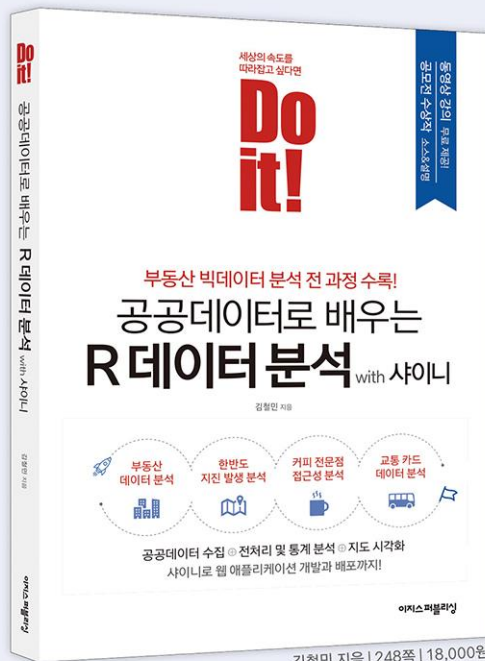


R로 공공데이터를 분석하는 전 과정 실습!
공모전 수상작으로 배우는 R 데이터 분석



김철민 지음 | 248쪽 | 18,000원

공모전
수상작
소스 & 설명

동영상
강의
무료 제공

R을 사용하는 데이터 분석 세계에 오신 것을 환영합니다. 첫 장에서는 경쟁력 있는 데이터 분석가로 성장하려면 무엇이 필요한지 알아보고, 컴퓨터에 R과 R스튜디오를 설치해 실습 환경을 준비해 보겠습니다.

01

데이터 분석가로 성장하기

01-1 경쟁력 있는 데이터 분석가가 되는 법

01-2 데이터 분석 환경 만들기

01-3 R 스크립트 실행해 보기

01-4 프로젝트 수행 과정 훑어보기

01-1 경쟁력 있는 데이터 분석가 되기

모든 것은 연결된다

‘커넥팅 더 닷^{Connecting the dots}’은 스티브 잡스의 스탠퍼드 대학교 졸업 연설문에 등장하는 유명한 문구입니다. 점이 모여서 선이 되듯이 과거 사건들이 조각조각 연결되어 현재를 구성한다는 의미입니다. 오늘날 우리의 모습은 과거 다양한 사건이 모여서 연결된 결과물인 거죠.

데이터 분석 세계에서든 마찬가지입니다. 이 책의 독자라면 아마도 기본적인 데이터 처리와 분석 방법을 경험해 보았을 것입니다. 지금부터는 이러한 경험과 기술을 하나로 모아서 새로운 무언가를 만들어 낼 때입니다. 이제 여러분은 데이터 분석의 입문자를 벗어나 중급자로 향하는 길에 들어섰습니다.

01-1 경쟁력 있는 데이터 분석가 되기

중급자의 함정

스키에서는 ‘중급자의 함정’이라는 말이 있습니다. 많은 돈과 시간을 들여서 기본 자세를 배우고 기술을 익혔지만 이후 동기가 약해지면서 어느 순간부터 흥미를 잃게 된다는 말입니다.

데이터 분석 세계에서도 비슷한 일이 생깁니다. 많은 사람이 처음 데이터 분석을 공부할 때는 신기하고 재미있어서 열심히 책을 보고 인터넷을 뒤져 가며 배웁니다. 그러나 어느 순간부터 데이터 분석에 투자한 노력과 시간이 무슨 의미가 있는지 모르겠다는 생각이 들면서 흥미를 잃어 갑니다.

이러한 슬럼프를 극복하려면 지금까지 배운 내용을 ‘도대체 어디에 사용할 수 있을까?’와 같은 질문을 스스로 던져 보아야 합니다.

01-1 경쟁력 있는 데이터 분석가 되기

데이터 분석 서비스를 개발해 보자

많은 시간과 노력을 들여서 배운 데이터 분석 능력을 어디에 쓸 수 있을까요? 학생이라면 지금까지 배운 지식을 살려 빅데이터 경진 대회나 해커톤 준비에 사용할 수 있습니다. 또 기업이나 연구 기관에 근무하는 직장인이라면 해당 분야의 데이터를 분석하여 보고서를 작성할 때 도움을 받을 수 있습니다.



그러나 모든 데이터 분석의 최종 목적은 분석 애플리케이션 개발이라고 생각합니다. 자료 수집, 전처리 그리고 서비스에 이르기까지 모든 분야를 알아야 하는 데이터 분석의 종합예술이자 끝판왕이기 때문입니다.

이제 여러분은 데이터 분석의 새로운 영역에 들어섰습니다. 자신의 아이디어와 다양한 분석 기법을 연결하여 지금까지 세상에 없던 새로운 분석 서비스를 창조해 보세요.

01-1 경쟁력 있는 데이터 분석가 되기

이 책에서 무엇을 얻을 수 있을까?

이 책은 데이터 분석에서 서비스 개발에 이르기까지 모든 과정을 다룸으로써 경쟁력 있는 데이터 분석가로 성장할 수 있도록 도와줍니다. R을 이용하여 공공데이터를 수집하고 분석한 다음 시각화하는 방법뿐만 아니라, 샤이니^{Shiny} 패키지를 이용하여 실제로 사용할 수 있는 애플리케이션을 개발하고 배포하는 과정까지 함께 살펴봅니다.



즉, 단편적인 분석 기술을 익히는 데 머무르지 않고 하나의 프로젝트를 완결하는 데 초점을 맞추었습니다. 분석 대상은 요즘 우리 사회에서 뜨거운 감자라고 할 수 있는 ‘서울특별시 부동산 데이터’입니다. R과 샤이니는 한 몸이므로 R만 어느 정도 알고 있다면 다른 프로그래밍을 따로 공부할 필요는 없습니다. 이 책을 덮는 순간 차별화된 데이터 분석 경쟁력을 갖출 수 있길 소원합니다.

01-1 경쟁력 있는 데이터 분석가 되기

데이터 분석 프로젝트는 일반적으로 다음처럼 다섯 단계로 구분됩니다. 데이터 분석 실습을 본격적으로 진행하기 전에 이 책의 프로젝트가 어떠한 순서로 진행되는지 살펴보겠습니다.

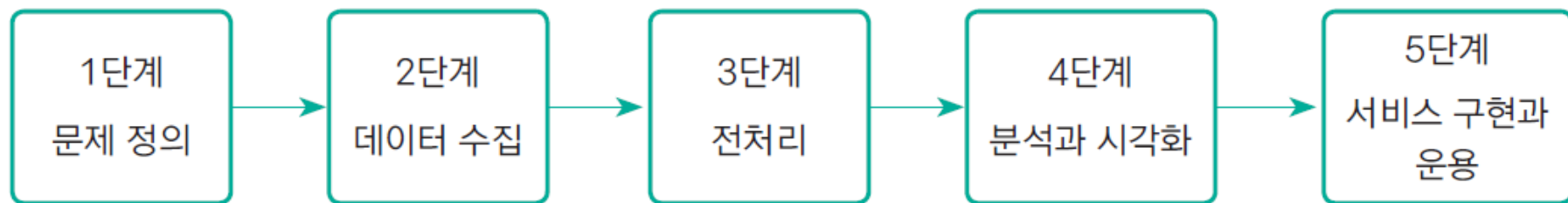


그림 1-11 데이터 분석 프로젝트 5단계

1단계 R 설치하기

R은 R 프로젝트의 공식 사이트(cran.r-project.org)에서 내려받을 수 있습니다. 공식 사이트 첫 페이지에서 Download R for Windows*를 클릭합니다.

* 맥OS 사용자라면 'Download R for macOS'를 클릭하세요.

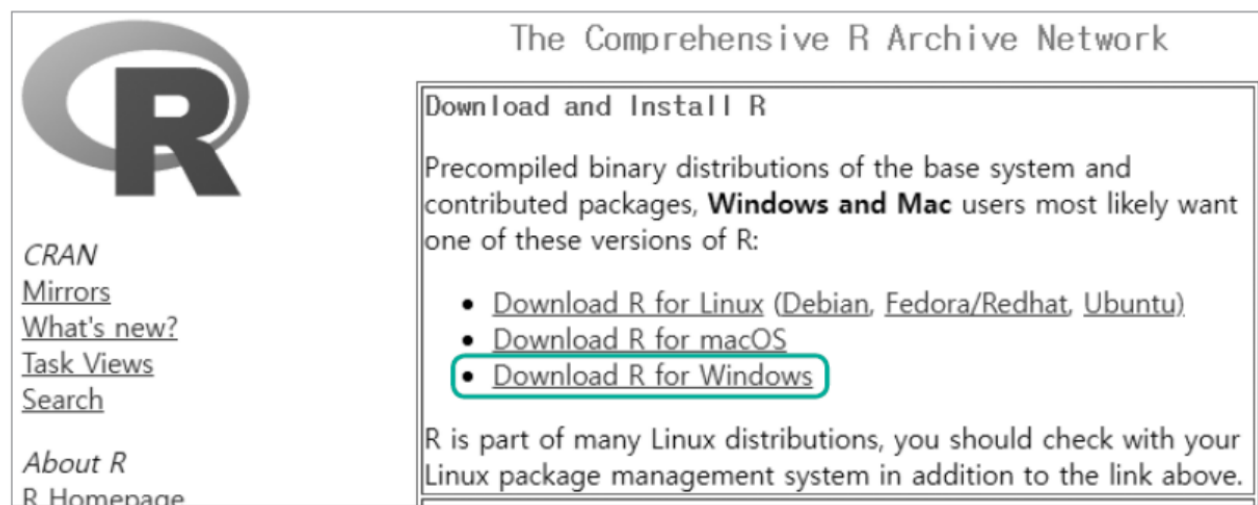


그림 1-1 R 공식 사이트

2단계 R스튜디오 설치하기

R스튜디오는 R을 사용하기 편리하게 만들어 주는 통합 개발 환경입니다. R스튜디오를 사용하면 다양한 부가 기능을 활용해 데이터를 효율적으로 분석할 수 있습니다. R스튜디오 공식 사이트에 접속한 후 내려받기 페이지(rstudio.com/products/rstudio/download)로 이동하면 다음과 같은 요금제를 확인할 수 있습니다. 우리는 학습이 목적이므로 무료인 RStudio Desktop 버전 아래 <DOWNLOAD>를 클릭합니다.

| RStudio Desktop | RStudio Desktop Pro | RStudio Server | RStudio Workbench® |
|----------------------------|----------------------------|----------------------------|---|
| Open Source License | Commercial License | Open Source License | Commercial License |
| Free | \$995 | Free | \$4,975 |
| | /year | | /year (5 Named Users) |
| DOWNLOAD | BUY | DOWNLOAD | BUY |
| Learn more | Learn more | Learn more | Evaluation Learn more |

그림 1-5 R스튜디오 요금제

01-2 데이터 분석 환경 만들기



R스튜디오를 실행할 때 오류가 발생해요

윈도우에서 R스튜디오를 실행할 때 크게 2가지 문제로 오류가 발생할 수 있습니다. 이럴 때는 다음처럼 설정을 바꾸어 보세요.

- 관리자 권한으로 실행되지 않을 때

R스튜디오를 관리자 권한으로 실행하지 않으면 오류가 발생할 수 있습니다. 이때는 다음처럼 설정을 변경하고 R스튜디오를 다시 실행해 보세요.

- 1 R스튜디오 아이콘을 마우스 오른쪽으로 클릭한 후 [속성 → 호환성]을 선택합니다.
- 2 '관리자 권한으로 이 프로그램 실행'에 체크한 후 <확인>을 누릅니다.

01-2 데이터 분석 환경 만들기



R스튜디오를 실행할 때 오류가 발생해요

윈도우에서 R스튜디오를 실행할 때 크게 2가지 문제로 오류가 발생할 수 있습니다. 이럴 때는 다음처럼 설정을 바꾸어 보세요.

• 윈도우 사용자 계정이 한글일 때

윈도우 사용자 계정이 한글이면 오류가 발생할 수 있습니다. 이때는 영문으로 사용자 계정을 새로 만들어 로그인한 후 R스튜디오를 다시 실행해 보세요.

- 1 작업 표시줄의 검색란에서 '설정'을 검색해 설정 앱을 열고 [계정 → 가족 및 다른 사용자 → 이 PC에 다른 사용자 추가]를 선택합니다.
- 2 '이 사람의 로그인 정보를 가지고 있지 않습니다.'를 클릭한 후 약관에 동의합니다. 'Microsoft 계정 없이 사용자 추가'를 클릭합니다.
- 3 '사용자 이름'에 영문으로 이름을 입력한 후 <다음>을 클릭합니다.
- 4 [시작 → 전원 → 종료 또는 로그아웃 → 로그아웃]을 클릭해 로그아웃합니다.
- 5 윈도우 시작 화면에서 화면 왼쪽 아래에 있는 새로 만든 영문 계정을 클릭해 로그인합니다.

01-3 R 스크립트 실행해 보기

지금까지 R과 R스튜디오 설치 방법을 살펴보았습니다. 본격적으로 실습에 들어가기 전에 몇 가지 사용법을 알아보겠습니다. 먼저 앞 절에서 설치한 R스튜디오를 실행하면 다음과 같은 화면을 볼 수 있습니다.

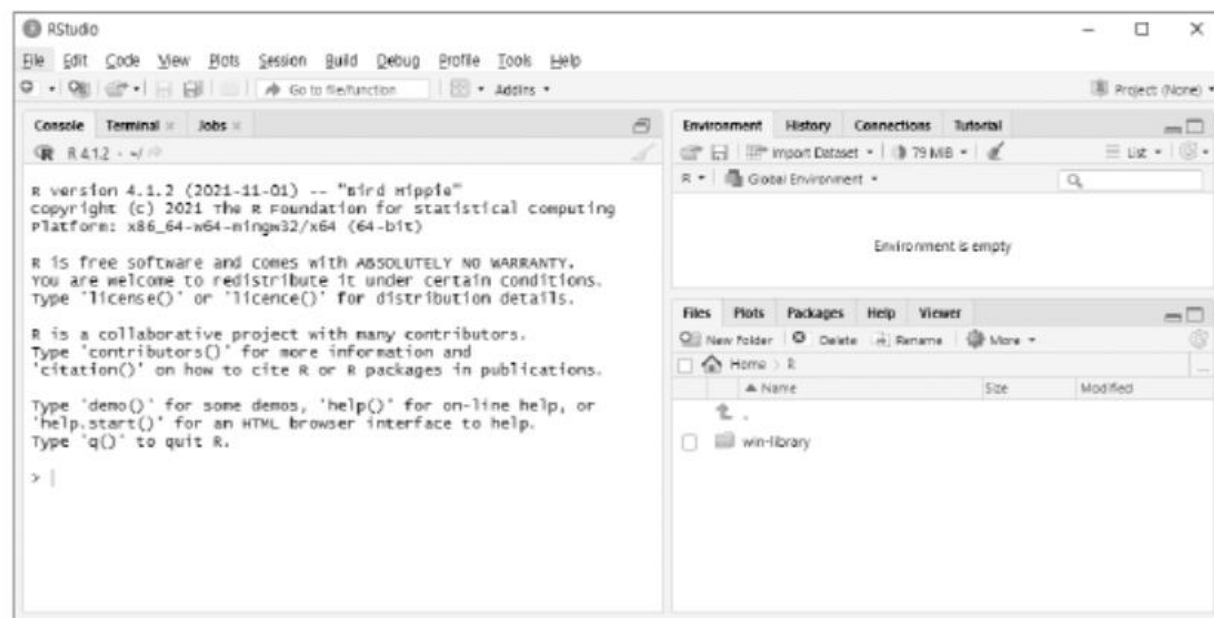


그림 1-8 R스튜디오 초기 화면

01-3 R 스크립트 실행해 보기



R 스크립트에서 한글이 깨져요

스크립트 내용 중 한글이 깨져서 나온다면 대부분 인코딩 설정 문제입니다. 이때는 R 스튜디오에서 [Tools → Global Options]를 선택하여 설정 창을 엽니다. 왼쪽 메뉴에서 [Code]를 클릭하고 위쪽 탭에서 [Saving]을 클릭하면 다음 화면이 보입니다.

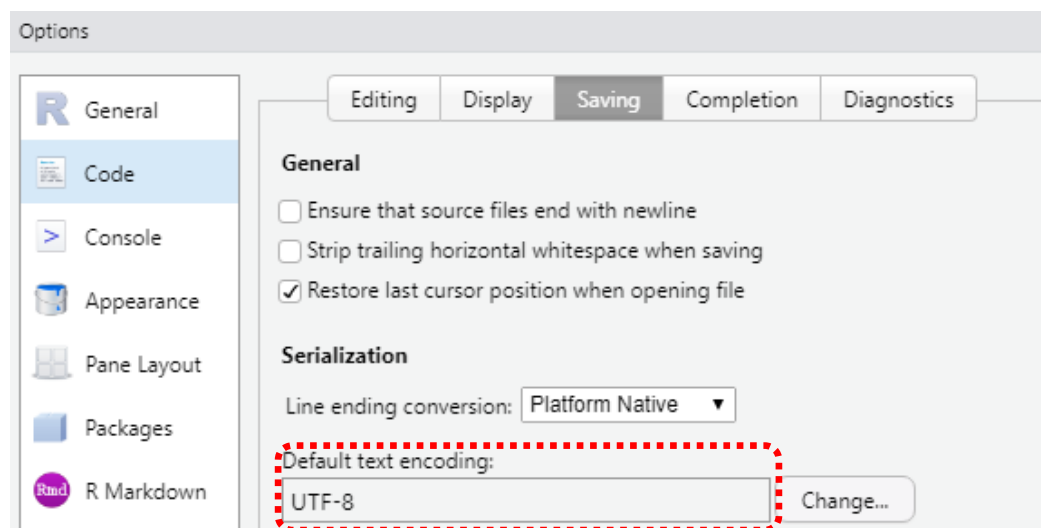


그림 1-10 R스튜디오 인코딩 설정

여기서 'Default text encoding' 항목을 UTF-8로 바꾼 후 <OK>를 클릭하여 R스튜디오를 다시 시작합니다. 이처럼 인코딩을 UTF-8로 바꾸면 한글로 작성한 스크립트도 정상으로 출력됩니다.

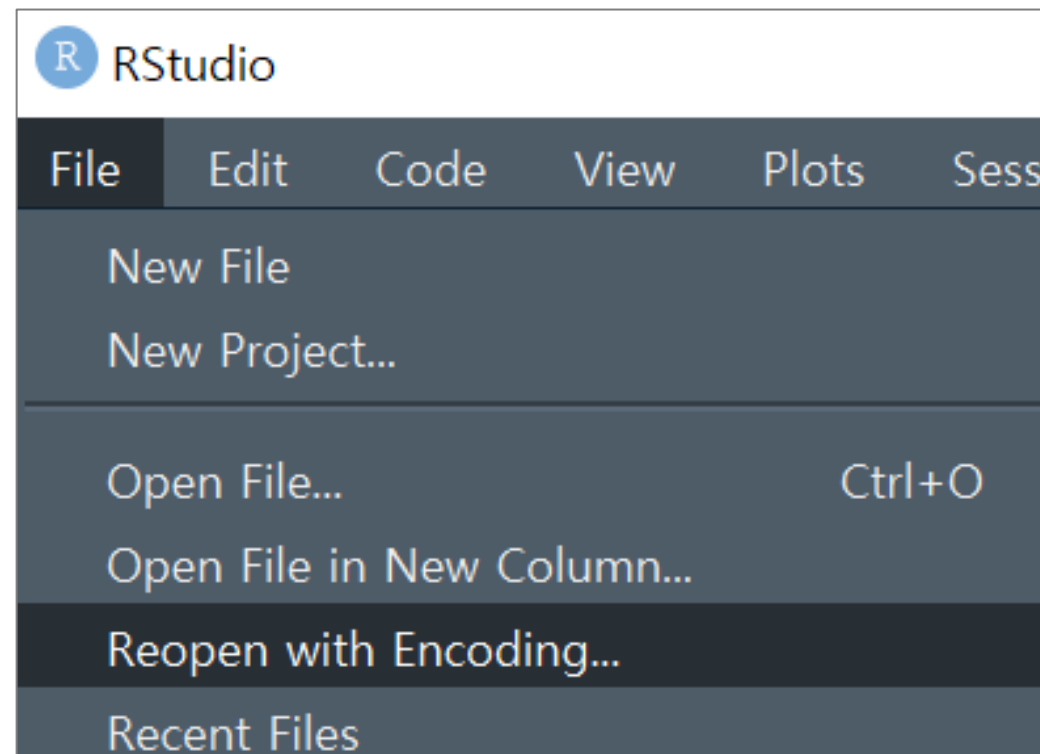
01-3 R 스크립트 실행해 보기



R 스크립트에서 한글이 깨져요

File → Reopen with Encoding

```
1  
2 #-----#  
3 # 03-1 ?궐濡ㅂ 쫑 以.鋤?: 兕덥뽳?쑤 以  
4 #-----#  
5  
6 #---# [1?뽳怨?: ?엑?뽳?룽?똥 ?좃?절?  
7  
8 # install.packages("rstudioapi") #  
9 setwd(dirname(rstudioapi::getSourceE  
10 getwd() # ?쑤?쑤  
11
```



01-4 프로젝트 수행 과정 훑어보기

데이터 분석 프로젝트는 일반적으로 다음처럼 다섯 단계로 구분됩니다. 데이터 분석 실습을 본격적으로 진행하기 전에 이 책의 프로젝트가 어떠한 순서로 진행되는지 살펴보겠습니다.

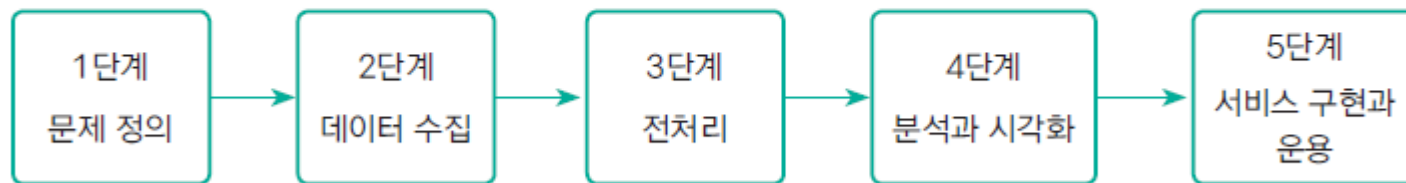


그림 1-11 데이터 분석 프로젝트 5단계