빅데이터 예측분석 4

서강대학교 정보통신 대학원 정 화민 교수 (MIS Ph.D)

R 설치

r studio 사용법 pdf

전체 이미지 동영상 뉴스 지도 더보기 설정 도구

검색결과 약 32,200개 중 3페이지 (0.32초)

R 설치, 기초 사용법 – Project Elite

projectl33t.xyz/archives/244 ▼

2016. 10. 17. - 또한 **R-Studio**는 관리자 권한으로 실행되도록 해주세요. ... 뿐만 아니라 출력된 시각화 자료를 이미지나 **pdf** 파일로 저장하는 기능도 있습니다.

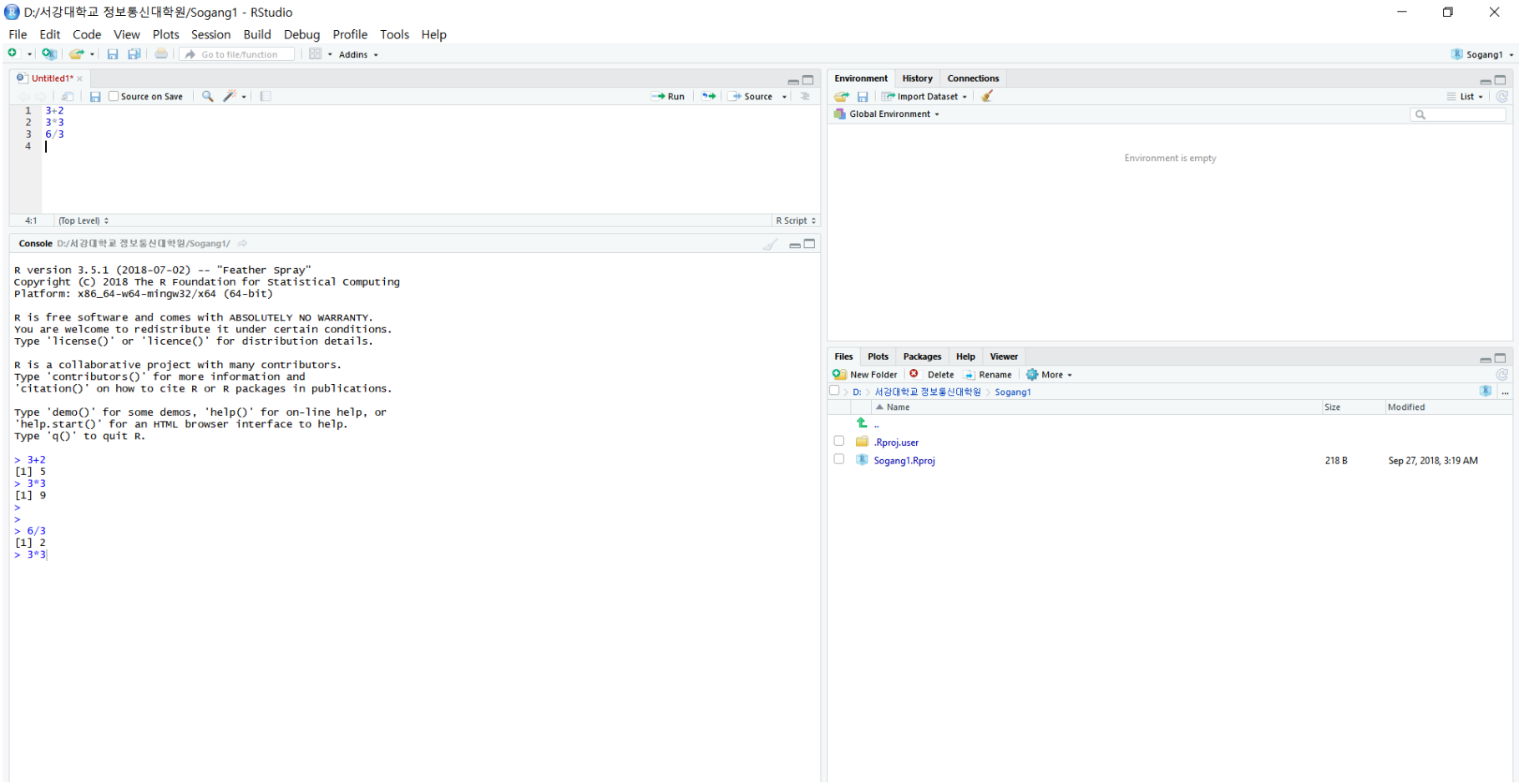
R markdown 사용법 - Amazon AWS

https://rstudio-pubs-static.s3.amazonaws.com/369288_2e88c27061494355977e770ece7...

R과 **RStudio** 설치방법: eclass의 동영상 참고; 라이브러리 설치 ... knitr을 이용하면 R코드와 설명을 html, latex, **pdf** 등 다양한 형태로 변환 가능; R script(확장자는 .

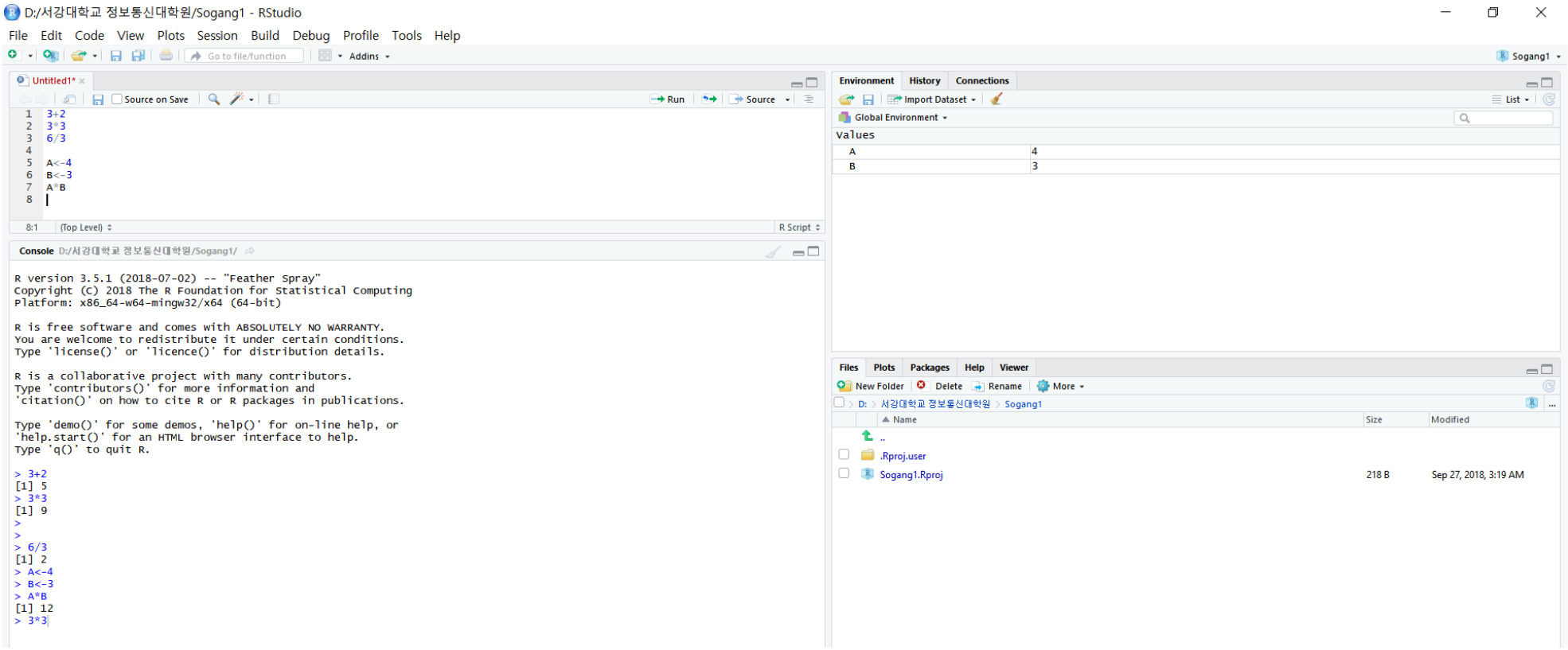
R Studio에서 New Project 실행

R studio 스크립트 창에서 간단한 연산 입력하여 콘솔창에 결과값이 나오는지 확인
명령어 실행은 Ctrl + Enter



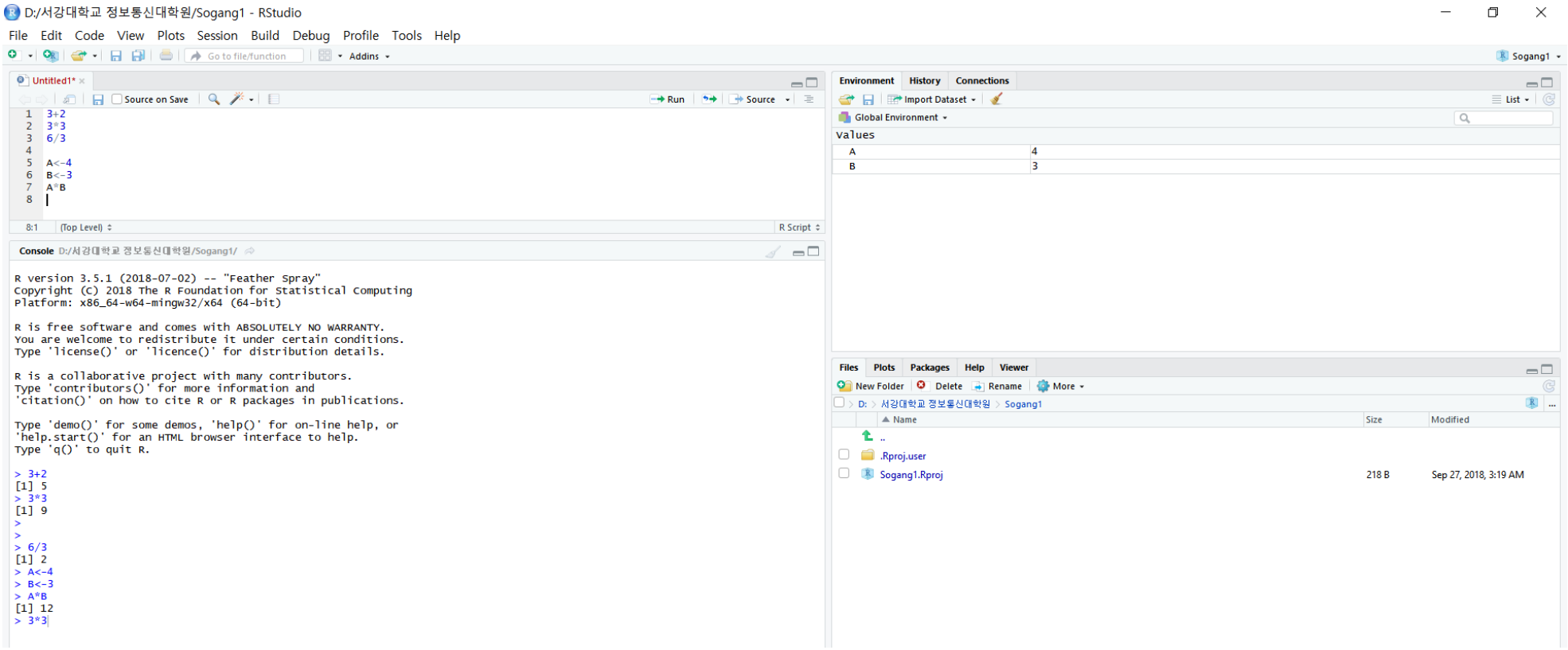
R Studio에서 변수 만들기

R studio 명령어를 이용해 변수 만들기
A라는 변수에 4를 넣어라 (A<- 4), 오른쪽 상단 환경 창에 A변수 생성

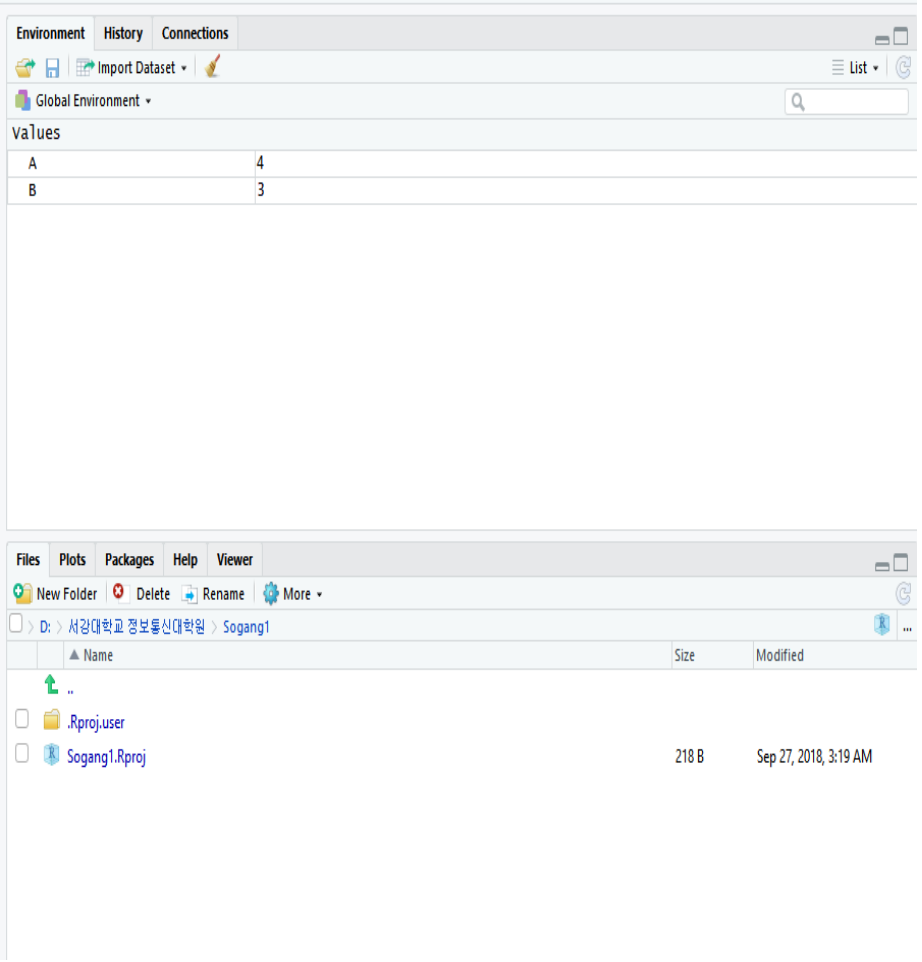


R Studio에서 변수 만들기

R studio 명령어를 이용해 변수 만들기
A라는 변수에 4를 넣어라 (A<- 4), 오른쪽 상단 환경 창에 A변수 생성



R Studio 파일창의 기능



Files : 현재 프로젝트 디렉토리

Plots : 그래프 보여주는 곳

Packages : 설치된 패키지 목록을 보여줌

Help: help( )함수를 실행하면 명령어를 설명하는 곳

Viewer : 분석결과를 HTML 과 같은 웹 문서로 출력

R Studio 데이터 프레임 만들기

```
english <- c(70, 90, 80, 60, 70, 60)
english

math <- c(50, 60, 100, 20, 50, 80)
math

df_midterm <- data.frame(english, math)
df_midterm

class <- c(1, 1, 1 ,2, 2, 2)
class

df_midterm <- data.frame(english, math, class)
df_midterm

mean(df_midterm$english)
mean(df_midterm$math)
```

Console D:/빅데이터기획에서분석/Data/

```
[Workspace loaded from D:/鍮낅뜲?댄꽣湲고쉷?먯꽌遺꾩꽍/Data/.RData]

> english <- c(70, 90, 80, 60, 70, 60)
> english
[1] 70 90 80 60 70 60
> math <- c(50, 60, 100, 20, 50, 80)
> math
[1]  50  60 100  20  50  80
> df_midterm <- data.frame(english, math)
> df_midterm
  english math
1      70   50
2      90   60
3      80  100
4      60   20
5      70   50
6      60   80
> class <- c(1, 1, 1 ,2, 2, 2)
> class
[1] 1 1 1 2 2 2
> df_midterm <- data.frame(english, math, class)
> df_midterm
  english math class
1      70   50     1
2      90   60     1
3      80  100     1
4      60   20     2
5      70   50     2
6      60   80     2
> mean(df_midterm$english)
[1] 71.66667
> mean(df_midterm$math)
[1] 60
>
```

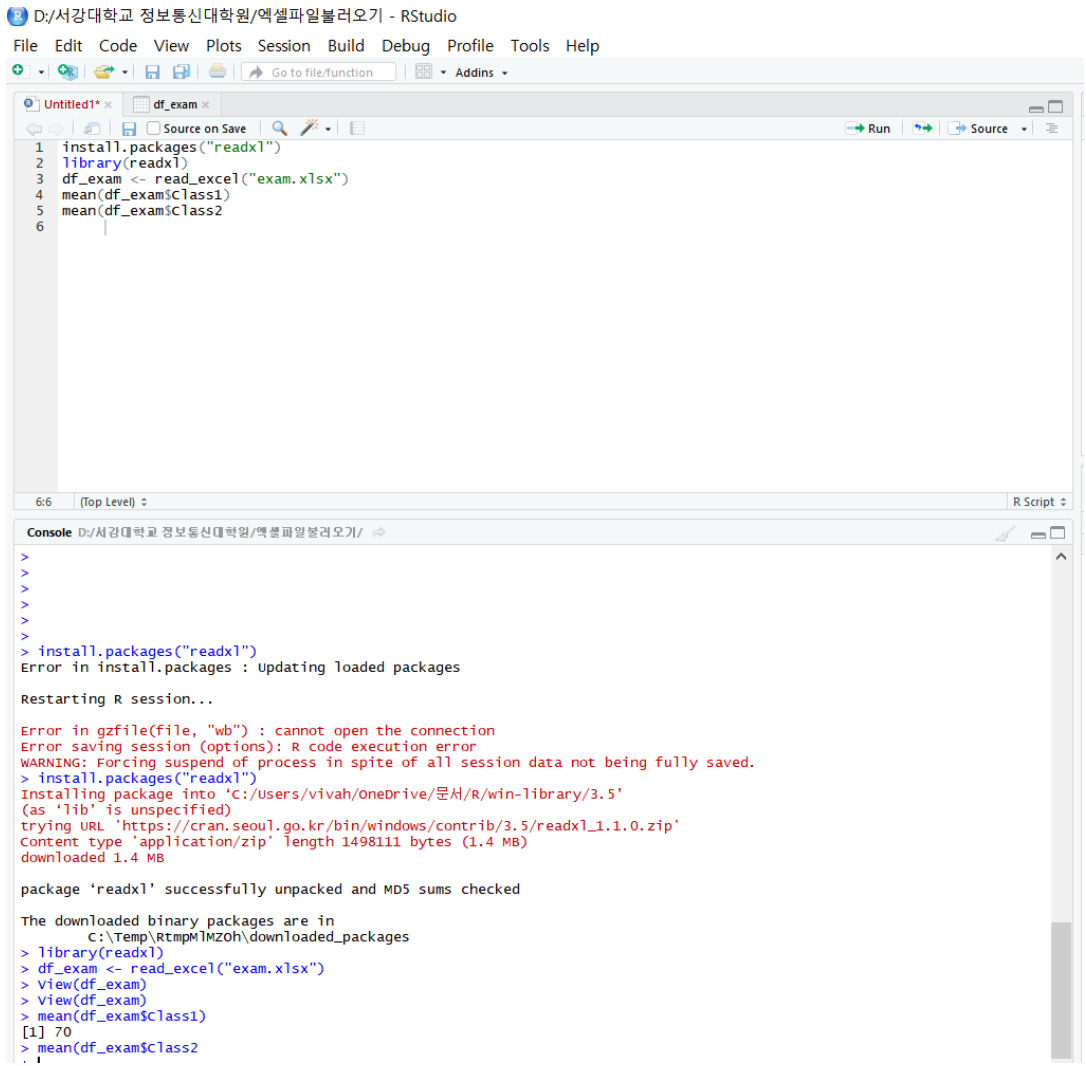
영어점수 변수생성
수학점수 변수생성
각 Class 변수생성

영어와 수학점수를 중간고사 데이터 프레임으로 할당할 것

평균을 구할 것

영어: 71.6점
수학: 60점

R Studio 엑셀 파일 불러오기 (readxl 패키지 설치하고 로드하기)

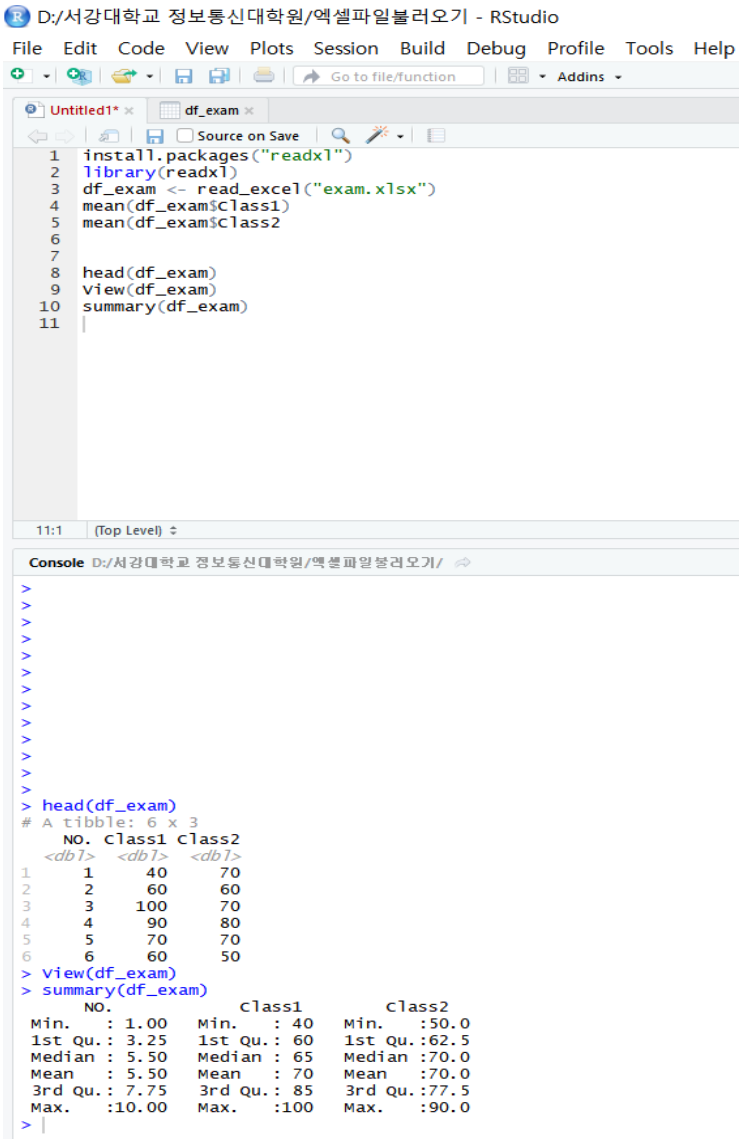


readxl 패키지 설치하고 로드하기

```
install.packages("readxl")
library(readxl)
df_exam <-
read_excel("exam.xlsx")
mean(df_exam$Class1)
mean(df_exam$Class2
```

Class 1/ Class 2의 평균은 각 70점
(단, 여기서 엑셀파일은 프로젝트 폴더안에 있어야 함)

R Studio 기술통계(요약통계량 구하기)



head(df_exam)
데이터의 앞부분 출력

View(df_exam)
뷰어창에서 데이터 확인
(단 V는 대문자, 소문자로 입력하면 에러 발생_주의)

summary(df_exam)
요약통계량 출력

요약통계량 : 최소값, 1사분위값 중앙값, 평균, 3사분위 값, 가장 큰 값.

R Studio 요인분석

```
library(xlsx)
install.packages("xlsx")
library(xlsx)
install.packages("psych")
library(psych)
install.packages("nFactors")
library(nFactor)
library(nFactors)
mydata=read.xlsx(FACT_ANALYSIS,1)
mydata=read.xlsx("FACT_ANALYSIS.xlsx",1)
View(mydata)
attach(mydata)
names(mydata)
x=cbind(Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12)
e_value=eigen(cor(x))
e_value
PCA=principal(x,nfactors = 2, rotate = "varimax")
print(PCA, digits=3, cutoff=.3, sort = true)
print(PCA, digits=3, cutoff=.3, sort = TRUE)
print(PCA, digits=3, cutoff=.3, sort = TRUE)
PCA=principal(x, nfactors = 2, rotate = "varimax")
print(PCA, digits=3, cutoff=.3, sort = TRUE)
```

```
eigen() decomposition
$`values`
 [1] 6.75147664 2.29192951 0.59300517 0.50986645 0.45256698
 [6] 0.29520845 0.26744906 0.24213504 0.21518752 0.16150030
[11] 0.13527080 0.08440408
```

패키지를 인스톨한 다음 라이브러리를 로드 한다.

아이겐밸류 값 1이상을 기준으로 요인이 몇 개 나올 수 있는지 검증한다.
아이겐밸류값 6.75, 2.29를 기준으로 2개의 요인으로 분류할 수 있다.

(분석용 데이터 파일 FACT_ANALYSIS는 네이버 카페에 올려 놓겠음)

R Studio 요인분석 결과

```
Principal Components Analysis
Call: principal(r = x, nfactors = 2, rotate = "varimax")
Error in print.psych.fa(x, digits = digits, all = all, cut = cut, sort = sort,
  object 'true' not found
> print(PCA, digits=3, cutoff=.3, sort = TRUE)
Principal Components Analysis
Call: principal(r = x, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
    item   RC1   RC2    h2    u2  com
Q10   10 0.897 0.180 0.837 0.163 1.08
Q8     8 0.860 0.215 0.786 0.214 1.12
Q12   12 0.857 0.236 0.791 0.209 1.15
Q9     9 0.856 0.197 0.771 0.229 1.11
Q11   11 0.852 0.281 0.805 0.195 1.22
Q7     7 0.831 0.270 0.763 0.237 1.21
Q4     4 0.187 0.882 0.813 0.187 1.09
Q6     6 0.190 0.835 0.733 0.267 1.10
Q1     1 0.082 0.832 0.700 0.300 1.02
Q3     3 0.250 0.805 0.711 0.289 1.19
Q2     2 0.271 0.783 0.686 0.314 1.24
Q5     5 0.314 0.742 0.649 0.351 1.35
```

요인분석 결과 2개의 요인으로 분류되었다.

2가지 요인 중 Q10은 요인1을 설명하고 있는 설명력이 89.7%를 나타내고, 요인2에서는 Q4가 두 번째 요인을 88.2%를 설명하는 변수로 나타났다.

(데이터에 대한 설명은 수업시간에 자세히 합니다)