

데이터 가져오기

텍스트 데이터 읽어오기

- 텍스트 데이터의 개념
 - 텍스트 데이터는 데이터와 데이터를 구분하기 위해서 공백(blank), 콤마(comma), 탭(tab)과 같은 구분자(separator)를 사용
- 텍스트 데이터를 R에서 읽어오기 위해서는 `read.table()` 함수를 사용

argument	설명
file	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
header	<ul style="list-style-type: none">• 논리형으로 TRUE를 지정하면 외부 데이터에 있는 변수명을 R 데이터에서도 동일하게 사용• FALSE를 지정하면 외부 데이터의 변수명을 사용하지 않고 R에서 자체적으로 변수명을 지정(이름은 “V1”, “V2” 식으로 됨)
sep	구분자로 문자형 형태로 지정
stringsAsFactors	<ul style="list-style-type: none">• 문자형 데이터를 요인(factor)으로 자동으로 변경할 지를 설정• FALSE를 지정하면 문자형 데이터를 요인으로 변경하지 않고 문자형으로 읽어 들임
na.strings	데이터에 결측치가 있거나 결측치로 지정할 내용을 문자형으로 지정

텍스트 데이터 읽어오기

- 구분자가 공백인 경우
 - 파일의 위치가 c드라이브에 있고, 파일이름이 data.txt로 되어 있을 때 R에서 읽어오는 방법
 - 데이터명 = `read.table(file="파일위치/파일명.txt", header=TRUE, sep=" ")`
`data_rblank = read.table(file="c:/data.txt", header=TRUE, sep=" ")`
- 구분자가 콤마인 경우
 - 파일의 위치가 c드라이브에 있고, 파일이름이 data.txt로 되어 있을 때 R에서 읽어오는 방법
 - 데이터명 = `read.table(file="파일위치/파일명.txt", header=TRUE, sep=",")`
`data_comma = read.table(file="c:/data.txt", header=TRUE, sep=",")`

텍스트 데이터 읽어오기

- 구분자가 탭인 경우
 - 데이터명 = `read.table(file="파일위치/파일명.txt", header=TRUE, sep="\t")`
data_tab = `read.table(file="c:/data.txt", header=TRUE, sep="\t")`

CSV 데이터 읽어오기

- read.csv() 함수의 사용
 - 엑셀 데이터의 특별한 형태로 콤마로 구분된 형태의 데이터
 - Comma Separated Value의 약자
 - CSV 데이터는 read.csv() 함수를 사용

argument	설명
file	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
header	논리형으로 CSV 데이터 있는 변수명을 사용하려면 TRUE 를 지정

CSV 데이터 읽어오기

- read.csv() 함수의 사용
- c 드라이브에 인구주택총조사2015.csv 데이터가 있을 때 R에서 읽어오는 방법
 - 데이터명 = read.csv(file="파일위치/파일명.csv", header=TRUE)
 - survey2015 = read.csv(file="c:/인구주택총조사2015.csv", header=TRUE)

엑셀 데이터 읽어오기

- 엑셀 데이터는 R의 기본 기능으로는 읽어오는 함수가 없기 때문에 엑셀 데이터를 읽어오는 기능을 제공하는 새로운 패키지를 설치해야 함
- 엑셀 데이터를 읽어오는 기능을 제공하는 패키지는 여러 개가 있지만 패키지를 위해서 별도의 작업을 해야 하는 불편함이 있음
- 해들리 위컴(Hadley Wickham)이 개발한 `readxl` 패키지에서 제공하는 `read_excel()` 함수를 이용하여 엑셀 데이터를 읽어오는 방법을 학습하도록 함

엑셀 데이터 읽어오기

- readxl 패키지의 설치와 로딩
 - readxl 패키지를 설치하고 로딩하는 방법
 - `install.packages("readxl")`
 - `library(readxl)`

엑셀 데이터 읽어오기

- read_excel() 함수 사용법

argument	설명
path	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
sheet	엑셀 파일의 있는 시트(sheet) 중에서 어떤 시트인지를 지정 <ul style="list-style-type: none">• 시트명을 알고 있으면 문자형으로 지정• 시트의 위치(index)를 알면 수치형으로 지정
col_names	<ul style="list-style-type: none">• 논리형으로 엑셀에 있는 변수명을 사용할 지를 의미• TRUE이면 엑셀에 있는 변수명을 사용

엑셀 데이터 읽어오기

- C드라이브에 big.xls가 있고, Sheet1에 데이터가 있다고 가정하면 다음과 같이 엑셀 데이터를 불러올 수 있음
 - 엑셀 2003 이하 버전인 경우에는 확장자가 xls이고, 2007 이상 버전인 경우에는xlsx임
- c 드라이브에 ebook.xlsx 데이터가 있을 때 R에서 읽어오는 방법
 - 데이터명 = read_excel(path="파일위치/파일명.xls", sheet="시트명", col_names=TRUE)
 - ebook = read_excel(path="c:/ebook.xlsx", sheet="data", col_names=TRUE)