇개인지 표로 요약해서 보여주고 이를 그림으로 나타내준다.

## 제 4 절 이상치



“이상치” 탭에서는 각 변수 별로 평균과 표준편차를 보여주고 이상치가 있는 경우 이상치를 보여준다. 웹R에서는 이상치를 판단할 때에 car 패키지의 outlierTest()를

사용하여 이상치를 판단하여 출력해준다.

## 제 5 절 dataTable

웹에서 하는 R 통계

그림번호와 행번호를 선택하여 쉽게 표를 만들 수 있으며 그래프를 통한 자료 탐색과 여러가지 통계분석이 가능합니다. 자신의 데이터를 csv 형식으로 업로드하여 분석을 할 수 있을 뿐 아니라 결과를 pdf 또는 HTML 형식으로 저장할 수 있습니다. 또한 Plot을 원하는 크기로 저장할 수 있습니다. 표가 보일때까지 잠시만 기다려주세요.

Web-R.org

데이터 선택 | 표만들기 | **기술통계** | Plots | 상관분석 | 회귀분석 | 로지스틱 회귀분석 | 생존분석 | ROC 커브 분석 | 나무분석

R 명령어 실행 | 보고서 / Plot 옵션 | WordCloud:Pubmed | 성향점수 맞추기 | 주성분분석

여러 가지 기술통계를 보여줍니다. 아래 탭에서 하나를 선택해 보세요.

download Report

자료구조 | 자료요약 | 결측치 확인 | 이상치 | **dataTable**

	age	sex	cardiogenicShock	entry	Dx	EP	height	weight	BMI	obesity	TC	LDL	HDL	TG	DM	HBP	smoking
1	62	Male	No	Femoral	STEMI	18.00	168.00	72.00	25.51	Yes	215.00	154	35	155	Yes	No	Smoker
2	78	Female	No	Femoral	STEMI	18.40	148.00	48.00	21.91	No				166	No	Yes	Never
3	76	Female	Yes	Femoral	STEMI	20.00				No					No	Yes	Never
4	89	Female	No	Femoral	STEMI	21.80	165.00	50.00	18.37	No	121.00	73	20	89	No	No	Never
5	56	Male	No	Radial	NSTEMI	21.80	162.00	64.00	24.39	No	195.00	151	36	63	Yes	Yes	Smoker
6	73	Female	No	Radial	Unstable Angina	22.00	153.00	59.00	25.20	Yes	184.00	112	38	137	Yes	Yes	Never
7	58	Male	No	Radial	Unstable Angina	24.70	167.00	78.00	27.97	Yes	161.00	91	34	196	Yes	Yes	Ex-smoker
8	62	Male	No	Femoral	STEMI	26.60	160.00	50.00	19.53	No	136.00	88	33	30	Yes	Yes	Ex-smoker
9	59	Female	No	Radial	Unstable Angina	28.50	152.00	67.00	29.00	Yes	239.00	161	34	118	Yes	Yes	Never
10	71	Male	No	Femoral	STEMI	31.10	168.00	60.00	21.26	No	169.00	88	54	141	Yes	No	Smoker
11	52	Male	No	Radial	Unstable Angina	31.10	175.00	60.00	19.59	No	272.00	212	32	52	Yes	No	Ex-smoker
12	52	Female	No	Radial	Unstable Angina	31.10	156.00	63.00	25.89	Yes	184.00	123	43	72	Yes	Yes	Never
13	66	Male	No	Femoral	Unstable Angina	31.50	164.00	64.00	23.80	No	239.00	150	42	156	Yes	Yes	Never

“dataTable” 탭에서는 현재 데이터 프레임의 모든 데이터를 볼 수 있다. 데이터가 많은 경우 전체 데이터를 보려면 몇초의 시간이 걸리니 잠시 기다리면 데이터를 볼 수 있다. 또한 녹색화살표가 가리키는 “download Report” 버튼을 누르면 기술통계의 보고서를 출력할 수 있다.

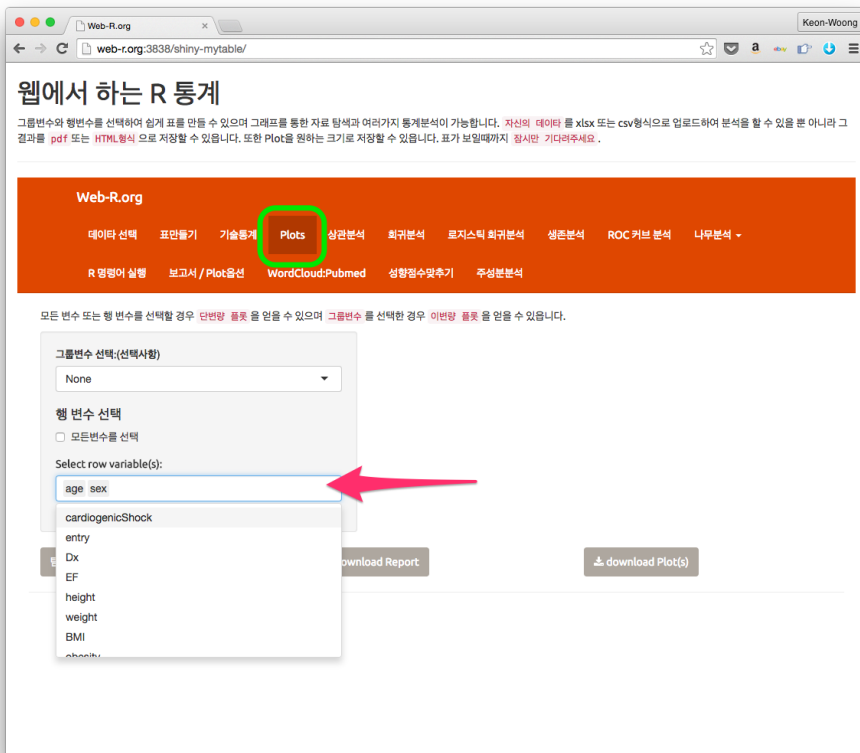




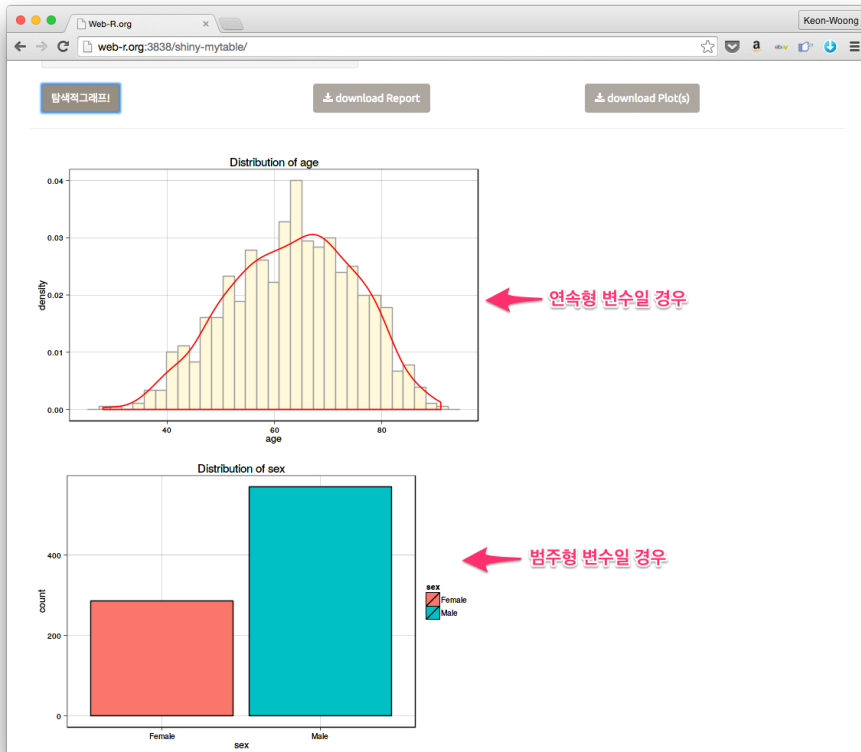
## 제 3 장

# 클릭만으로 탐색적 그래프 보기

“Plots” 메뉴에서는 클릭만으로 탐색적 그래프를 볼 수 있다. 다음 그림과 같이 행변수를 선택하고 “탐색적그래프” 버튼을 클릭해보자.



## 제 1 절 그룹변수를 선택하지 않았을 때

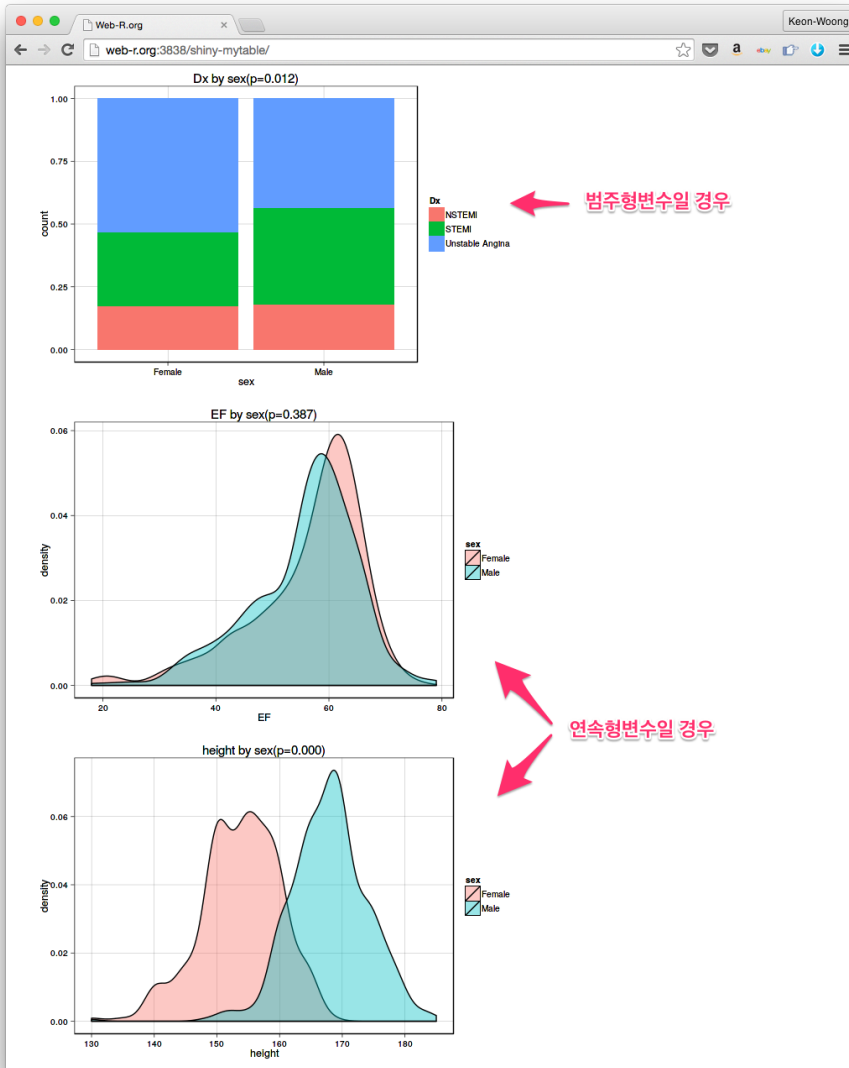


그룹변수를 선택하지 않았을 경우 탐색적그래프를 누르면 아래 그림 과 같은 그래프를 얻을 수 있다. 즉 연속형 변수인 경우 histogram과 density 곡선을 보여주며 범주형 변수인 경우 각 변수의 분포를 보여준다.

## 제 2 절 그룹변수를 선택할 경우

The screenshot shows the Web-R.org website interface. At the top, there's a navigation bar with 'Web-R.org' and a menu with options like '데이터 선택', '표본들기', '기술통계', 'Plots', '상관분석', etc. Below the navigation bar, there's a section titled '웹에서 하는 R 통계' (R Statistics on the Web) with a brief introduction. The main content area contains a form for selecting variables. The form has a dropdown menu for '그룹변수 선택:(선택사항)' (Group variable selection: optional) with 'sex' selected. Below it, there's a '행 변수 선택' (Row variable selection) section with a checked box for '모든변수를 선택' (Select all variables). There are also buttons for '탐색적그래프!' (Exploratory graph!), 'download Report', and 'download Plot(s)'. At the bottom, a density plot titled 'age by sex(p=0.000)' is displayed, showing two overlapping density curves for 'sex' (Male and Female).

그룹변수를 선택할 경우 조금 다른 그래프를 얻을 수 있다. 위의 그림과 같이 (1) 그룹변수를 선택하고 (2)행변수로는 “모든변수를 선택” 체크박스를 체크한 후 (3) 탐색적그래프 버튼을 누른다. 변수가 많을 경우 약간의 시간이 걸리니 잠시만 기다리면 선택한 모든 변수에 대한 그래프를 얻을 수 있다.

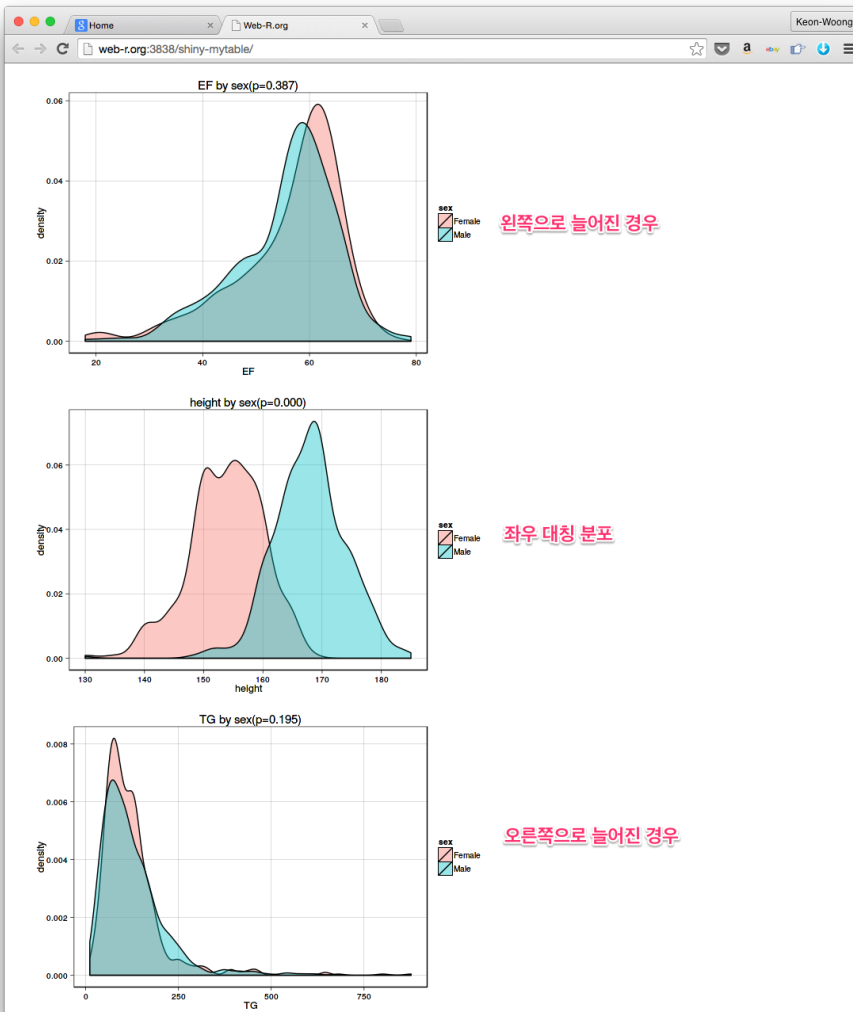


범주형 변수일 경우 분포를 알 수 있는 백분율 그래프가 나오고 연속형 변수의 경우 각각의 그룹에 대해 density plot이 그려진다. 또한 p값이 표시되는데 여기서 표시되는 p값을 얻는 통계적 방법은 표만들기에서와 같다. 즉 연속형 변수인 경우 정규분포하는 것은 모수통계인 t검정(두군일 경우)이나 ANOVA의 결과(세군이상인 경우)이고 정규분포하지 않는 경우는 비모수 통계인 Wilcoxon Rank-Sum 검정(두군일 경우) 또는 Kruskal-Wallis Rank-Sum 검정(세군 이상인 경우)의 결과이다. 이렇게 탐색적 그래프를 보는 것과 수치 요약만 보는 것은 어떤 차이가 있을까? 단순히 평균±표준편차로 되어 있는 숫자만 보는 것과 실제 분포를 보는 것은 큰 차이가 있다. 다음

페이지 그림을 보자.

Table 1. Clinical Characteristics according to sex

	Female (N=287)	Male (N=570)	P
EF	56.3 ± 10.1	55.6 ± 9.4	0.387
height	153.8 ± 6.2	167.9 ± 6.1	0.000
TG	119.9 ± 76.2	127.9 ± 97.3	0.195



단순히 숫자로 되어있는 표를 볼 때에는 자료의 분포에 대해 전혀 알 수 없었지만 탐색적그래프를 보면 세계의 분포가 전혀 다른 것을 알 수 있다. 첫번째 그래프의 EF은 좌심실구혈률(ejection fraction)으로 정상인 56-78%에 많은 자료가 분포하고 있지만 심장기능이 나쁜 환자의 경우 왼쪽으로 분포하고 있어 왼쪽으로 늘어진 모양을 하고 있다. 두번째 그래프인 키의 경우에는 비교적 좌우가 대칭인 분포를 하고 있으며 세번째 그래프인 TG의 경우 오른쪽으로 늘어진 모양을 하고 있다. 특히 의학 영역에서 피검사등의 데이터를 보면 이와같이 오른쪽으로 늘어진 모양을 하는 경우가 많은데 이와 같은 분포의 특징은 (1)음수가 없다 (2) 정상범위에 많은 자료가 있지만 정상범위를 벗어난 자료가 특히 오른쪽에 많이 분포한다는 특징을 가지고 있다. 간기능검사인 AST/ALT, 신장기능과 관련된 creatinine, 심장효소 검사 등도 이런 분포를 보이는 경우가 많다. 이런 경우 통계처리를 위해서는 정규분포에 가깝게 자료를 변환하여 주어야 하는 경우가 있다.(오른쪽으로 치우친 경우는 제곱근변환 또는 로그변환 등이 필요하다.) 자료를 볼때 요약된 숫자 또는 표만 볼 때에는 자료의 분포에 대해 전혀 알 수 없으므로 반드시 통계 처리하기 전에 탐색적 그래프를 통해 자료의 분포를 확인할 것을 권한다.

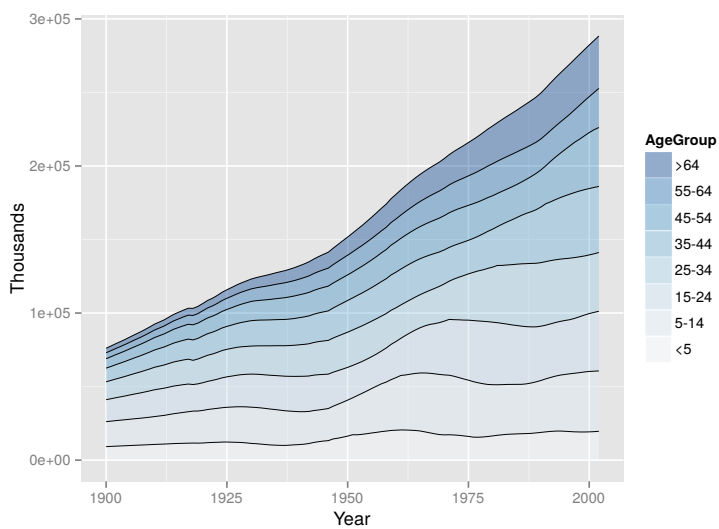
## 제 4 장

# Area plot with palette

이번 장에서 그리는 plot은 Winston Chang의 “R Graphics Cookbook”에서 인용하였음을 밝힌다.

### 제 1 절 목표

이번 장에서 그릴 그래프는 palette를 적용한 area plot이다.



## 제 2 절 웹R의 방법

### 2.1 step 1 : 예제불러오기

예제 갤러리에서 area plot with palette를 선택해보자. 그림이 나타나면 변수/ 옵션초기화 버튼(화살표)를 눌러 그림 대신 표가 나타나게 한다.

Web-R.org ggplot2 갤러리 MultiPlot

데이터를 입력 하시거나 자료를 업로드 하세요. (xlsx 또는 csv화일을 읽을 수 있습니다.)

자료 업로드 (\*.xlsx, \*.csv)  
파일 선택 선택된 파일 없음

예제 데이터를 선택 하시기 전 꼭 변수/옵션초기화를 누르세요

데이터 선택

- Salaries
- acs
- radial
- iris
- heightweight

변수/옵션 초기화

데이터 Preprocessing하기

자료의 preprocessing이 필요한 경우는 이곳에 입력하십시오.  
예제갤러리의 전처리 예제를 참조하세요.  
업로드한 파일은 다음과 같이 접근하실 수 있습니다.  
my\_readfile(input\$file)  
명령어를 고칠때는 체크박스를 해제하고 고친후 다시 체크하세요

사용할 데이터를 입력하세요

uspage

예제갤러리

- area plot with palette
- 선곡 시노 인구분포도
- 동별 서울 인구분포도
- area plot with palette
- polar plot
- heat map
- boxplot
- barplot with errorbar
- linechart with errorbar

Thousands

Year

c5

- 5-14
- 15-24
- 25-34
- 35-44
- 45-54
- 55-64
- >64

myplot (8).pdf

모두 표시



## 2.2 step 2 : 변수 할당하기

x축 변수에 Year를 할당하고(1) y축 변수에 Thousands를 할당한다(2). fill변수에 AgeGroup을 할당한다(3).

The screenshot shows the Web-R.org shiny-ggplot2 interface. At the top, there is a data table with columns Year, AgeGroup, and Thousands. Below the table are several control panels for configuring the plot. The '변수선택' (Variable Selection) panel on the left has three dropdown menus highlighted with red boxes and numbered 1, 2, and 3. The 'x축 변수 선택:' dropdown is set to 'Year' (1). The 'y축 변수 선택:' dropdown is set to 'Thousands' (2). The 'fill 변수 선택:' dropdown is set to 'AgeGroup' (3). Other panels include 'geometry options', 'color 변수 선택:', 'alpha 변수 선택:', 'linetype 변수 선택:', 'size 변수 선택:', and '원형그래프프리기' (Faceted Plot) options.

	Year	AgeGroup	Thousands
815	2001	55-64	25315
816	2001	>64	35352
817	2002	<5	19609
818	2002	5-14	41037
819	2002	15-24	40590
820	2002	25-34	39928
821	2002	35-44	44917
822	2002	45-54	40084
823	2002	55-64	26602
824	2002	>64	35602

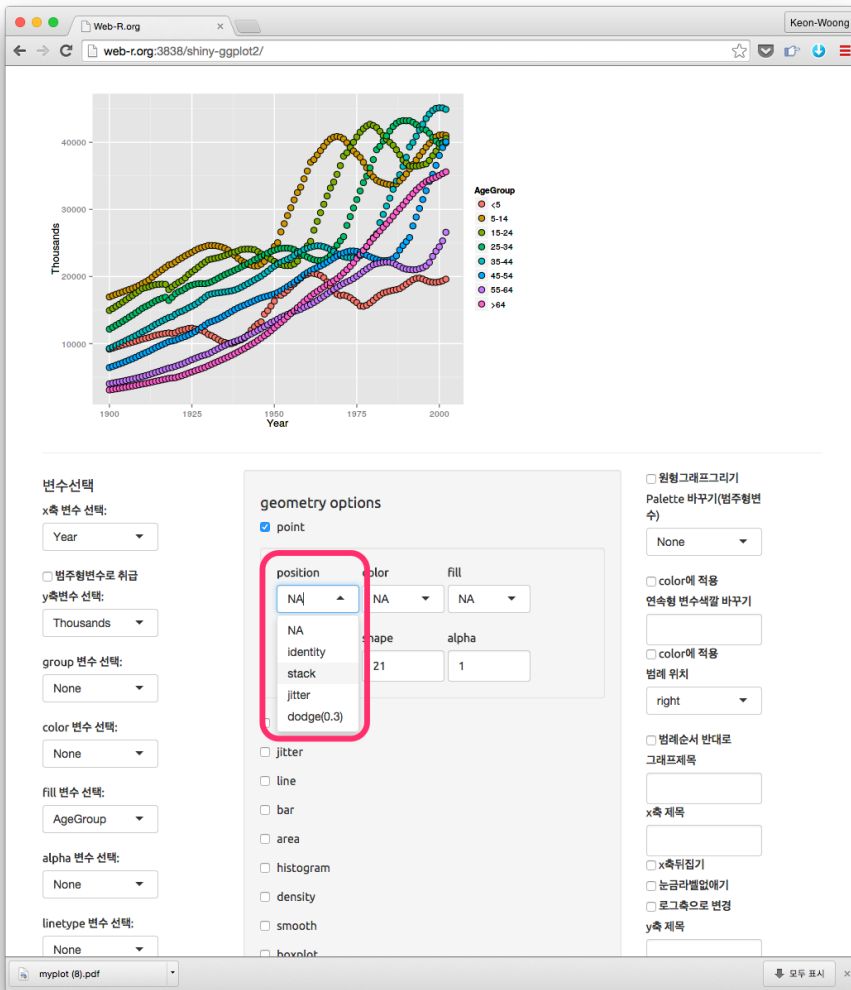
## 2.3 step 3 : 점그래프 그리기

geometry options 중에서 point 체크박스를 클릭한다(1). 그러면 화면과 같은 그래프가 나타나는데 point의 shape을 21로 바꿔준다(2).

The screenshot shows a web browser window with the URL `web-r.org:3838/shiny-ggplot2/`. The main plot area displays a line graph with 'Year' on the x-axis (1900 to 2000) and 'Thousands' on the y-axis (0 to 40,000). Multiple lines represent different age groups, with a legend on the right titled 'AgeGroup' listing categories: <5, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, and >64. Below the plot is a control panel. In the 'geometry options' section, the 'point' checkbox is checked and circled in red with a '1' next to it. The 'shape' dropdown menu is also circled in red with a '2' next to it, showing the value '16'. Other options like 'text', 'jitter', 'line', 'bar', 'area', 'histogram', 'density', 'smooth', and 'boxplot' are listed but unchecked. The bottom of the browser shows a tab for 'myplot (8).pdf' and a '모두 표시' button.

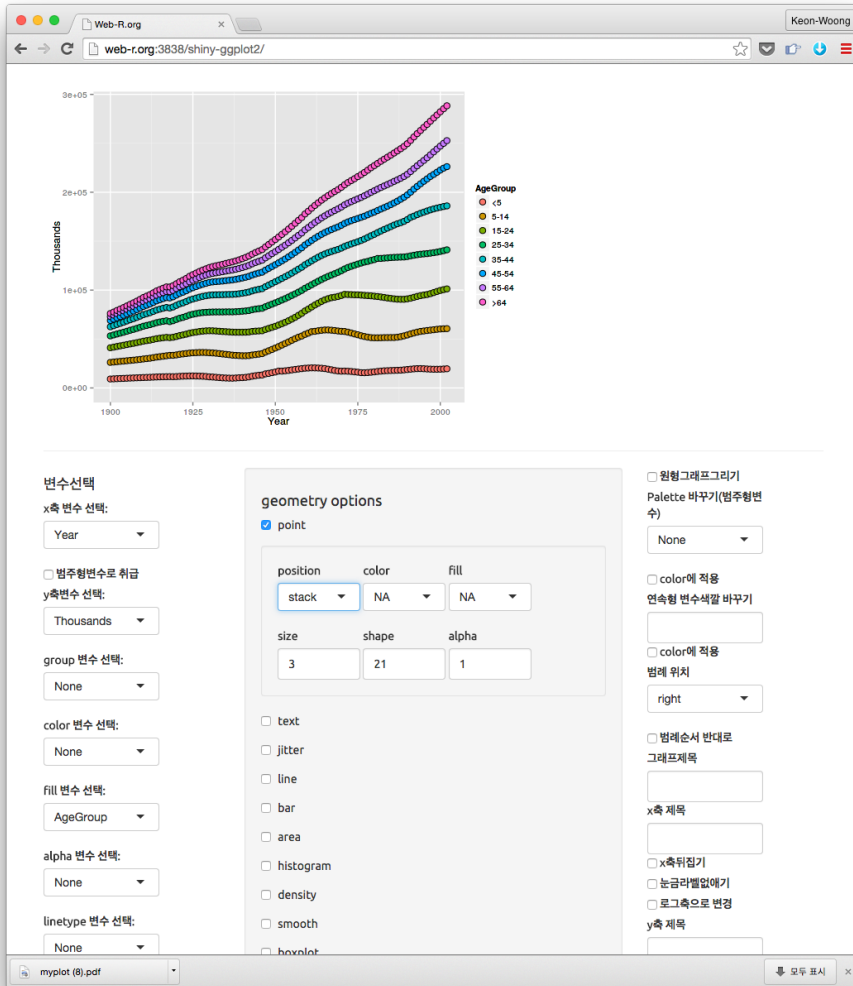
## 2.4 step 4 : 점모양 바꾸기

point의 모양이 21(속이 채워진 동그라미)로 바뀌어 연령대별로 구분은 가지만 위치가 모두 겹쳐있다. 우리가 그리고자 하는 그래프는 연령별로 구분된 그래프가 아닌 연령대별로 합쳐서 전체 인구의 변동을 그래프를 보고자 하므로 위치(position)를 stack으로 바꾼다.



## 2.5 step 5 : 위치 바꾸기

point이 stack으로 바뀌면서 우리가 의도한 대로 연도변화에 따른 전체 인구 구성비 및 전체인구의 변동을 한눈에 알 수 있게 되었다. 하지만 우리가 그리고자 하는 것은 점그래프가 아닌 area plot이다.



## 2.6 step 6 : area plot 보기

점그래프가 아닌 area plot을 그리기 위해 point 체크박스를 해제하고(1) area 체크박스를 선택한다(2). 화면과 같은 그래프가 나타나는데 배색이 조금 마음에 들지 않는다.

The screenshot shows a web browser window displaying a ggplot2 area plot. The plot shows population trends from 1900 to 2000, categorized by age groups. The y-axis is labeled 'Thousands' and ranges from 0e+00 to 3e+05. The x-axis is labeled 'Year' and ranges from 1900 to 2000. The legend for 'AgeGroup' includes categories: <5, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, and >64. Below the plot is an interactive settings panel for 'geomtry options'. In this panel, the 'point' checkbox is unchecked and circled in red with a '1' next to it, and the 'area' checkbox is checked and circled in red with a '2' next to it. Other settings include 'stat' (NA), 'color' (NA), 'fill' (NA), 'size' (0.2), and 'alpha' (1). The panel also includes options for 'histogram', 'density', 'smooth', and 'hvorlot'. On the right side of the panel, there are options for '원형그래프그리기', 'Palette 바꾸기', 'color에 적용', '연속형 변수색깔 바꾸기', '범례 위치', '범례순서 반대로', '그래프제목', 'x축 제목', 'x축뒤집기', '눈금리벨업예기', '로그축으로 변경', and 'y축 제목'.

## 2.7 step 7 : 색상 palette 선택

화면 오른쪽에 있는 Palette바꾸기에서 Blues를 선택한다.



## 2.8 step 8 : line을 추가하기

각 영역을 구분하기 위한 line을 추가하고 싶다. 먼저 line 체크박스를 체크하면(1) 화살표처럼 여러개의 line이 영켜서 나오게 된다(2). 점그래프에서와 마찬가지로 line의 position을 stack으로 바꾸어 준다(3).



## 2.9 step 9 : 투명도 조절

선이 너무 굵으므로 선의 size을 0.4로 하고 영역의 투명도(alpha)를 0.4로 한다. 또한 그래프에서의 영역의 순서와 범례의 순서가 서로 반대이므로 화면 오른쪽의 범례순서 반대로를 체크한다. 이것으로 그림이 완성된다.





## 제 3 절 R의 표준방법

이 그림을 그린 R 명령어는 다음과 같다. 명령어를 익히는 것이 더 간단할 수도 있지만 중간과정을 볼 수 없으므로 우리가 원하는 그래프를 얻기 위하여 여러 번 시행착오를 겪을 수도 있다. 반면에 웹R을 이용하여 그래프를 그리는 경우 옵션을 바꿀 때마다 그때그때 그래프의 변화를 눈으로 볼 수 있다는 장점이 있다.

```
require(ggplot2)
require(gcookbook) # uspopage데이터의 사용을 위해
ggplot(uspopage, aes(x=Year, y=Thousands, fill=AgeGroup)) +
  geom_area(size=0.2, alpha=0.4) +
  geom_line(position='stack', size=0.2) +
  scale_fill_brewer(palette="Blues") +
  guides(fill=guide_legend(reverse=TRUE))
```

