2. R은 통계 계산을 위한 프로그래밍 언어, 데이터 분석 소프트웨어

AT&T의 벨연구소 S라는 통계언어에서 문법부분 통계처리부분을 참고해서 만듬

1993년 통계학과 교수 "로스 이하카"와 "로버트 젠틀맨" 2명이 개발(통계처리 부분만)

=일반 사람들도 쉽게 통계적 데이터분석을 할 수 있도록 개발.(자바/C/파이썬과 다름)

3. R로 무엇을 할 수 있나요?

통계 분석 : 기초 통계, 가설 검정

머신러닝 모델링 : SVM, Random Forest, Deep Learning

텍스트 마이닝 : 문장속의 단어 빈도를 이용하여 글의 성향을 분석(긍정/부정, 보수/진보 등등)

이미지 분석, 사운드 분석 등등

4. 데이터 분석을 할 수 있는 다른 패키지는 무엇이 있나요? [ 데이터 분석 툴적인 측면에서 비교]

SAS, SPSS : 유료@ 기능이 더 좋아요(특정 목적에 최적화가 되어있음-사회과학/ R은 속도가 느리다)

Excel : 무료, 통계처리 가능, 기능이 빈약함, 데이터 용량 한계가 있다

R : 유료에 버금가는 기능, 무료(학계에서 주로 사용)

5. 데이터 분석을 할 수 있는 다른 프로그래밍 언어는 어떤게 있나요?

Python : 데이터 분석 후 해당 결과를 다른 시스템에 연동하는 목적으로 사용(자율주행)

R : 일반적으로 데이터 분석 후 결과를 어딘가에 reporting/ 차트로 표현하는데 최적

6. R의 장점

오픈 소스 : 무료 사용 가능 > 오픈 소스 생태계 > 계속 새로운 분석 기법이 나옴/공유해줌/사용해봐~

CRAN 사이트에서 여러 패키지를 공유하여 최신 분석기법들을 손쉽게 이용할 수 있어요.

프로그래밍 방식의 데이터 분석 : 재현성 확보 가능

7. R 다운로드

CRAN 사이트(Comprehensive R Archive Network)

CRAN에서 platform에 맞는 파일을 다운로드 후 설치

=컴퓨터/노트북에 윈도우 아이디를 영어로 만들어서 설치해야함

IDE(개발툴 integrated development environment)이 필요함 : Rstudio를 다운로드하여 설치

8. R studio 실행하기

- Source 창 : 스크립트 파일을 생성해서 코드를 작성함

- Console 창 : 코드의 실행결과가 출력됨(ctrl 누른 채로 enter)

- Environment 창(환경 창) : 코드 실행 시 사용된 변수의 값을 알려줌

- File 창 : 현재 사용하는 폴더 (wroking directory )를 확인할 수 있음

ex ) plot(차트) package(추가기능목록)

9. R프로젝트 생성

R 프로젝트 생성 시 한글이 들어 있는 경우 오류 발생 가능

프로젝트 명과 폴더 명에 한글이 들어가지 않도록 처리하기

프로젝트 생성 후 약간의 설정을 해보자(한글 인코딩 설정/ 줄바꿈 설정)

- 줄바꿈 설정 (tools > global option > code > editing > soft-wrap R source files)

- 유니코드 인코딩 설정 (tools > global option > code > saving > default text encoding)-모두 유니코드로 저장(utf8)

(tools > project option > code editing > utf8)