빅데이터 AI - R programming (류정원 강사 - [ryuzy@naver.com](mailto:ryuzy@naver.com))

WIFI: KOREA\_G / korea123g

1. 7.14.

R 프로그래밍 – 데이터 관리/분석 적합 cf) Python – 머신러닝/딥러닝

* OpenSource를 기반으로 사용하므로 version에 민감하다.
* Opensource를 통해 도움을 받을 수 있다.
* 데이터 핸들링이란 Txt, csv, x니, sas, spss, Stata, db 등의 다양한 데이터를 읽어오고 처리가능
* \*데이터 전처리\*
* 메모리 사용: 데이터는 메모리(RAM)에서 작동하기 때문에 데이터 처리가 굉장히 빠르다. 메모리 내에서 CPU를 통해 연산하고 화면상에 결과를 보여줌.
* 메모리: 임시기억장치
* CPU: 계산 및 제어 역할 수행
* R 프로그램 설치

-스타트업 옵션: 사용자 환경 변경 (Yes) - 디스플레이 모드 선택.

* R studio 설치 (rstudio.com)
* 오픈소스 기반으로 R을 사용하는데 편리하고 유용한 기능 제공
* 명령어를 직접 입력하는 방식으로 작동
* 컴퓨터 이름은 한글(특수문자, 띄어쓰기) 로 하지 않고 영어로 지정
* 사용자 이름은 한글(특수문자, 띄어쓰기) 로 하지 않고 영어로 지정
* Option🡪 code🡪 saving 🡪 UTF-8로 변경
* Environment 창: 메모리에 올라간 내용들을 볼 수 있다.
* Plots 창: 도움말, 그래프 등을 보여준다.
* Scripts 창: 많은 양의 프로그램을 작성할 때 쓰는 창.
* R 문법

1. script에서 커맨드 실행 (ctr + Enter)
2. 실행경로를 설정해준다. C:/KyoungPark/R (Tools 🡪 Global Options)
3. 주석 ( # ): 프로그램의 전반적인 설명이나 내용이 무엇을 의미하는지 알 수 있도록 사용자가 설명을 달아주는 기능입니다.
4. 7.16.

* Num{1:5} 🡪 index 1부터 5까지의 data가 들어가있다.,
* Matrix 관련함수
  + cbind(), rbind()

열, 또는 행을 중심으로 백터값을 행렬로 만들어줌

1. 7.21

* 데이터 프레임 형식
* 데이터 프레임 요소 접근
* read.csv()
* header=F 옵션
* read\_excel() 🡪 tibble 형태로 가져옴
* 외부 패키지 이용하는 방법
* 패키지를 다운로드하고 메모리에 로드 시켜줘야한다.

(install.packages(‘ ’), library( ))

* Xlsx 읽기 쓰기를 위해 java JDK 를 설치해줘야한다.

1. 7.23

* 배열과 리스트 -모든 자료 형태 공부 완료
* 파생 변수
* Subset, transform, mutate
* 데이터 처리 및 분석하기
* Filter, select, table-

1. 7.24

* Summarise() ,group\_by()
* Left\_join() ,bind\_rows()
* 데이터 정제
* 결측치 처리하는 방법
* 실습
* 이상한데이터 : outliers
* 이상치 제거하기 : 결측치로 대체하기
* 그래프 그리기 (ggplot2)
* Geom\_point(), geom\_col()

1. 7.28

* 예제 문제 실습
* 한국 복지 패널 데이터 분석
* 데이터 열이름을 쉽게 바꿔준다.
* 주제 설정
  1. 성별에 따른 급여 차이
  2. 나이에 따른 급여 차이
  3. 연령대 따른 급여 차이
  4. 성별과 연령대별 급여 차이
* 데이터 전처리
  1. 성별
  2. 급여
* 표를 만들어준다.

1. 8.6

* 과제 실습
* 정적 웹 스크레이핑(크롤링)
* 네이버 영화 사이트 실습