

R 과 RStudio 설치 추가내용

0.선수작업 : JDK설치

-JDK다운로드, 설치 및 환경변수 설정

=>Java종속 패키지인 한글 텍스트 마이닝에 필요한 KoNLP패키지 사용 등등에 필요

(1)JDK다운로드

-<http://www.oracle.com/technetwork/java/javase/downloads/index.html> 에서 최신항목을 다운로드 받음.
java.sun.com을 사용해도 됨.

참고:윈도우즈 운영체제에 따른 자바설치 파일

32bit : jdk-버전번호-windows-i586.exe

64bit : jdk-버전번호-windows-x64.exe

(2)설치

-다운로드 받은 jdk-버전-운영체제종류.exe 파일을 더블클릭해서 설치

예) jdk-8u162-windows-x64 또는 jdk-8u162-windows-i586.exe

참고: 레가시 버전 - 실제 실무 버전 : JDK6

(3)환경변수 설정 : 자바명령어 위치 및 다른프로그램(Oracle, 웹프로그래밍, R...)과의 연동을 위해서 자바의 위치를 시스템에 알림

-방법:[제어판]-[시스템]-[고급시스템설정]-[고급]탭에서 [환경변수]버튼을 눌러서 PATH, CLASSPATH, JAVA_HOME 환경변수를 설정

①PATH <-[편집]

:C:\Program Files\Java\jdk1.8.0_162\bin;

②CLASSPATH <-[새로만들기]

.:C:\Program Files\Java\jdk1.8.0_162\lib\tools.jar

③JAVA_HOME <-[새로만들기] :다른프로그램과의 연동

C:\Program Files\Java\jdk1.8.0_162

(4)환경변수 설정 확인

- [시작]버튼 클릭 후 cmd 입력하고 엔터
- javac 입력후 엔터
- java 입력후 엔터

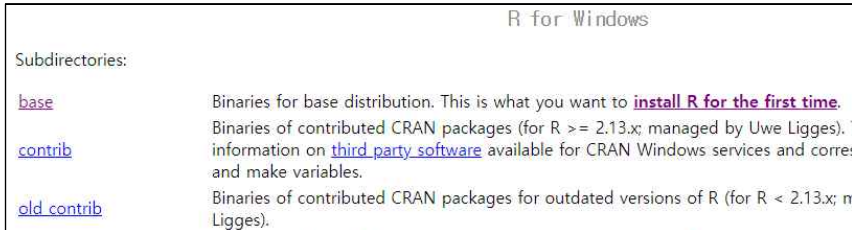
I. R다운로드, 설치 및 패키지 업데이트 - <https://www.r-project.org/>

1.

① <https://cran.r-project.org/mirrors.html> 사이트에서 미러사이트중 하나 선택 후 [Download R for Windows]링크 클릭. 예시)<https://cran.seoul.go.kr/>



② [install R for the first time]링크 클릭



③ [Download R 3.5.1 for Windows]링크 클릭. R프로그램이 다운로드 됨



④ 다운로드 받은 R-3.5.1-win.exe 파일을 더블클릭해서 설치. 기본 값을 그대로 사용하고 [다음]버튼을 클릭하다가 마지막에 [완료 또는 닫기]버튼 클릭

2. R 설치 추가 내용 : R설치 후 R 아이콘에서 마우스 오른쪽 버튼을 눌러 [속성]-[호환성]에서 [관리자 권한으로 이 프로그램 실행]체크하고 [확인]버튼 클릭

3. R 패키지 업데이트 : 최신 버전을 설치한 경우에도 함.

- R아이콘 더블클릭. R이 실행되면 [패키지들]-[패키지 업데이트]메뉴 클릭 또는 RStudio의 [Packages]탭에서 [Update]버튼 클릭

II. RStudio 다운로드 및 설치 : R통합개발환경

1. RStudio 다운로드 - https://www.rstudio.com/

- [RStudio]항목하단의 [Download]버튼 클릭

- [RStudio Desktop(Open source License)]항목의 [DOWNLOAD]버튼 클릭.

- [Installer for Supported Platforms]에서 [RStudio 버전번호 - Windows Vista/7/8/10]항목 클릭. 여기서는 RStudio-버전번호.exe 파일이 다운로드 됨.

->예)RStudio-1.1.453.exe가 다운로드 됨

2. RStudio 설치 : 예) RStudio-1.1.453.exe

- RStudio-1.1.453.exe더블클릭

- RStudio 설치 후 RStudio 단축아이콘 만들기. 경로 예) C:\Program Files\RStudio\bin\rstudio.exe

- 단축아이콘에서 마우스 오른쪽 버튼을 눌러 [속성]-[호환성]에서 [관리자 권한으로 이 프로그램 실행]체크하고 [확인]버튼 클릭

3. RStudio 실행 : [시작]-[모든 프로그램]-[RStudio]에서 RStudio항목 선택 또는 [RStudio]단축아이콘 더블클릭

* [Console] : 패키지 추가설치 및 명령어 실행

4. 프로젝트 관리 폴더 작성 : 탐색기에서 [project] 폴더 생성

5. R 프로젝트 작성

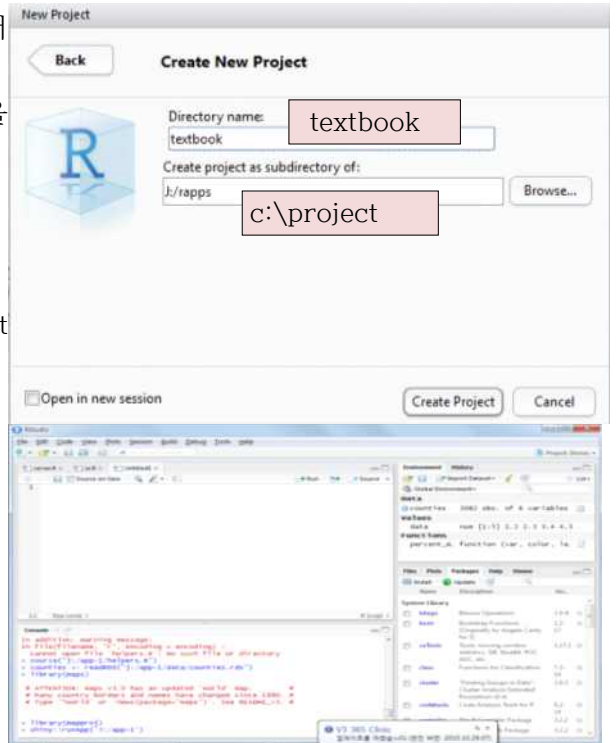
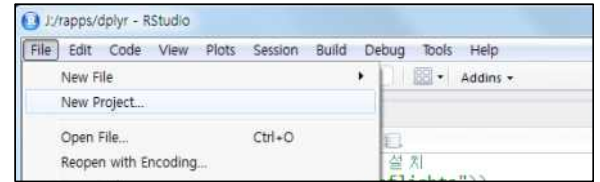
-[File]-[New Project...]메뉴 선택

-[New Directory]항목 선택

-[Empty Project]항목 선택

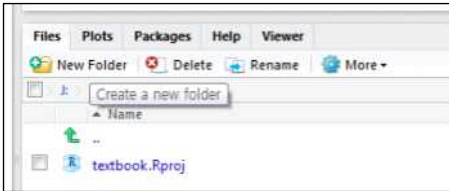
-[Empty Project]창에서 다음과 같이 선택 후 [Create Project]버튼 클릭

- [Create Project at subdirectory of]항목 : [Browse]버튼을 눌러 [project] 폴더 선택
- [Directory name]항목 : textbook



6. R script 사용

-[script]폴더 생성: [File]탭에서 [New Folder]항목 클릭 후 script 폴더 생성



- 생성된 [script]폴더 클릭

- [File]-[New File]-[R script]메뉴 선택

* [R script] : 코드 작성

- [R script]에서 [Console]처럼 내용 입력 후 실행할 경우 : 실행부분을 블록설정 후 Ctrl+ Enter을 눌러 실행 예) 2+3

7. 패키지 추가 설치 및 호출

-설치: `install.packages("패키지명")` 예)한국어 자연어 처리(KoNLP) 설치 : `install.packages("KoNLP")`

-호출(사용) : `library(패키지명)` 또는 `library("패키지명")` 예) 호출: `library(KoNLP)`

-도움말: `help("패키지명")`

8. 내부망에서 패키지 수동설치

-설치 : `install.packages("패키지명", repos = NULL)`

예시) `install.packages("패키지명", repos = NULL)`

* 그밖의 유용한 패키지

참조: 원문 사이트

<https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages>

1. To load data (데이터 로드)

- RODBC, RMySQL, RPostgresSQL, RSQLite - 데이터베이스로부터 데이터를 읽어 들이는데 유용
- XLConnect, xlsx - 엑셀파일을 읽고 쓰는데 유용
- foreign - SAS, SPASS 데이터를 읽어 들이는데 유용
- R에서 일반 텍스트 파일 읽기 - 특별한 패키지가 필요 없으며, `read.csv`, `read.table`, `read.fwf` 함수를 사용한다.

2. To manipulate data (데이터 조작(가공)하기)

- dplyr - 부분화(그룹화,subsetting), 요약(summarizing), 재배열(재조정, rearranging), 취합 데이터 셋 조인(취합

데이터세트 결합, joining together data sets)을 위한 손쉬운 방법 제공. 데이터마이닝에서 데이터 전처리기로 사용

- tidy - 데이터 셋의 레이아웃을 변환시킬 때 사용. 수집(gather)과 분산(spread)함수를 사용해서 데이터를 정돈된 형태로 변환.
- stringr - 정규표현식(regular expressions)과 문자열(character strings) 작업을 위한 툴.
- lubridate - 날짜와 시간데이터를 쉽게 작업할 수 있게 해주는 툴.

3. To visualize data(데이터 시각화하기)

- ggplot2 - 그래픽 툴
- ggvis - 상호동작적인 웹기반의 그래픽 툴
- rgl - 상호동작적인 3D 시각화 툴
- htmlwidgets - 상호동작적인 시각화 툴
- leaflet (maps)- 지도
- dygraphs (time series)- 시계열
- DT (tables)- 표
- diagrammeR (diagrams)- 다이어그램
- network3D (network graphs, 네트워크), 네트워크
- threeJS (3D scatterplots and globes)- 3D 산포도와 구체
- googleVis - 구글 차트 툴 사용

4. To model data(데이터 모델링하기)

- car - 다 변량 분산 분석 툴
- mgcv - 일반 가산 모형(Generalized Additive Models(GAM), 일반화모형에 다른 모형이 가산된 모형)
- lme4/nlme - 선형 및 비선형이 혼합된 효과모형(Linear and Non-linear mixed effects models)
- randomForest - 기계학습으로부터 랜덤 포레스트 방법(Random forest methods). 랜덤 포레스트 방법(Random forest, RF)- 디시전트리(의사결정나무, decision tree)의 확장 개념으로 목표변수 예측에 사용
- multcomp - 다중 비교 테스트(multiple comparison testing) 툴
- vcd - 분류별 데이터(categorical data) 테스트와 시각화 툴
- glmnet - Lasso 와 elastic-net 회귀 방법. 연속적인 변수 선택과 계수 추정을 동시에 활용할 수 있는 특성.
- survival - 생존분석(survival analysis)툴. 시간-사건 분석. 어떤 특정한 사건이 발생할 때까지의 시간을 분석하여 생존함수 또는 생존곡선을 추정하는 방법
- caret - 훈련 회귀(training regression)와 분류 모형(classification models)툴

5. To report results (결과 리포트하기)

- shiny - 상호 동작적 웹 앱을 생성
- R Markdown - 재현할 수 있는 리포팅을 위한 워크플로우. RStudio에 포함되어 있음.
- xtable - xtable함수를 사용해서 latex 또는 HTML코드를 반환. R Markdown과 짝을 이뤄서 사용

6. For Spatial data(공간데이터)

- sp, maptools - 공간데이터 로드와 사용 툴
- maps -작도를 위한 맵 다각형 사용
- ggmap - 구글맵에서 스트리트 맵 다운로드

7. For Time Series and Financial data (시계열과 금융 데이터)

- zoo - 시계열 객체를 구함
- xts - 시계열 데이터 셋을 조작
- quantmod - 금융데이터 다운로드, 차트 작도 및 기술적인 분석 제공.

8. To write high performance R code (고성능 R코드 쓰기)

- Rcpp - 빠른 속도의 C++ 코드 호출하는 R함수.
- data.table - 정리된 데이터를 만들어주는 매우 빠른 명령어를 제공.
- parallel - 병렬처리에 사용

9. To work with the web (웹 작업 - 웹 데이터 읽고 쓰기)

- XML - XML문서를 읽고 생성
- jsonlite - JSON데이터 테이블을 읽고 생성
- httr - http통신 작업에 유용

10. To write your own R packages(자신만의 R패키지 작성)

- devtools - R패키지에 자신의 코드를 넣기 위한 툴.
- testthat - 자신의 코드 프로젝트를 위한 유닛 테스트를 작성
- roxygen2 - 자신의 R패키지를 위한 문서(document) 작성.