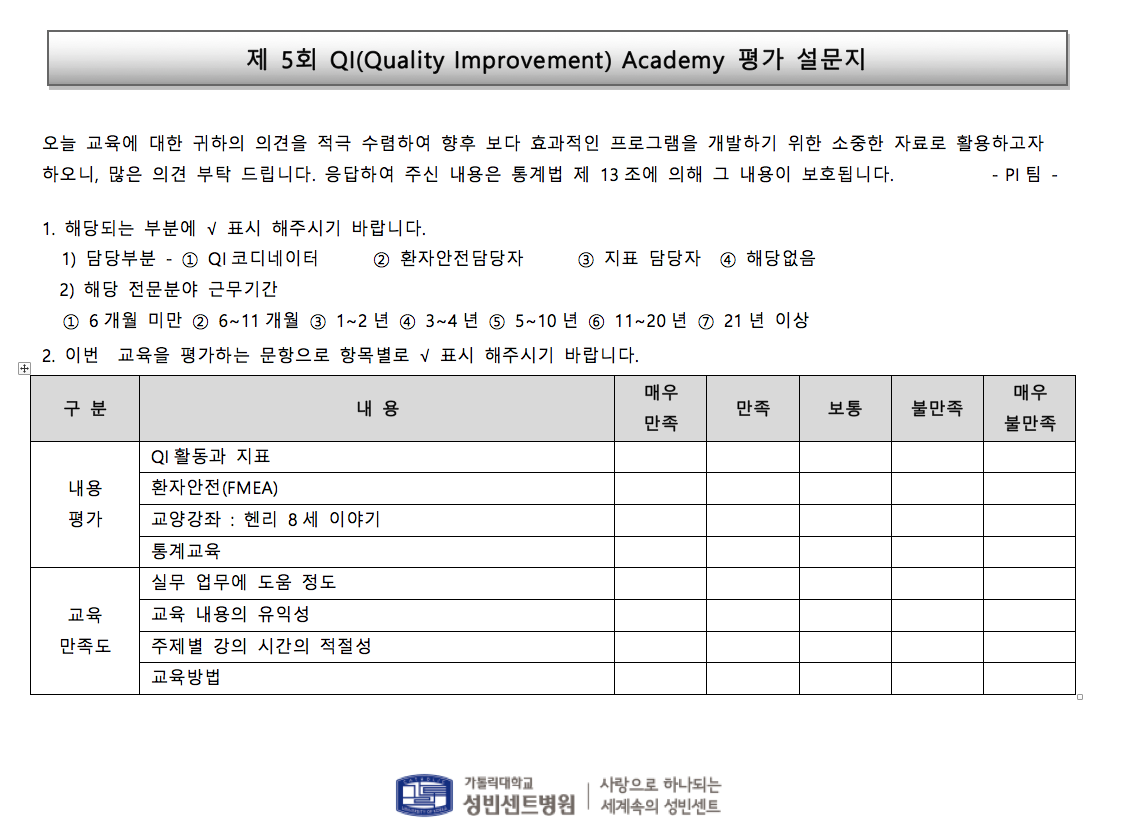
웹R을 이용한 설문조사 데이터 분석

문건웅

2017년 4월 20일

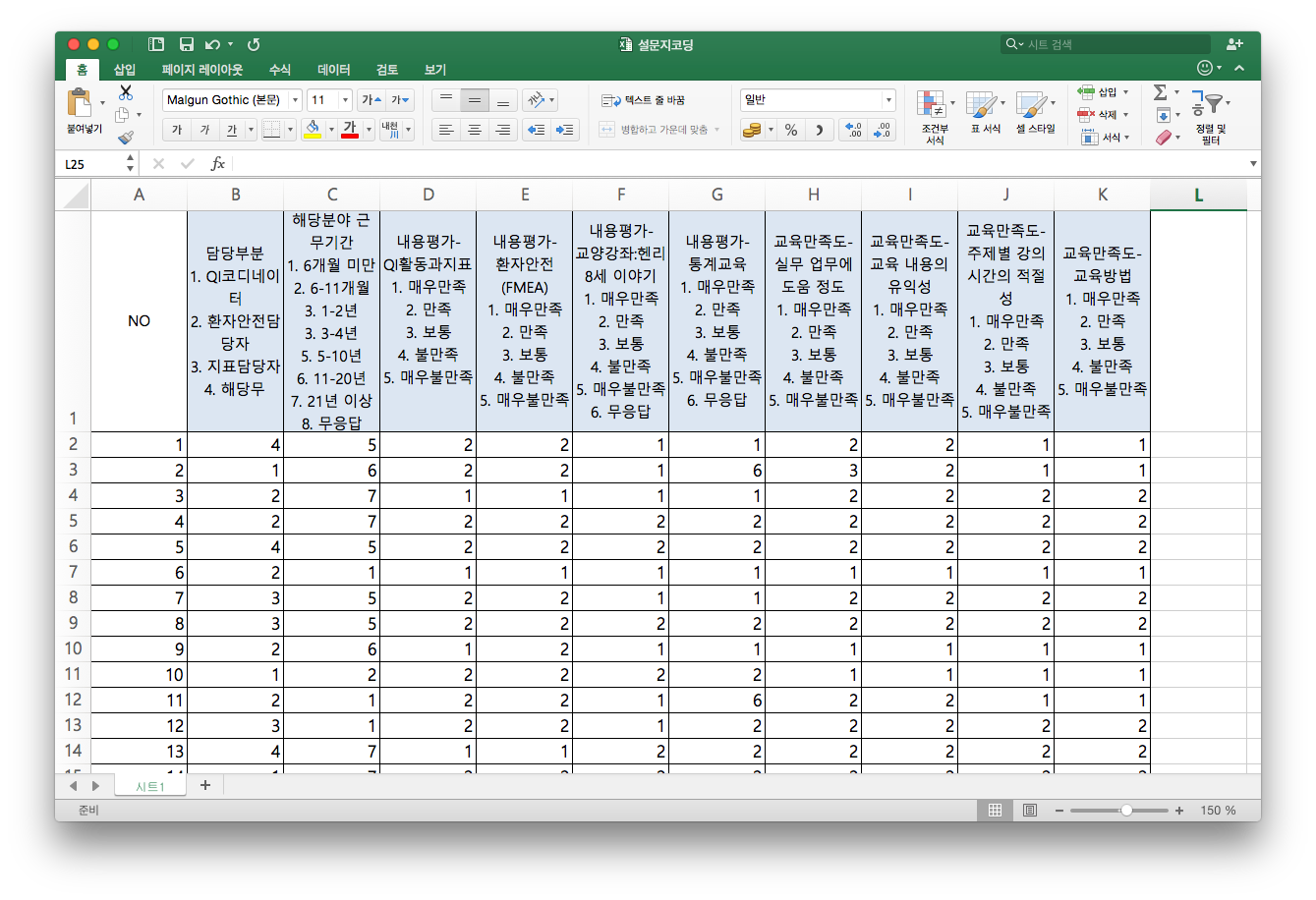
## 예제

가톨릭의과대학 성빈센트병원에서는 제5회 Quality Improvement(QI) Academy를 한 후 다음과 같은 설문 조사를 하였다.



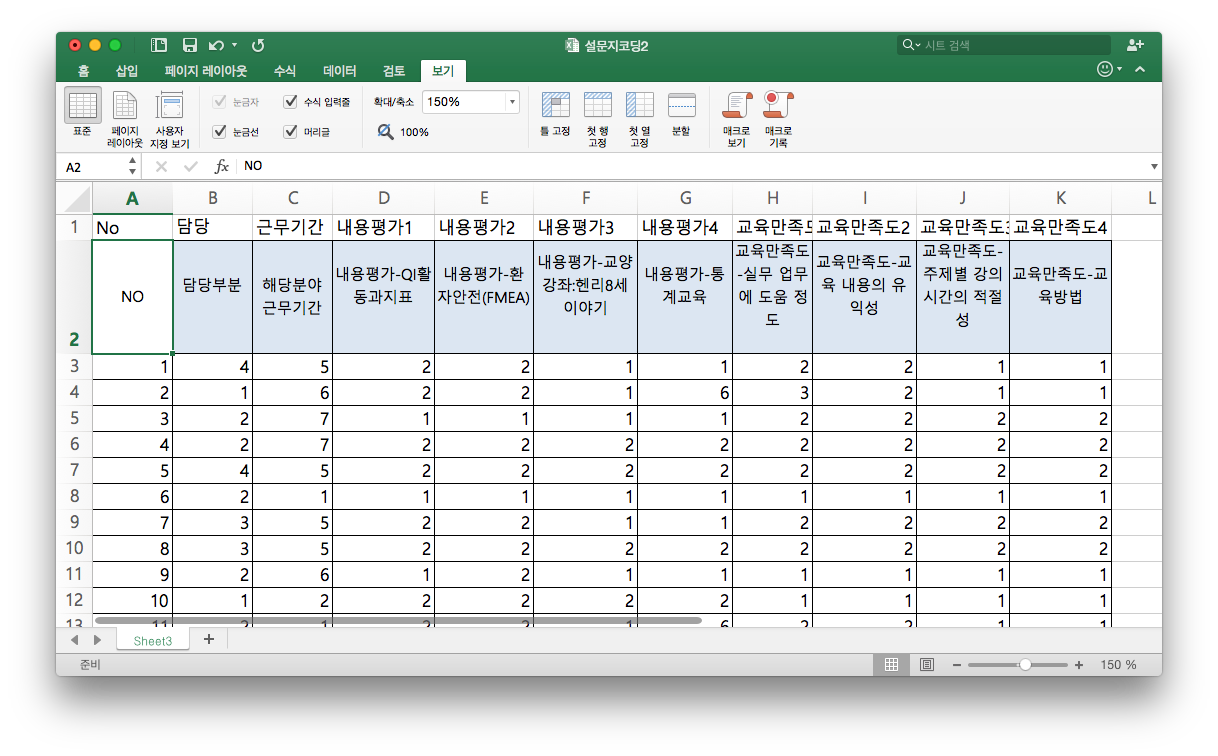
## 자료의 입력(1)

모두 55명에게 설문조사가 이루어졌으며 이후 자료의 통계 처리를 위하여 MS Excel 을 이용하여 다음과 같이 정리를 하였다.



이 자료파일의 이름은 설문지코딩.xlsx이다. 이 자료는 비교적 정리가 잘 되어 있으나 이 자료를 그대로 쓰기에는 문제가 있다. 첫번째 행의 열이름이 너무 길어 이후에 이 열 이름을 사용하여 통계 처리 하기가 어렵다. 또한 입력단계에서는 각 항목 라벨이 의미있으나 열이름에 라벨이 포함될 필요는 없다.

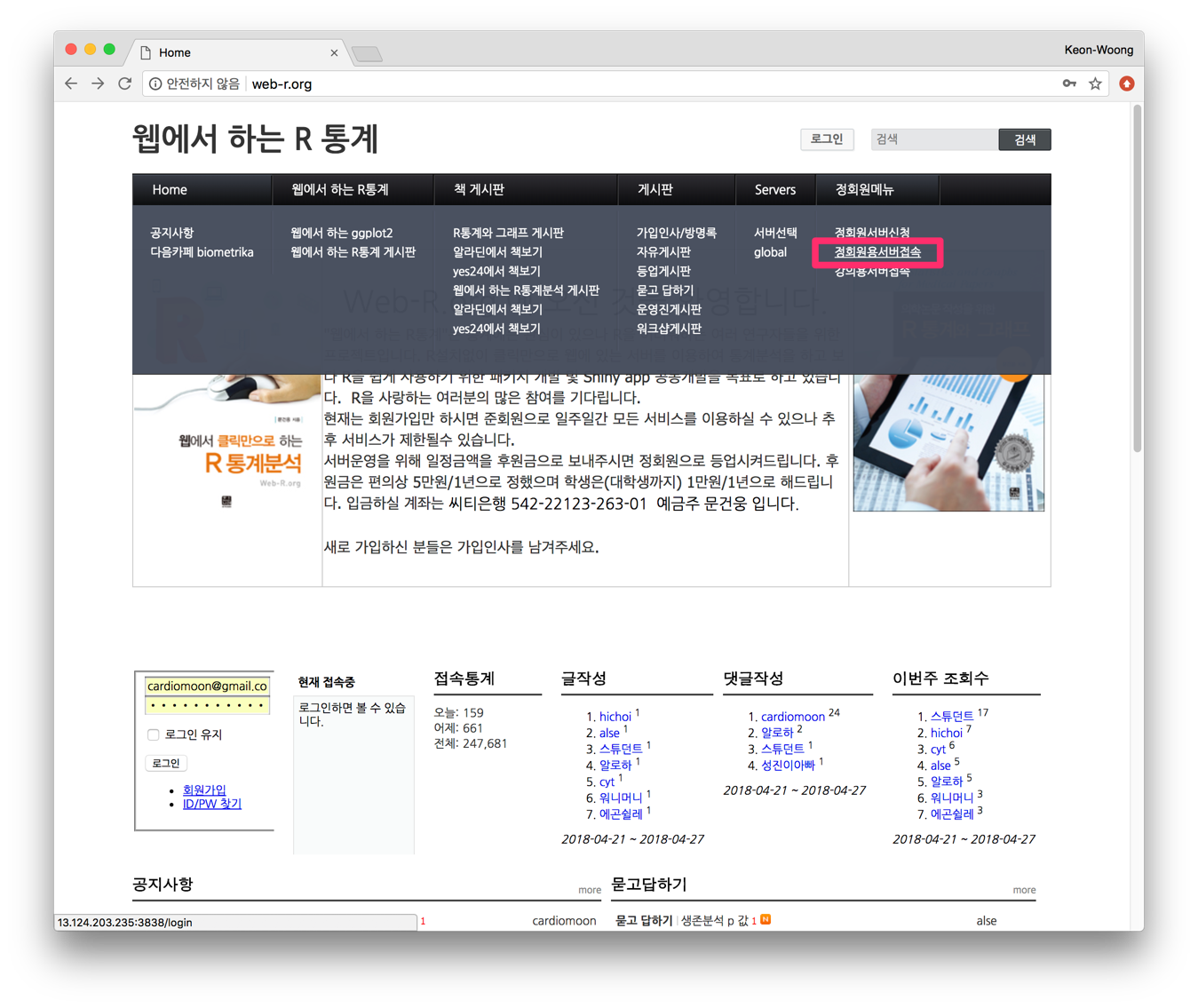
## 자료의 정리(2)



가장 권하고 싶은 방법은 이와 같이 정리를 하는 것이다.열 이름은 짧을수록 좋으며 빈칸이 있으면 안된다. 또한 숫자로 시작하지 않는 것이 좋다. 단어 사이를 띄어쓰기를 하고 싶은 경우 언더바를 사용하는 것이 좋다. (예: 언더\_바) 먼저 열이름을 최대한 간단하게 이름 붙이고 두번째 행에 자세한 내용을 적는다. 두번째 행의 내용은 이후 웹R에서 열이름 라벨로 사용된다. 이 자료파일의 이름은 설문지코딩2.xlsx이다.

## 웹R의 실행

Chrome 등의 웹브라우져에서 주소창에 web-r.org를 입력한다. Microsoft explorer는 웹표준인 HTML5를 지원하지 않으므로 chrome이나 safari, firefox, opera등의 브라우져를 사용할 것을 권장한다.

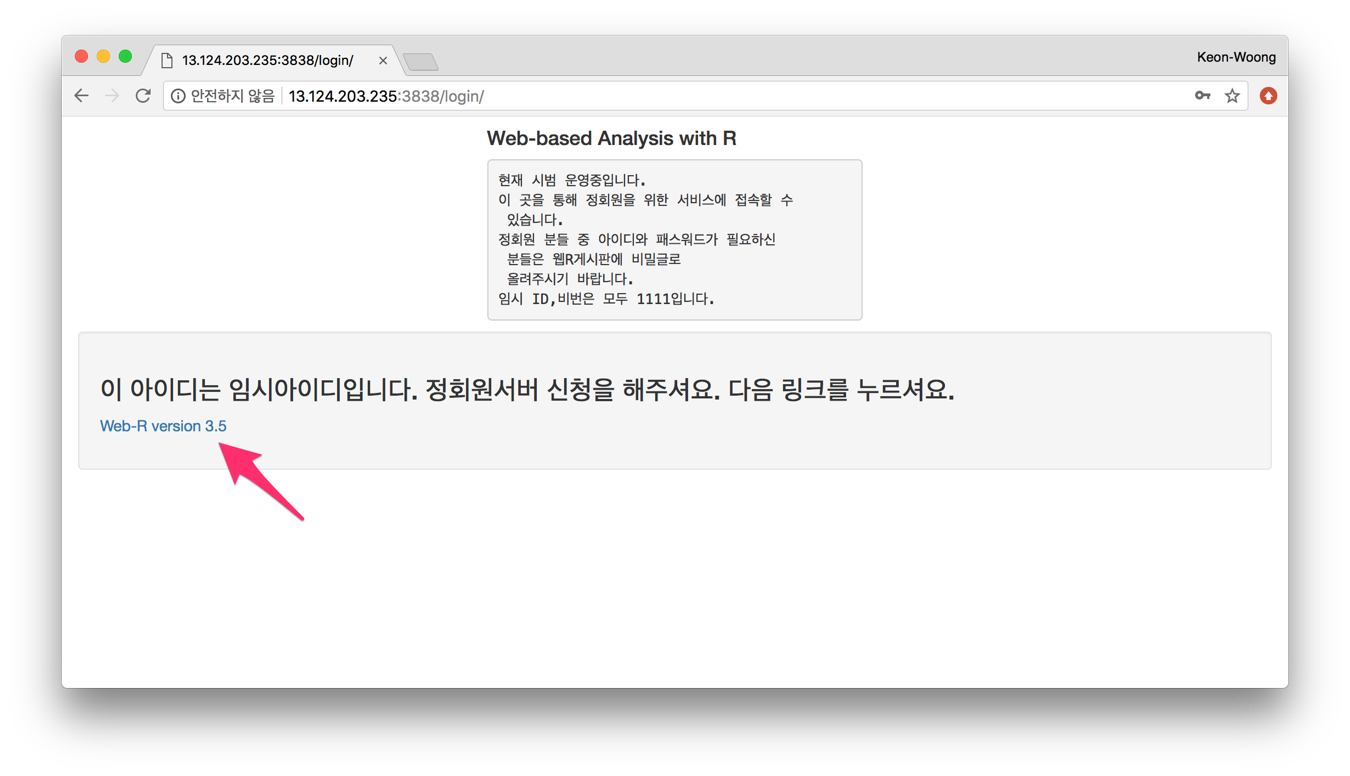


가장 위의 메뉴 중 정회원메뉴 -> 정회원용서버접속을 클릭한다.

UserName과 Password모두 1111을 입력하고 Enter버튼을 누른다.

## 

다음 화면에서 Web-R version 3.5를 누른다.



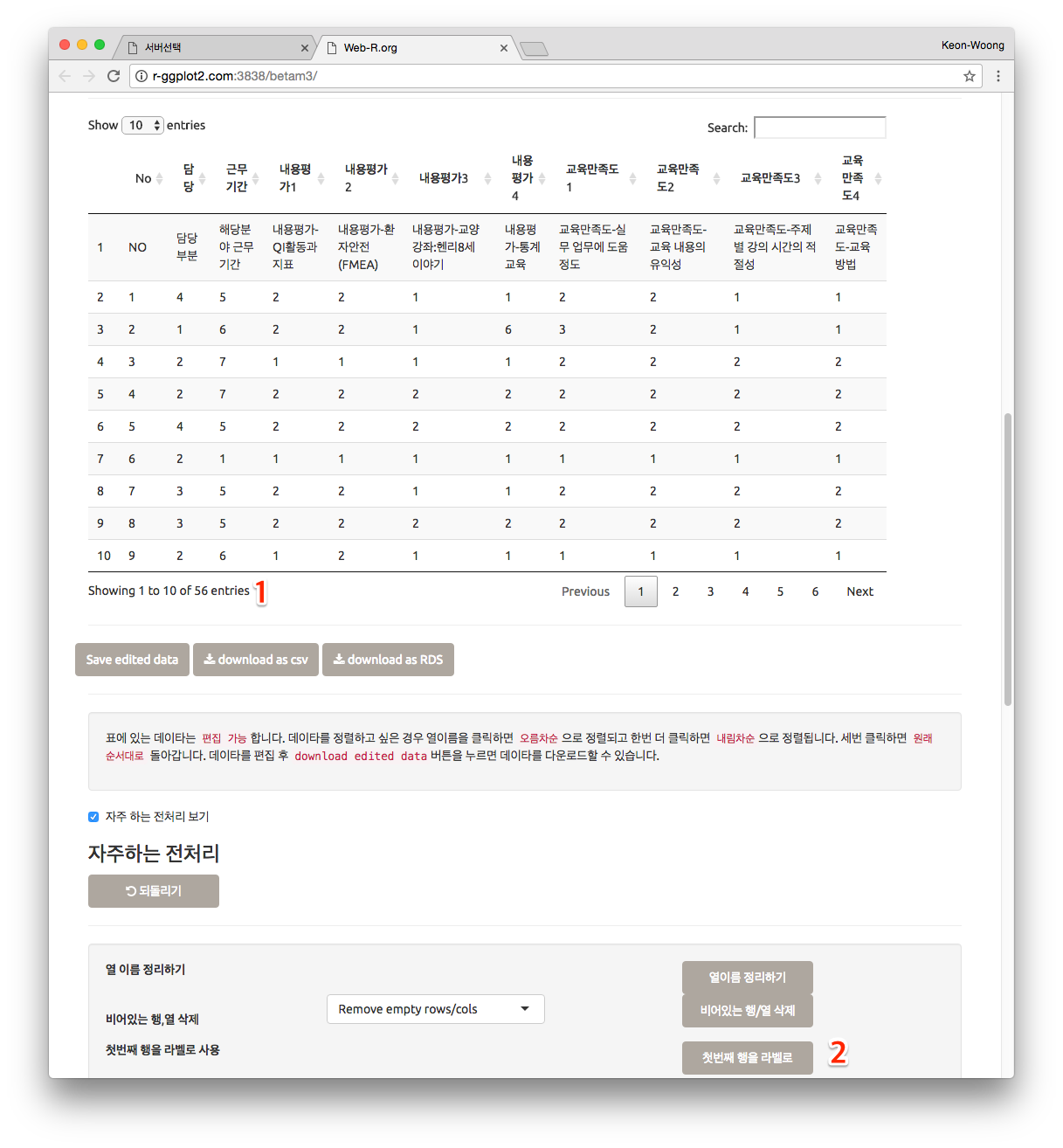
## 자료파일 업로드

1. 언어선택에서 한국어를 선택하고
2. Browse 버튼을 누른 후 아까 준비한 설문지코딩2.xlsx파일을 업로드한다.
3. 사용할 데이타의 이름이 uploaded로 바꾸는 것을 확인한다.



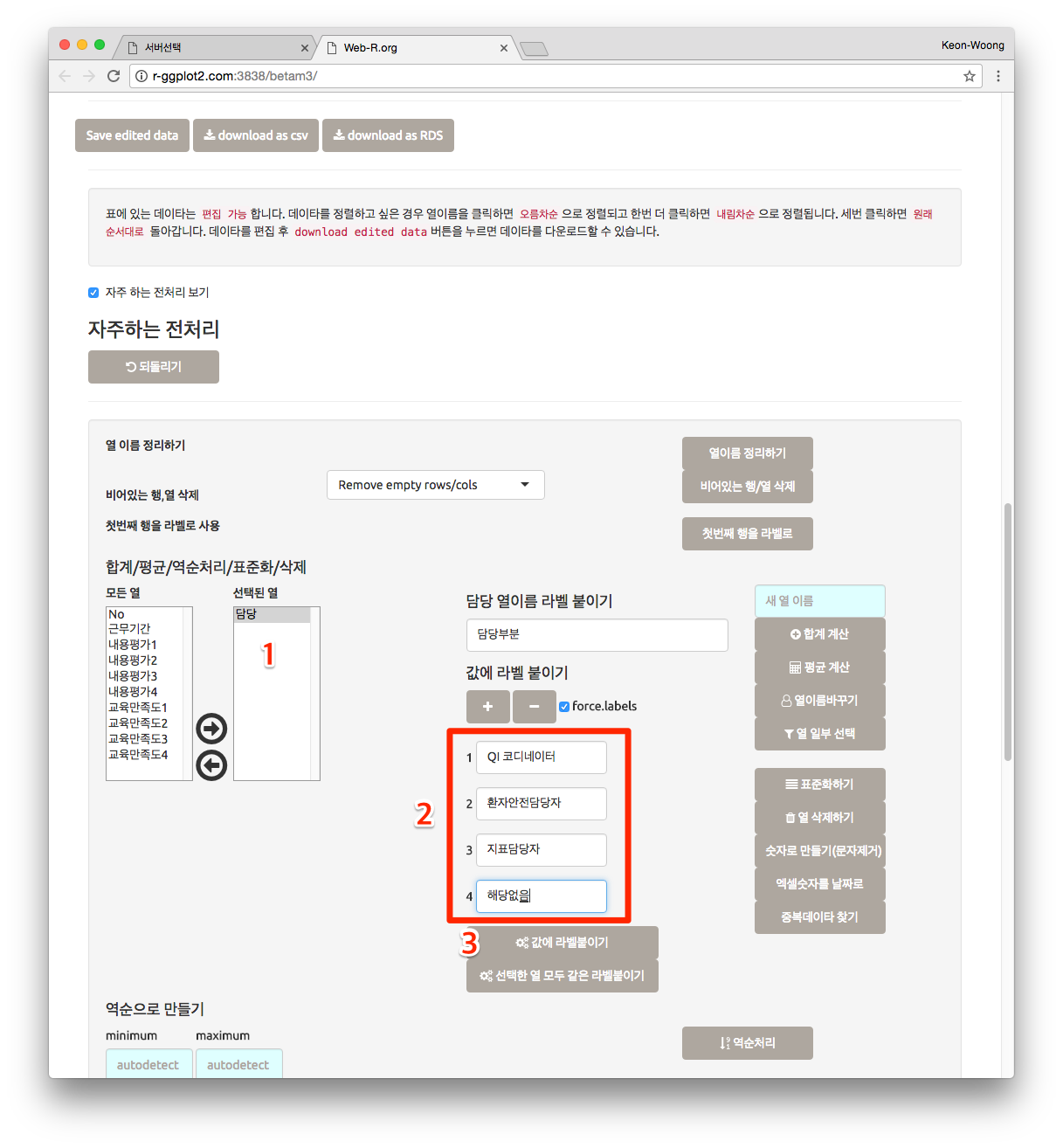
## 첫번째 행을 라벨로

화면을 아래로 이동해보면 다음과 같이 데이타 테이블이 보인다. 모두 55개의 자료이지만 열이름 라벨을 자료로 생각하기 때문에 모두 56개의 자료가 있는 것으로 나타난다(1). 데이타테이블의 아래에는 자주하는 전처리가 있다. 이중 첫번째 행을 라벨로 버튼을 누른다(2).



## 라벨 붙이기- 담당

화면을 조금 아래로 이동한 후 담당을 선택하고(1) 1-4까지 해당하는 라벨을 입력한 후(2) 값에 라벨붙이기 버튼을 누른다(3).



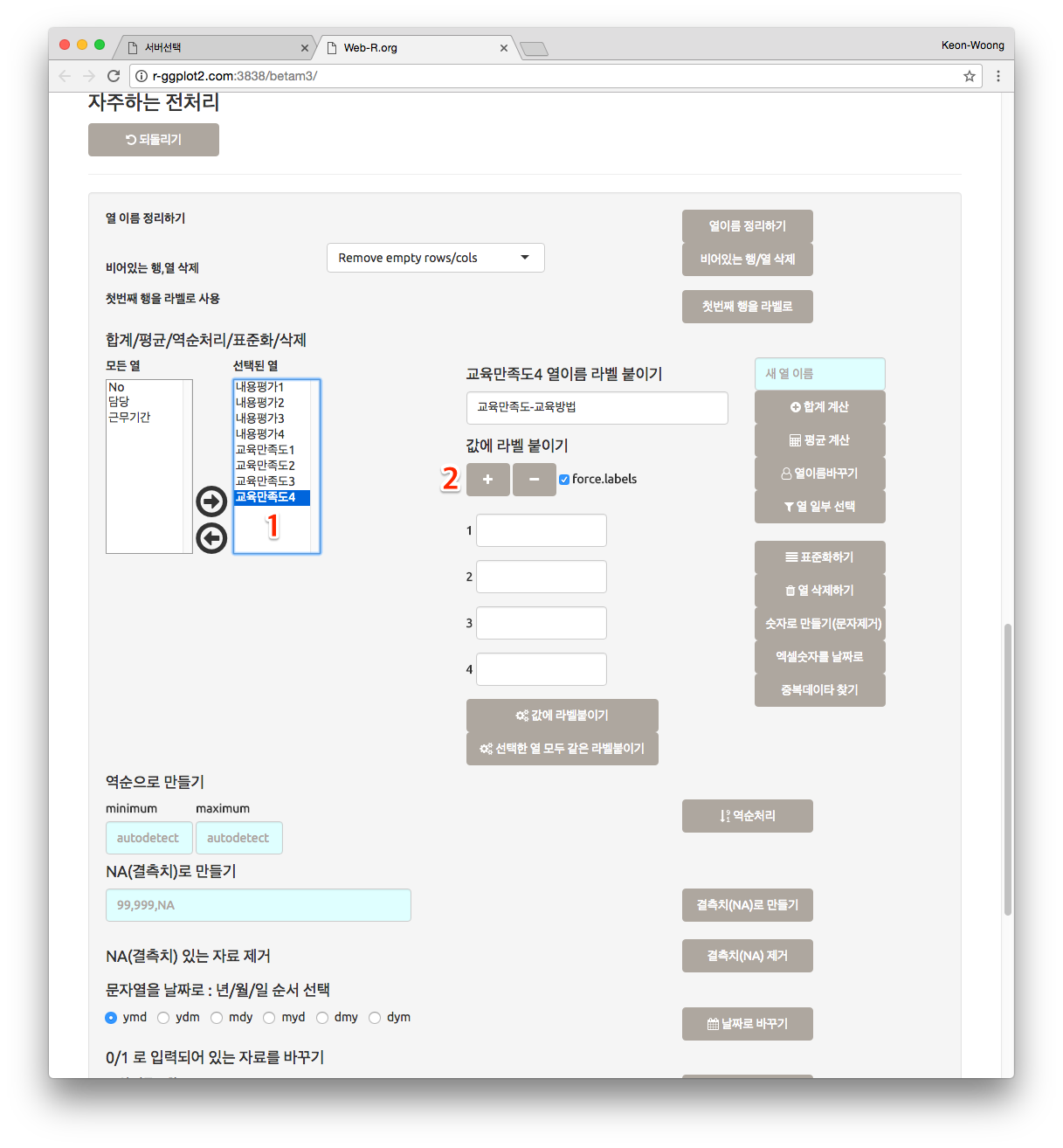
## 라벨 붙이기 - 근무기간

근무기간을 선택하고(1) 1-8까지 해당하는 라벨을 입력한 후(2) 값에 라벨붙이기 버튼을 누른다(3).

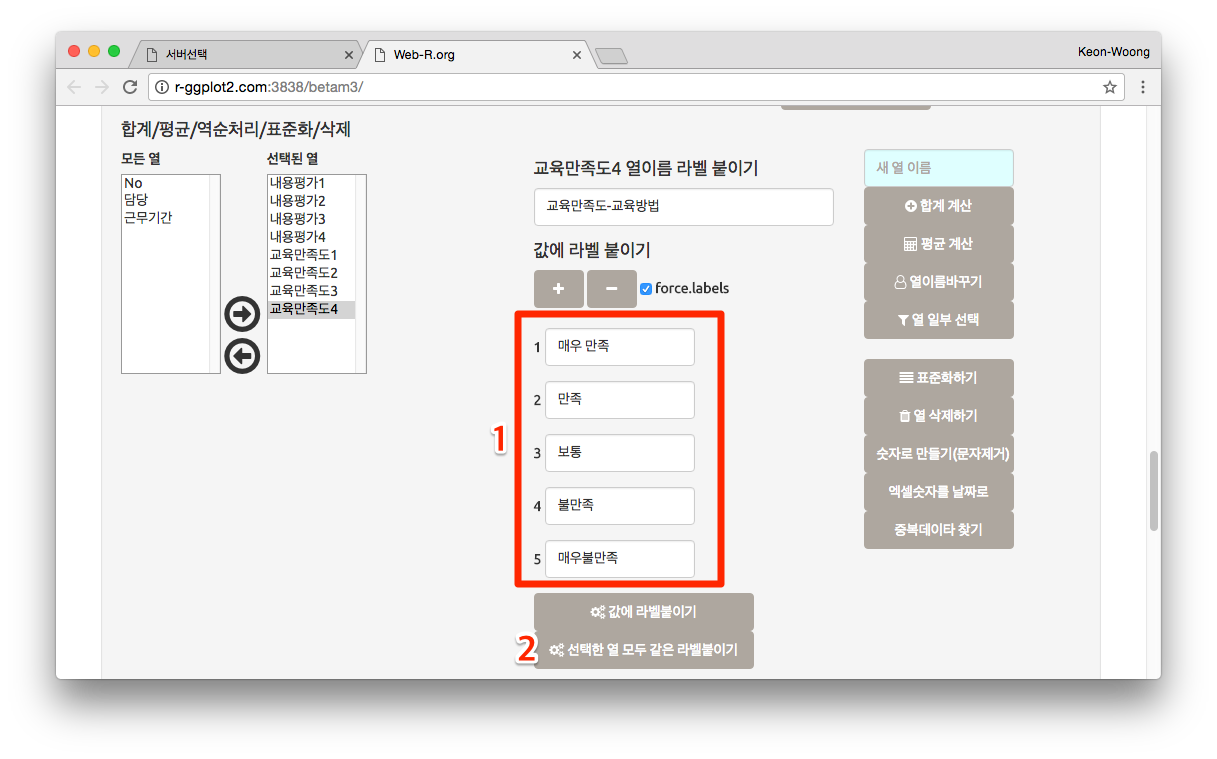


## 한꺼번에 라벨 붙이기 - 내용평가 및 교육만족도

내용평가1-4, 교욱만족도1-4를 선택한 후 교육만족도4 하나만 선택해보면(1) 1-4까지 빈칸이 보인다. 이 항목들은 5단계 리커트 척도를 사용하였는데 리커트 척도를 사용하는 경우 응답자는 극단적인 선택을 피하려는 경향이 있기 떄문에 매우 불만족을 선택한 사람이 아무도 없는 것을 알 수 있다. 모두 5단계가 있으므로 + 버튼(2)을 누른다.



계속해서 1-5단계까지 라벨을 입력하고(1) 선택한 열 모두 같은 라벨붙이기 버튼을 누른다.

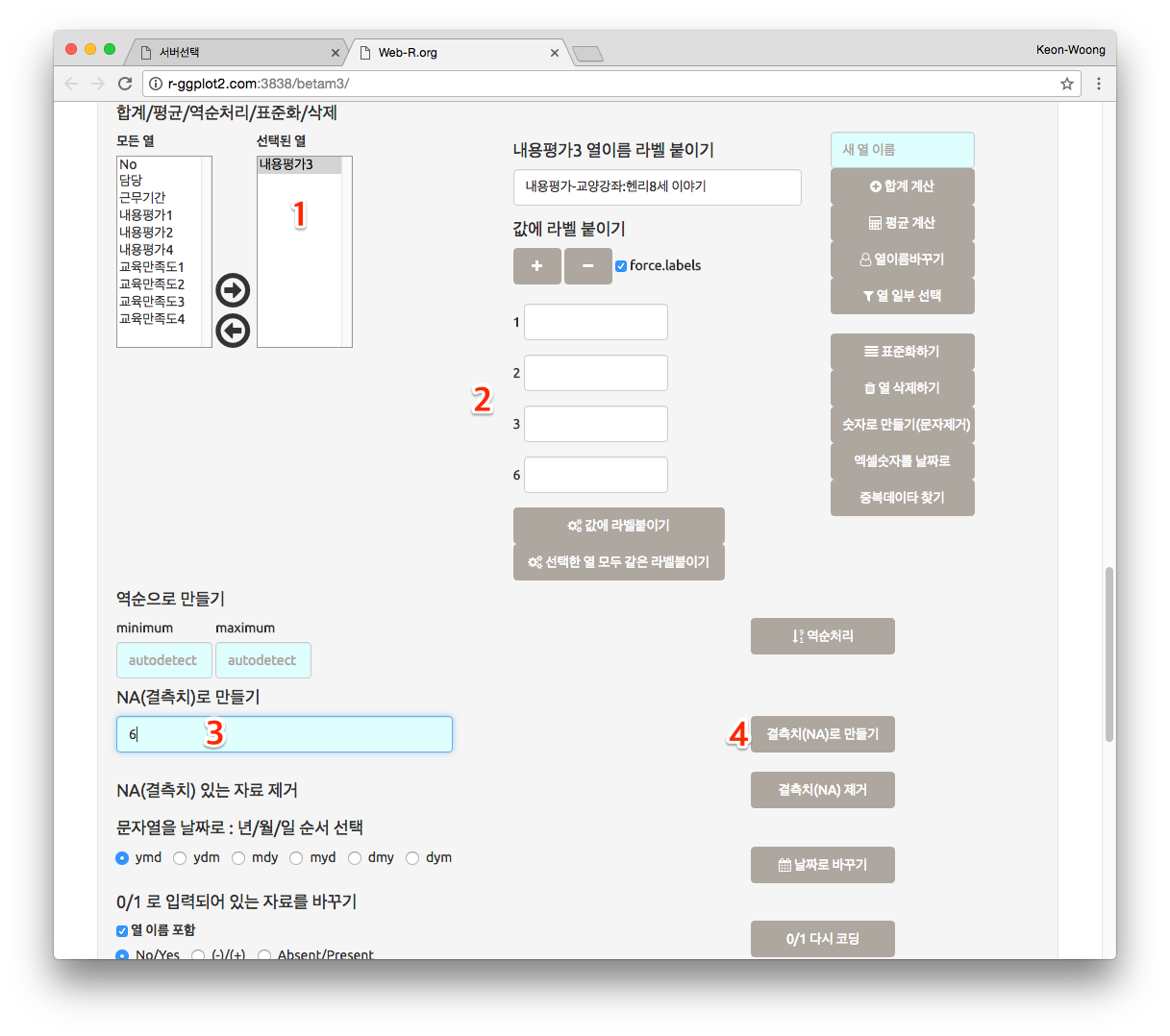


다음과 같은 경고메시지가 나타난다. 내용평가3은 한명이 값을 입력하지 않아 6으로 코딩한 값이 있어 이와 같은 경고메시지가 나타난다. 당황하지 말고 확인 버튼을 누른다.



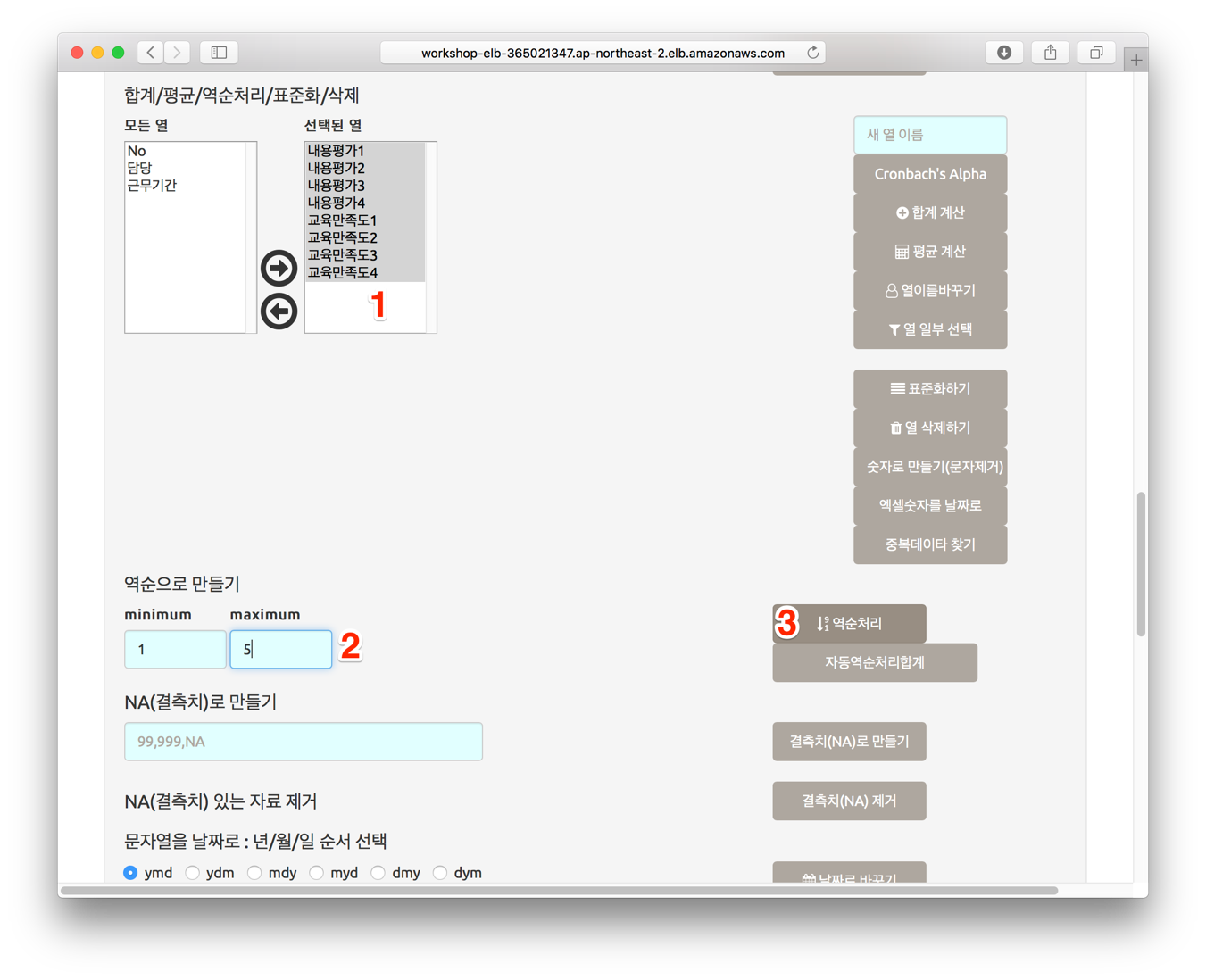
## 결측치의 처리

내용평가3을 선택하면(1) 1,2,3,6 이 있는 것이 보인다(2). 이중 6은 결측치이다. 화면 아래쪽에 있는 NA(결측치)로 만들기에 6을 입력하고(3) 결측치(NA)로 만들기 버튼을 누른다(4).



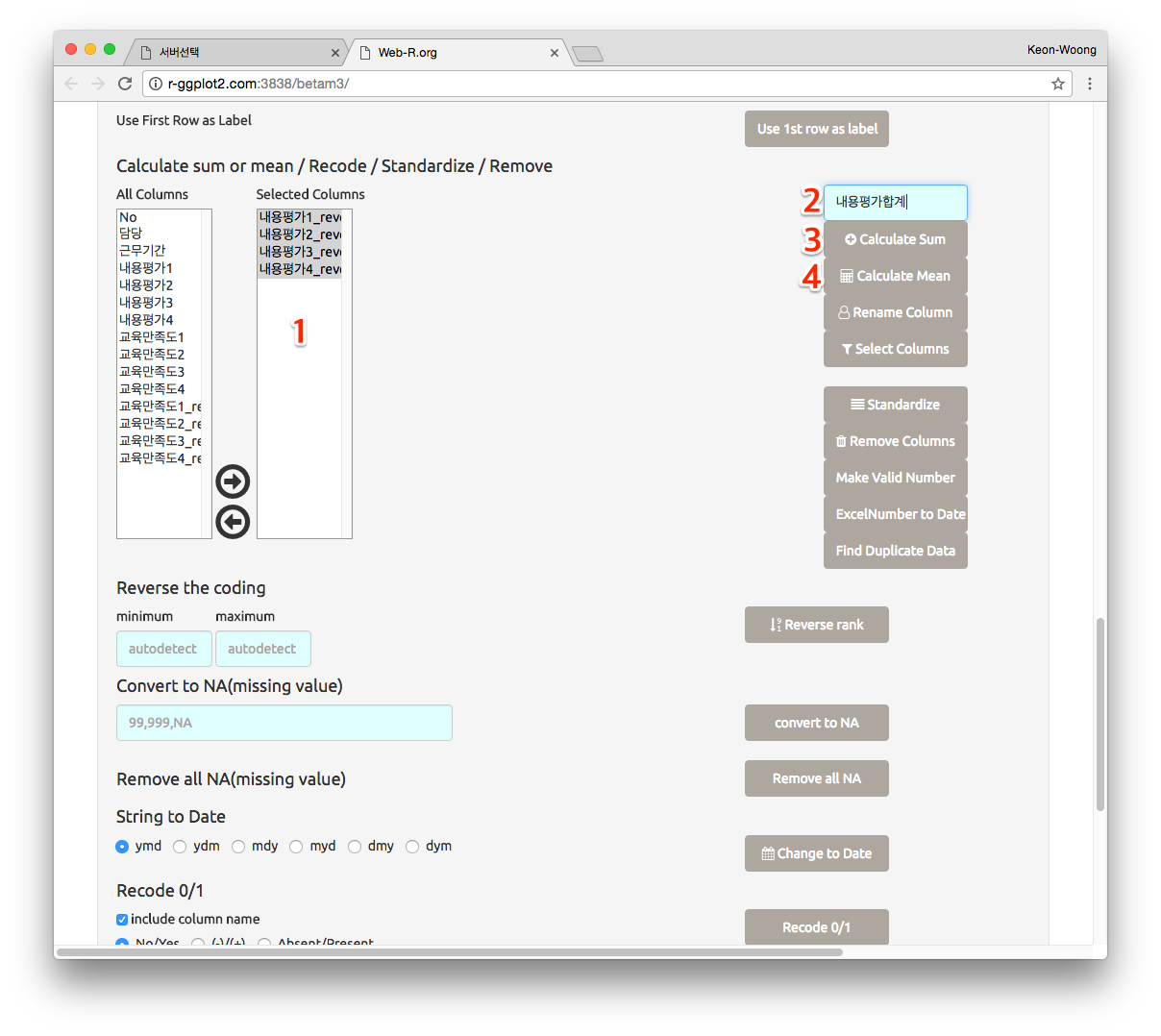
## 코딩 순서 뒤집기 (Reverse the coding)

내용평가1-4 와 교육만족도1-4는 모두 매우 만족이 1, 매우 불만족이 5로 코딩되어 있다. 이를 뒤집어 매우 만족이 5, 매우 불만족이 1로 바꾼다면 점수가 높을수록 만족도가 높다는 것을 알수 있을 것이다. 이렇게 항목의 일부의 순서를 뒤집으려면 다음과 같이 한다. 먼저 내용평가1-4, 교육만족도1-4를 모두 선택하고(1) 역순으로 만들기 의 minimum에 1, maximum에 5를 입력한 후(2) 역순처리 버튼을 누른다(3).



## 합계 및 평균 구하기

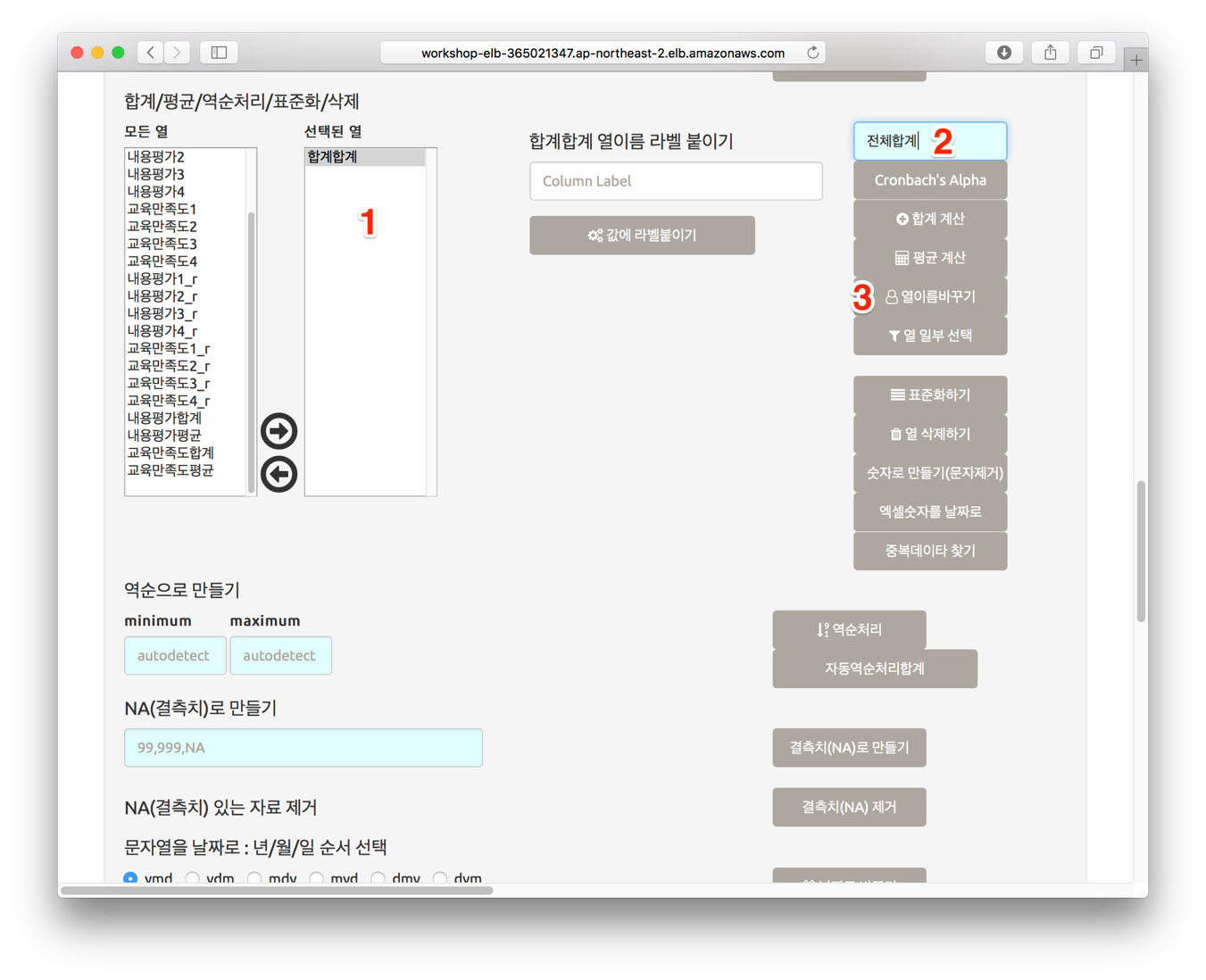
내용평가 및 교육만족도의 의 합계 및 평균을 구해보자. 먼저 내용평가1\_r부터 내용평가4\_r을 선택한 후 합계계산버튼을 누른다(2).



내용평가합계라는 항목이 자동으로 만들어진다. 계속해서 내용평가1-4\_r을 선택한 후 평균계산버튼을 누른다. 교육만족도 합계 및 평균도 같은 방법으로 구한다. 또한 내용평가1-4\_r, 교육만족도1-4\_r 모두의 합계 및 평균을 구한다.

## 전체합계/평균 구하기

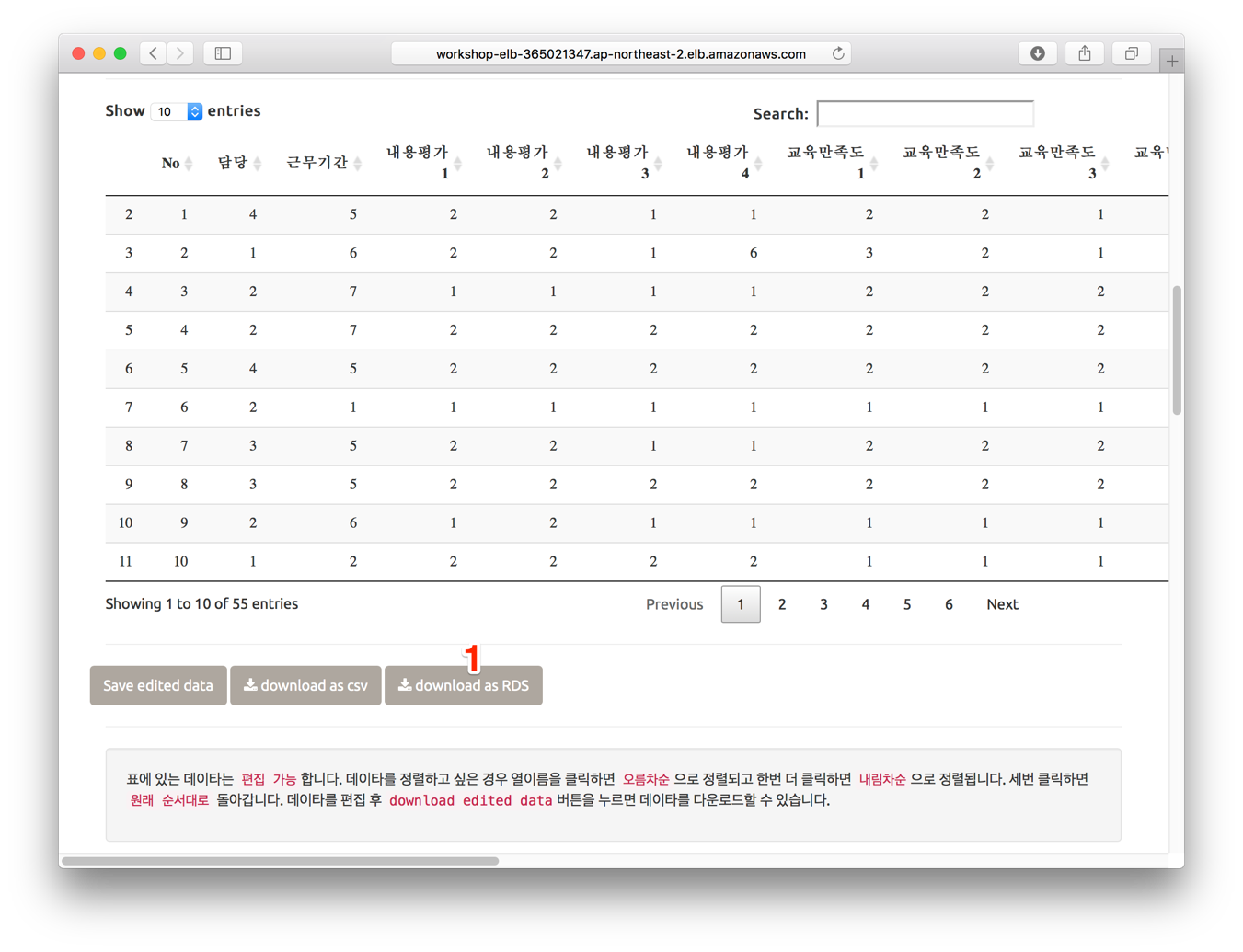
전체합계를 구하기 위해 내용평가 합계와 교육만족도 합계를 선택하여 합계계산을 한다. 이떄 자동으로 두개 항목의 공통되는 단어인 합계를 이용하여 합계합계라는 열이 만들어진다. 합계합계를 선택하고 새열이름에 전체합계를 입력한후 열이름바꾸기 버튼을 누른다.



계속해서 전체평균도 마찬가지 방법으로 구한다.

## 자료의 저장

라벨을 모두 붙인 자료를 다음에 이용하기 위해 내 컴퓨터로 다운로드한다. download as RDS 버튼(1)을 누르면 “Mydata.RDS”라는 이름으로 저장된다. 참고로 download as csv로 저장하면 csv형식으로 저장되나 이때에는 라벨은 저장되지 않는다. 저장된 파일을 확인하기 위해 다시 자료업로드 버튼을 누르고 저장한 파일을 업로드한다.

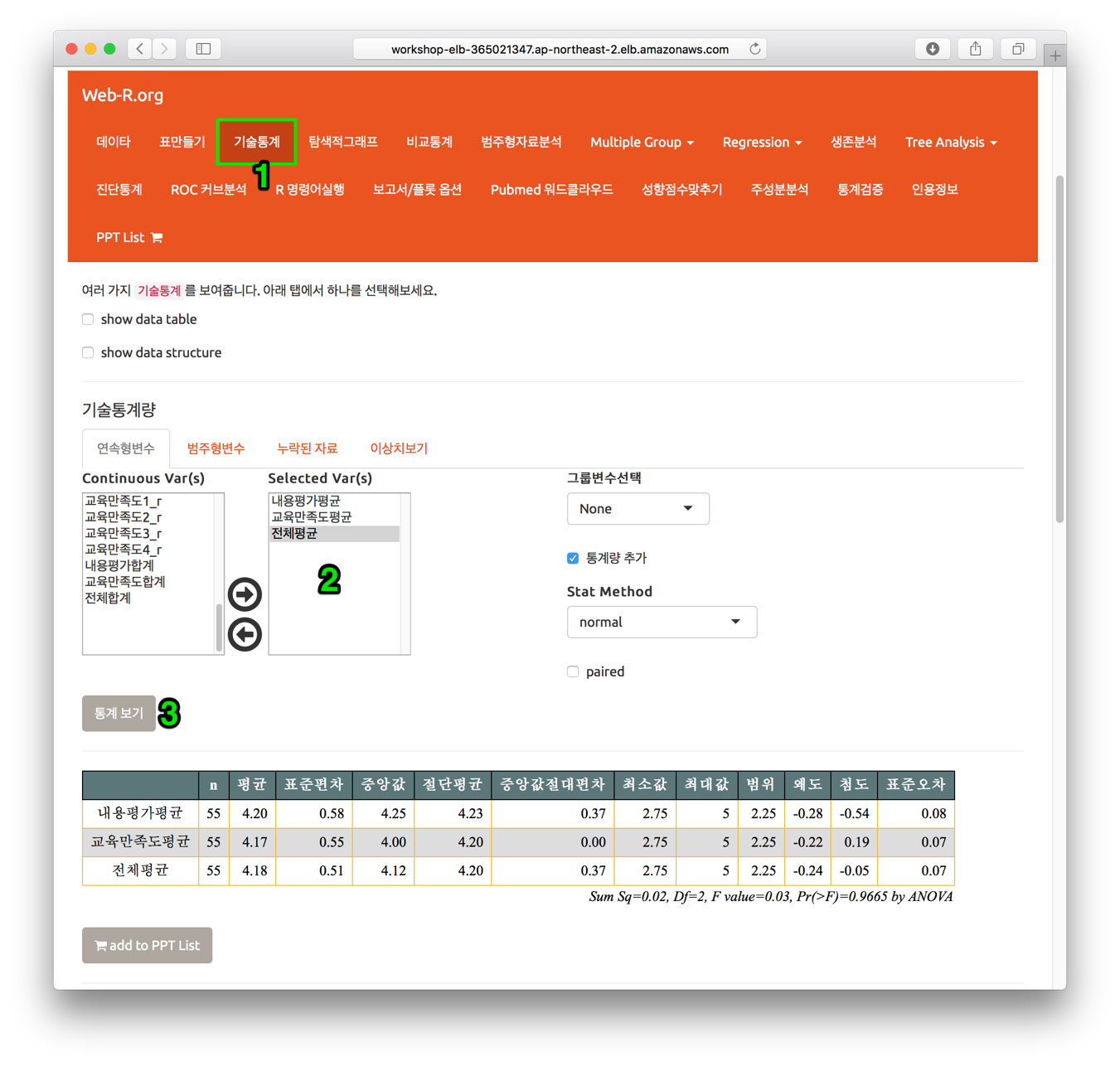


## 자료 분석

담당부분 별로 교육의 만족도에 차이가 있는지 분석해본다. 메인 메뉴에서 표만들기를 선택한 후 그룹변수에 담당을 선택하고(1) 행변수선택에서 select all variables(2)를 선택한후 표만들기 버튼을 누른다(3).****

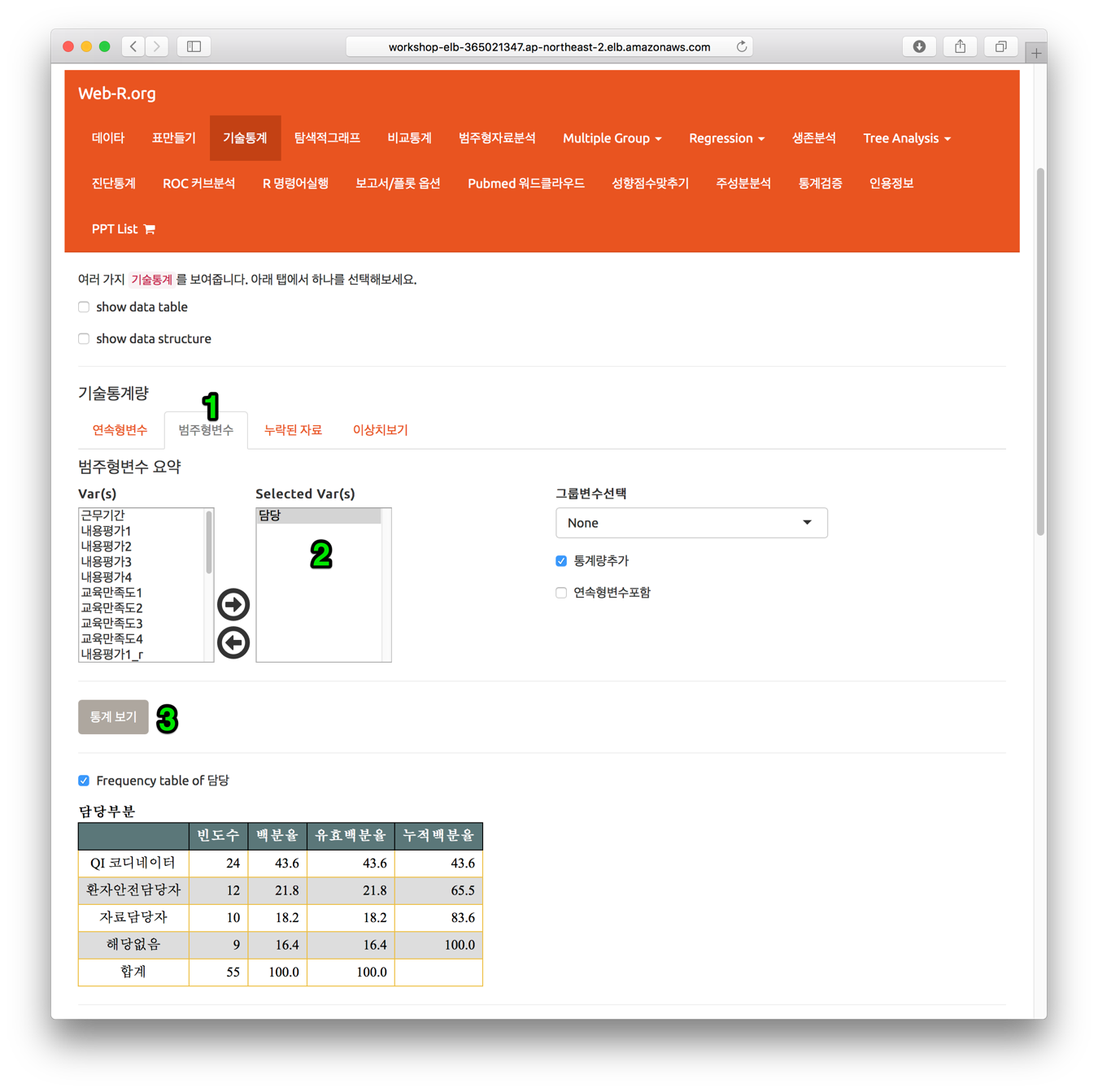
## 기술통계보기 - 연속형변수

메인메뉴에서 기술통계를 선택하고(1) 기술통계량 중 내용평가평균, 교육만족도평균, 전체평균을 선택한 후(2) 통계보기 버튼을 누르면 다음과 같은 표를 얻을 수 있다.

****

## 기술통계보기 - 범주형변수(1)

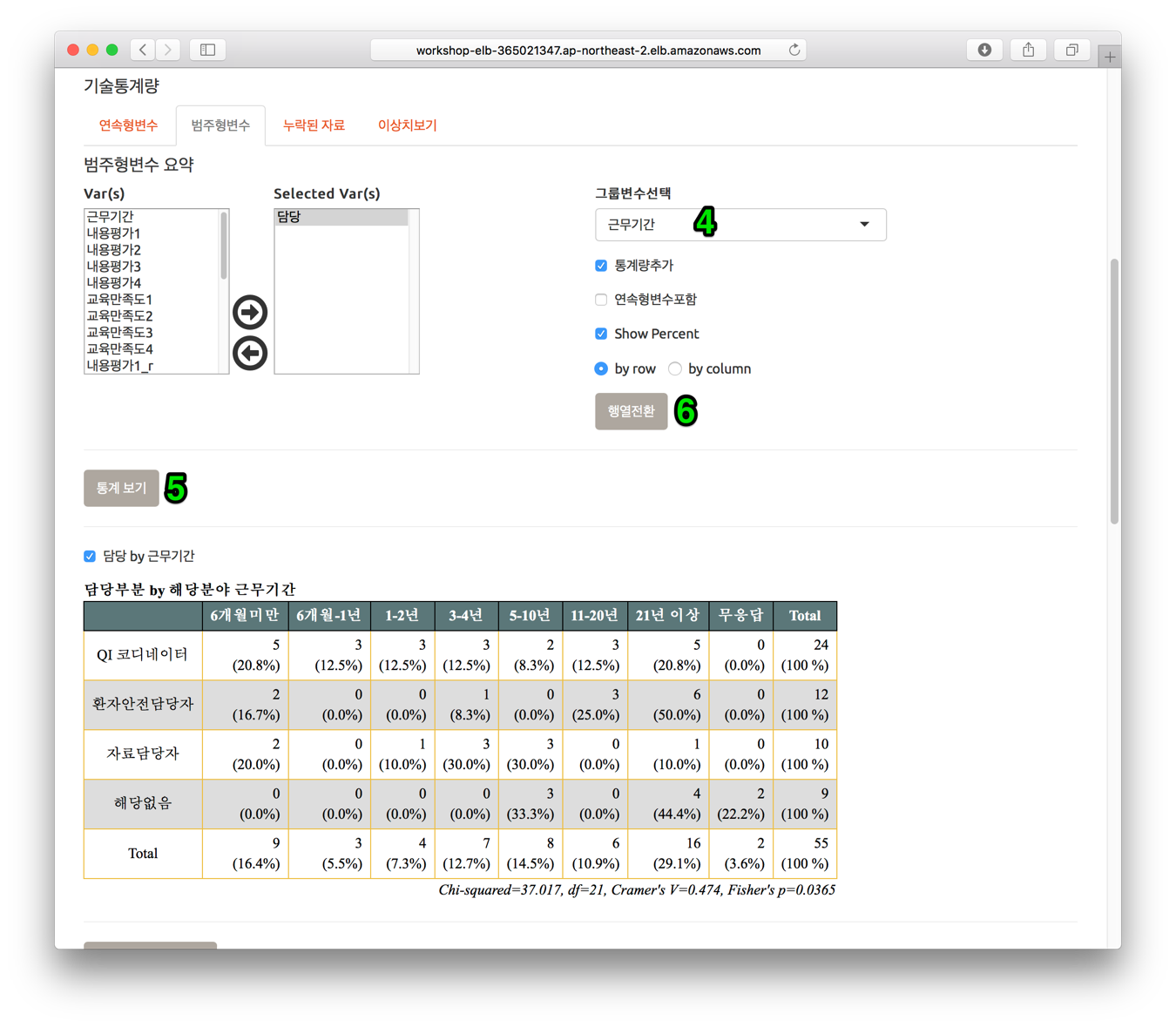
기술통계량 중 범주형변수 탭을 선택하고(1) 먼저 ‘담당’ 변수를 선택하고(2) 통계보기를 누르면(3) 통계량을 볼 수 있다.



## 

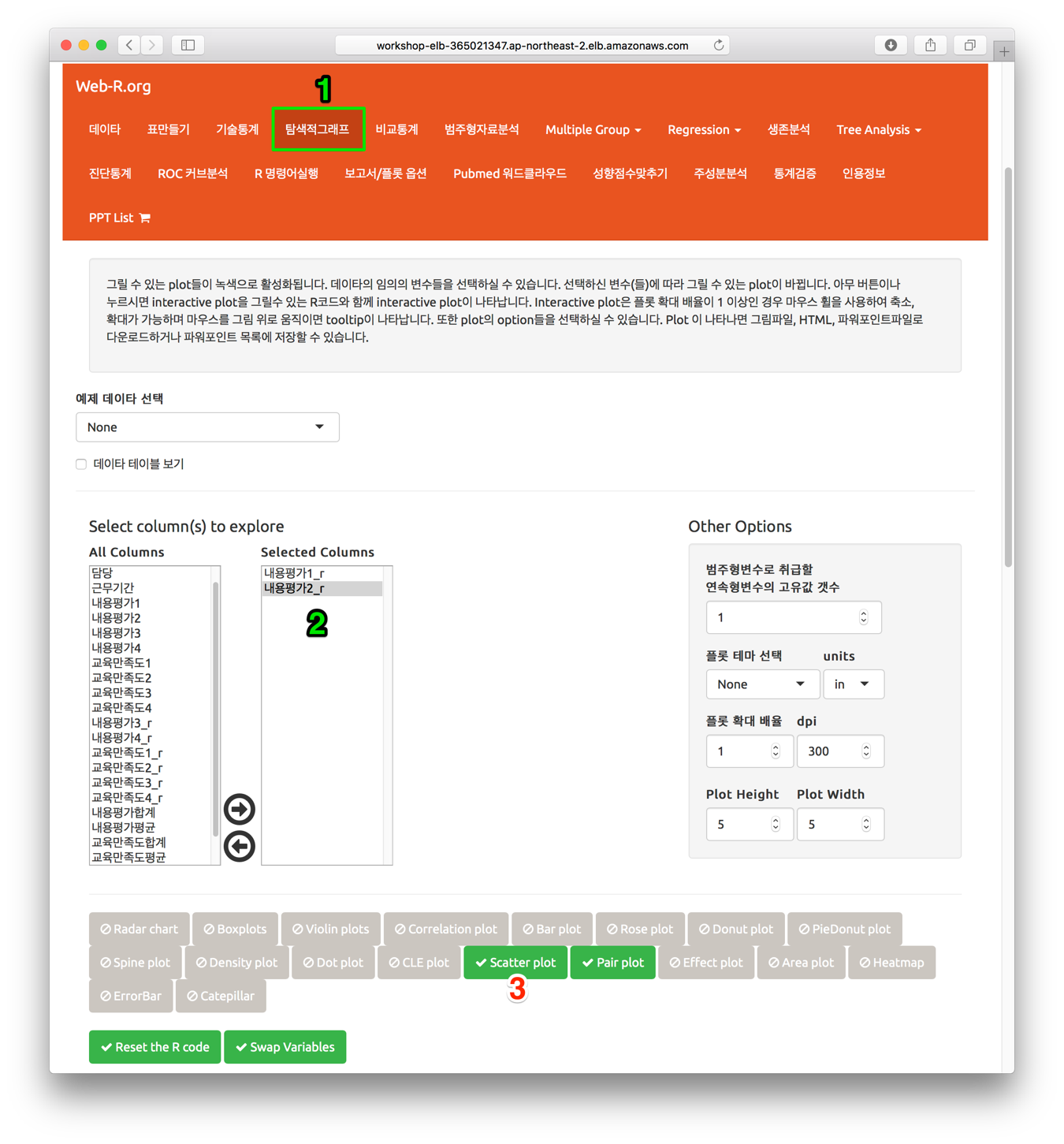
## 기술통계보기 - 범주형변수(2)

계속해서 그룹변수선택에 근무기간(4)을 선택하고 통계보기버튼을 누르면(5) 그룹별로 요약된 표를 볼 수 있다. 또한 행열전환 버튼(6)을 누르고 통계보기 버튼을 누르면 표의 행과 열이 바뀌는 것을 알 수 있다.



## 탐색적그래프

메인 메뉴 중 탐색적그래프를 선택한다(1). 열 이름 중 내용평가1\_r과 내용평가2\_ r을 선택한 후(2) 활성화된 버튼들 중 Scatter plot 버튼을 누른다(3).

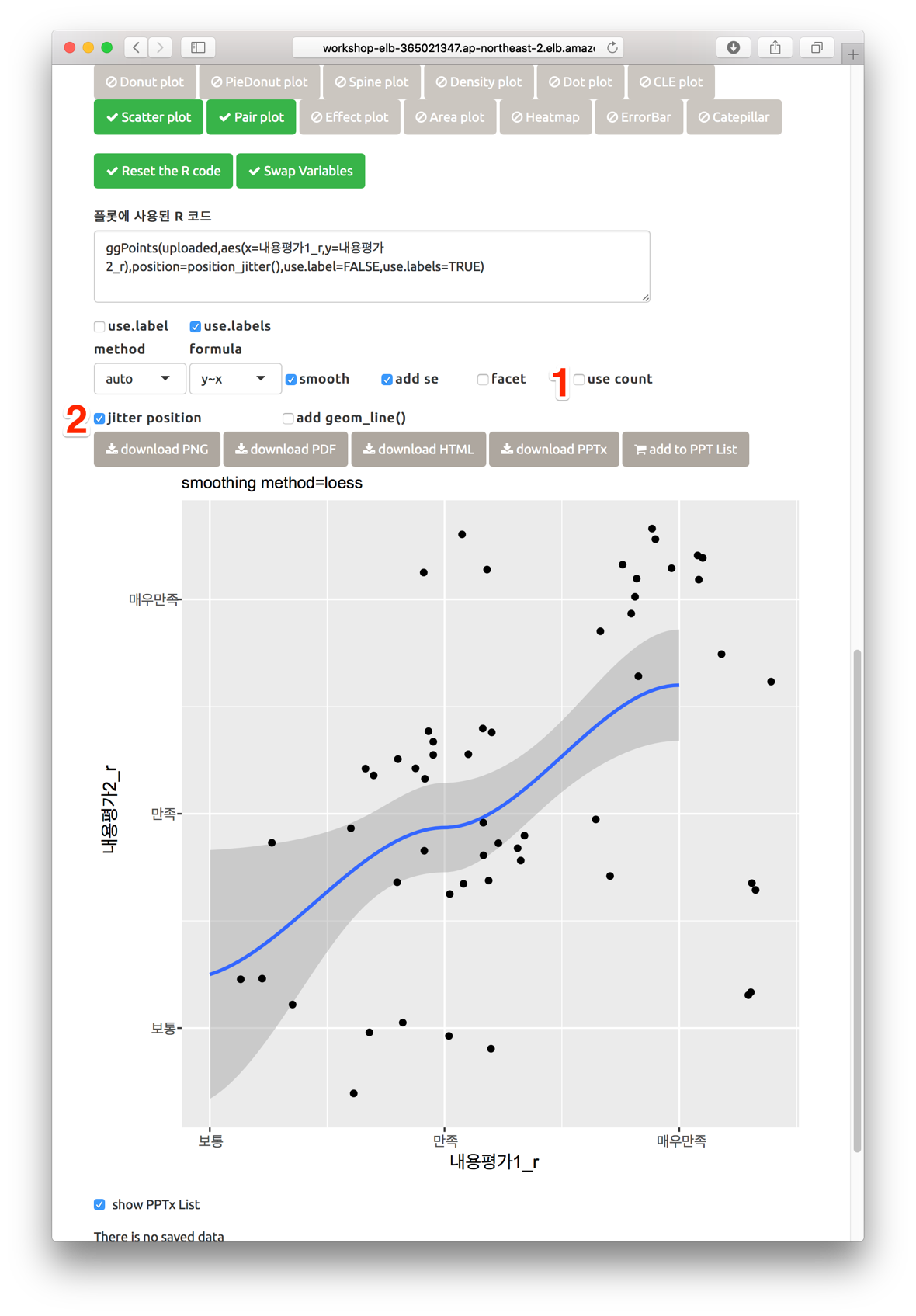


화면을 아래쪽으로 이동해보면 플롯에 사용된 R코드(1)와 함께 플롯(2)이 보인다. 이때 보이는 플롯은 상호작용이 가능한 플롯으로 점 위에 마우스커서을 위치시키면 툴팁을 볼 수 있다. 이 플롯을 그림파일로 저장(3)하거나 PDF 또는HTML파일로 저장할 수 도 있고(4) 파워포인트 파일로 다운로드(5)하거나 파워포인트리스트에 저장(6)하여 나중에 한꺼번에 파워포인트 또는 워드 파일로 다운로드 할 수 있다.



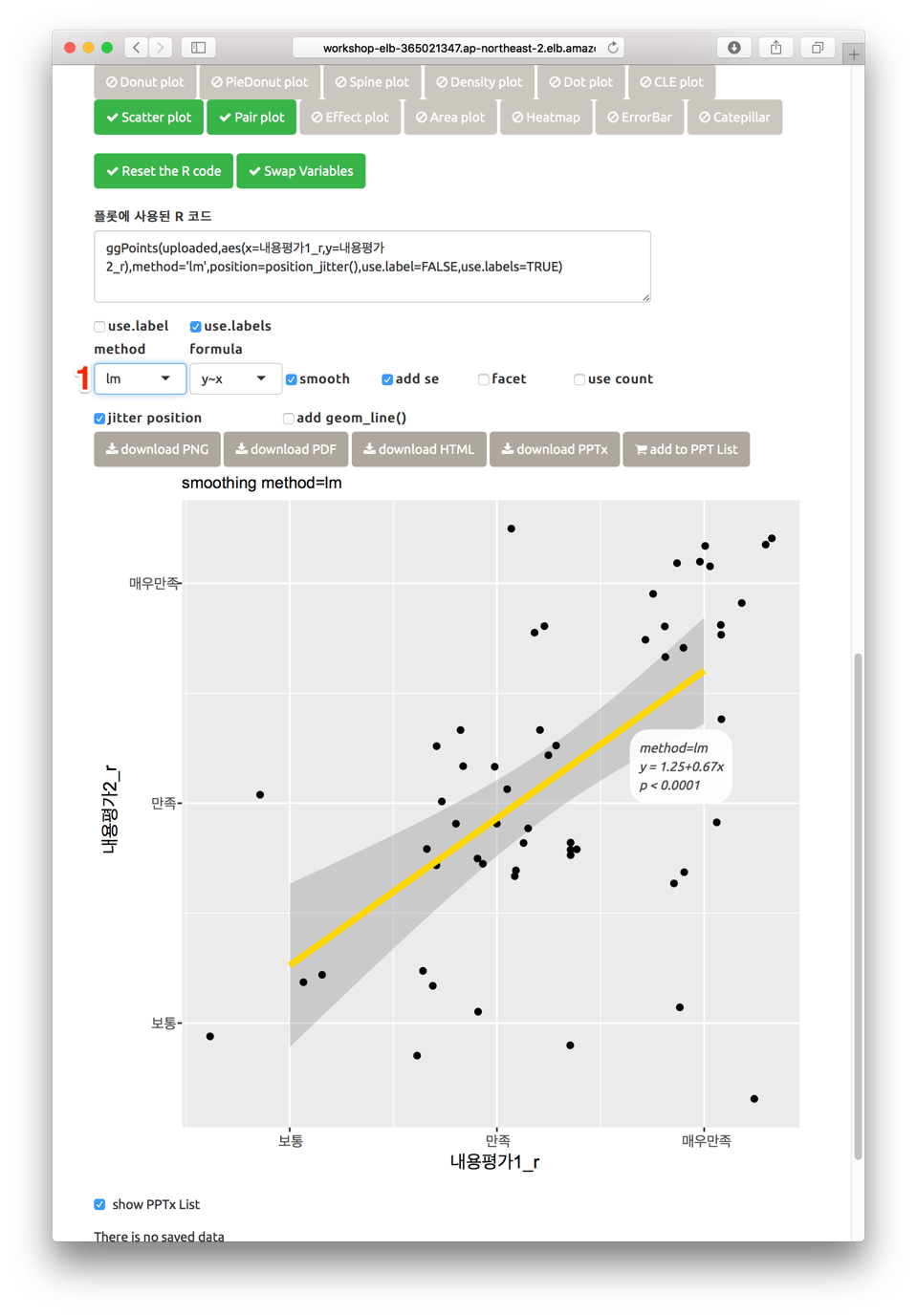
그런데 모두 55개의 자료인데 점의 갯수는 9개밖에 보이지 않는다. 그 이유는 중복된 자료가 많아 겹쳐 보이기 때문이다. 이럴 때 사용할 수 있는 방법은 겹쳐진 갯수에 따라 점의 크기를 달리하거나 점의 위치를 무작위로 흩뜨려서 점을 보이게 하는 방법이 있다. 먼저 겹쳐진 자료의 갯수에 따라 점의 크기를 조절하기 위해서는 use count 체크박스를 선택(1)하면 된다.



두번 째로 점의 위치를 무작위로 흩뜨리기 위해서는 use count 체크박스를 선택해제(1)하고 jitter position 체크박스를 선택하면 된다(2).****

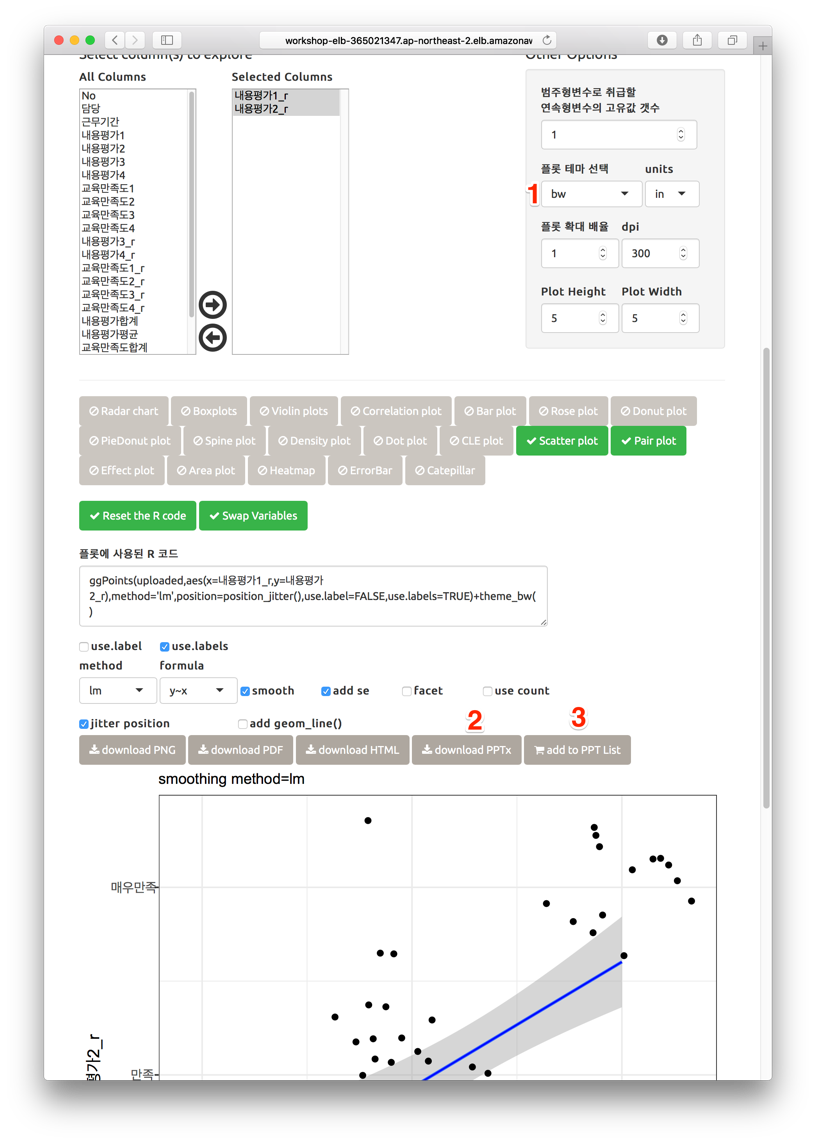
### 회귀방법의 선택

웹R을 이용해 산점도를 그리면 회귀선을 볼 수 있는데 디폴트로는 비선형회귀인 loess회귀가 보여진다. 이를 익숙한 선형회귀로 바꾸려면 method를 lm(=linear model)로 바꾸면 된다(1). 선형회귀인 경우 회귀선 위에 마우스커서를 위치시키면 회귀공식을 볼 수 있다.



## 플롯의 테마의 선택 및 파워포인트 파일로 저장

탐색적그래프에서는 플롯의 테마를 선택할수 있다. 디폴트로 플롯의 테마는 gray 테마로 되어있는데 이는 회색바탕의 그래프이다. 플롯테마 선택(1)에서 bw로 바꾸면 흑백에 맞는 테마로 바뀐다. 다양한 테마를 선택해보고 그림이 어떻게 바뀌는지 보기 바란다. 특히 wsj(wall street journal 테마)및 economist테마 등을 꼭 한번 실행해보기를 바란다. download pptx 버튼(2) 을 누르면 플롯이 파워포인트 형식으로 다운로드 된다 이때 다운로드 되는 파일은 그림이 아니라 벡터형식으로 되어 있어 수정이 가능하다. 하나의 그림을 파워포인트 파일로 저장할 수 도 있고 파워포인트 리스트에 넣었다가 나중에 한꺼번에 파워포인트 파일로 출력할 수 도 있다. add to PPT List 버튼(3)을 누르면 파워포인트 리스트에 저장할 수 있다.



## 연속형변수를 범주형변수로 취급

여기서 다루는 교육 만족도는 1-5로 코딩되어 있어 숫자로 취급할 수도 있으나 각 등급을 갖는 범주형변수로 취급할 수도 있다. 범주형 변수인 경우 막대그래프 등을 그릴 수 있는데 그러기 위해서는 먼저 범주형변수로 취급할 연속형 변수의 고유값 갯수를 6으로 입력한다(1). 활성화되는 버튼들이 달라지는데 barplot 버튼(2)을 누르면 다음과 같은 플롯을 얻을 수 있다.

