
R & HIVE 데이터 사용 매뉴얼

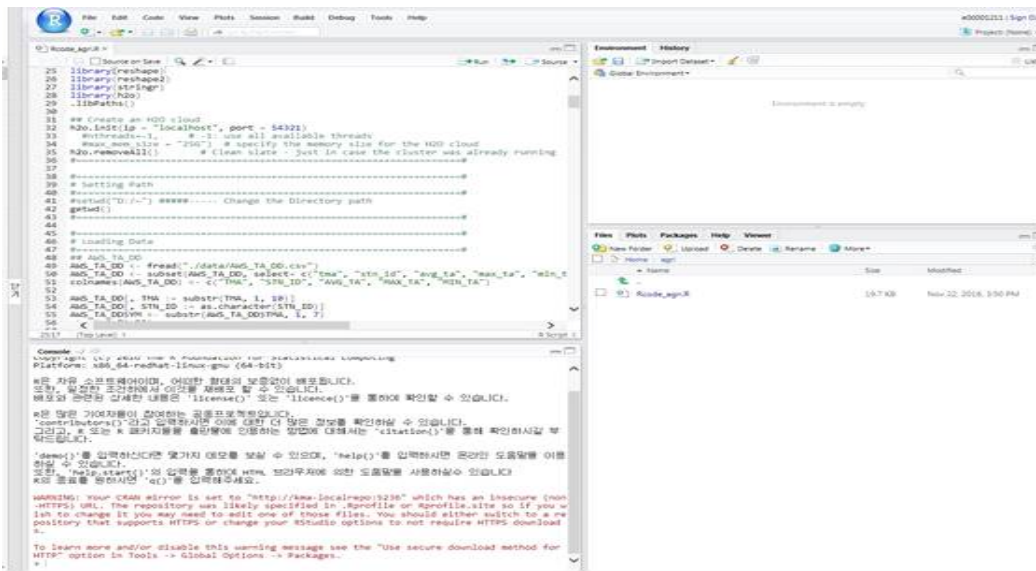
[2019 날씨 빅데이터 콘테스트]

0. 날씨마루(<https://bd.kma.go.kr>)에 접속



0-1. 날씨마루 내 분석환경 ▶ Rstudio 이용하기

- 분석환경 이용을 위한 계정신청이 필요합니다.



1. 개요

날씨 빅데이터 콘테스트 제공 데이터를 분석환경(R)에서 HIVE(HADOOP용 데이터웨어하우스 시스템)에 접속하여 사용 할 수 있는 방법을 설명 합니다.

1.1 HIVE에 저장되어 있는 유통데이터와 소셜데이터를 안내 합니다.

데이터정보	테이블명
기상데이터	bigcon_weather

<기상데이터>

데이터정보	테이블명
GS25 편의점	korea_cvs
랄라블라	korea_hnb

<유통데이터>

데이터정보	테이블명
맛집	social_br
데이트	social_date
건강	social_health
취미	social_hobby
미세먼지	social_pm

<소셜언급량 데이터>

2. 데이터 확인

2.1 날씨마루 분석환경(RStudio)에 접속하게 되면 HIVE와 자동으로 연결이 되며, conn변수와 dbGetQuery() 함수를 이용하여 데이터를 조회 및 연산을 수행 할 수 있습니다.

```
1 # 테이블 목록 확인
2 dbGetQuery(conn, "show tables")
```



```
> dbGetQuery(conn, "show tables")
      tab_name
1      aws_hr_hm
2      aws_hr_rn
3      aws_hr_ta
4      aws_hr_wd
5      db_aws_cloud_dd
6      db_aws_cloud_mnh
7      db_aws_cloud_tdom
8      db_aws_dd_old
9      db_aws_hm_tim_old
10     db_aws_icsr_ss_dd
11     db_aws_icsr_ss_mnh
12     db_aws_icsr_ss_tdom
13     db_aws_lwt_tg_dd
14     db_aws_nmyr_dd
15     db_aws_nmyr_mnh
16     db_aws_nmyr_sesn
17     db_aws_prsr_dd
18     db_aws_rhm_dd
19     db_aws_rhm_mnh
20     db_aws_rhm_tdom
21     db_aws_rn_dd
22     db_aws_rn_mnh
23     db_aws_rn_tdom
24     db_aws_rn_tim
25     db_aws_ta_dd
26     db_aws_ta_mnh
27     db_aws_ta_tdom
28     db_aws_te_dd
29     db_aws_ts_dd
```

예시)

```
database <- dbGetQuery(conn, "show tables")
```

3. 분석환경 데이터 로딩

- 데이터가 HIVE에 저장되어 있어 SQL을 이용한 분석이 바로 가능하지만, 분석환경으로 로딩이 필요한 경우 변수를 생성하여 저장 할 수 있습니다.

```
8 # 데이터 불러오기
9 korea_cvs <- dbGetQuery(conn, "select * from korea_cvs")
10 korea_hnb <- dbGetQuery(conn, "select * from korea_hnb")
```

예시)

```
test_select <- dbGetQuery(conn, "select * from aws_hr_hm limit 50")
```

4. 데이터 확인

- 분석환경에 로딩한 데이터가 정상적으로 불러왔는지 확인하기 위해 head, tail, summary 함수를 이용합니다.

```

11
12 # 데이터의 상위(head)/하위(tail) 확인
13 head(korea_cvs)
14 tail(korea_cvs)
15 head(korea_hnb)
16 tail(korea_hnb)
17
18 # 데이터 요약 정보
19 summary(korea_cvs)
20 summary(korea_hnb)
21

```

```

> head(korea_cvs)
bigcon01.pvn_nm bigcon01.sale_dt bigcon01.gen_cd bigcon01.age_cd bigcon01.category
1 "서울특별시"      20160101      "F"      "00~19"      "라면"
2 "서울특별시"      20160101      "F"      "20~39"      "과자"
3 "서울특별시"      20160101      "F"      "20~39"      "라면"
4 "서울특별시"      20160101      "F"      "20~39"      "마스크"
5 "서울특별시"      20160101      "F"      "20~39"      "맥주"
6 "서울특별시"      20160101      "F"      "20~39"      "생리대"
bigcon01.adj_qty bigcon01.bor_nm
1 7 "종로구"
2 655 "종로구"
3 282 "종로구"
4 35 "종로구"
5 183 "종로구"
6 78 "종로구"
> summary(korea_cvs)
bigcon01.pvn_nm bigcon01.sale_dt bigcon01.gen_cd bigcon01.age_cd
Length:2707786 Length:2707786 Length:2707786 Length:2707786
Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character

bigcon01.category bigcon01.adj_qty bigcon01.bor_nm
Length:2707786 Min. : 7.00 Length:2707786
Class :character 1st Qu.: 14.00 Class :character
Mode :character Median : 35.00 Mode :character
Mean : 86.66
3rd Qu.: 92.00
Max. :49938.00

```

예시)

head(korea_cvs)

tail(korea_cvs)

summary(korea_cvs)

5. 데이터 저장

- 분석결과 및 중간결과 확인을 위해 파일형태로 저장하여 반출을 할 수 있으며, write.csv 함수를 이용하여 분석의 용이한 CSV파일로 저장할 수 있습니다.
- 사용자 개개인의 환경이 다르므로 write.csv 함수 사용시, fileEncoding 파라미터를 추가하여 데이터를 저장하여야 합니다.

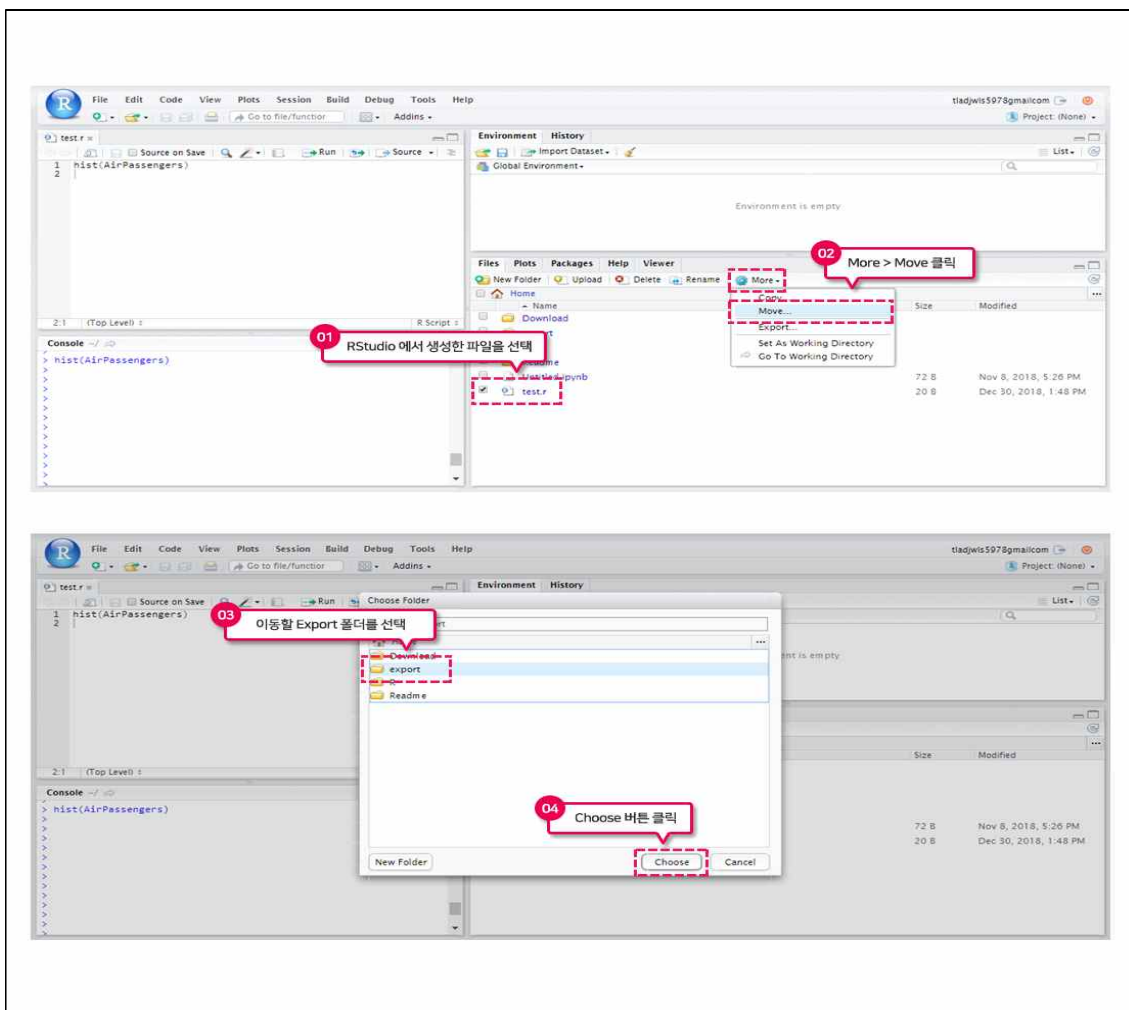
```

47
48 # CSV파일로 저장
49 write.csv(korea_cvs, "korea_cvs.csv", row.names = FALSE, fileEncoding = "cp949")
50 write.csv(korea_hnb, "korea_hnb.csv", row.names = FALSE, fileEncoding = "euc-kr")
51 write.csv(korea_hnb, "korea_hnb.csv", row.names = FALSE, fileEncoding = "utf-8")
52

```

5.1 분석결과 데이터 다운로드

- 분석결과 데이터를 다운로드하기 위해서 파일을 이동하여 분석결과 데이터 다운로드를 신청합니다.



기상청날씨마루
기상기후 빅데이터 분석 플랫폼

소개

융합서비스

분석환경

게시판

콘텐츠

01 상단 메뉴바에서 분석환경 선택

02 분석결과 다운로드

분석환경

기상기후 빅데이터를 다른 분야 데이터와
접목시켜 분석할 수 있는 환경을
제공합니다.

바로가기

분석교육실습

R을 활용한 분석 | R 교육 동영상 | Python을 활용한 분석 | Fortran을 활용한 분석

비정형 도구

데이터 시각화

빅데이터 분석도구

R studio | Python | Fortran

데이터

기상데이터 | 업로드 데이터 | 웹데이터

마이페이지

나의 이용 현황 | 1:1 상담 | 비밀번호 재설정

분석결과 다운로드

분석결과 다운로드

■ 분석결과 다운로드 신청 순서

- 01 step** 분석결과파일을 R-Studio에서 분석서버 홈 디렉터리 아래의 export 폴더에 이동시킵니다.
- 02 step** [분석결과 다운로드 신청] 화면에서 다운로드 할 파일을 선택하고, 다운로드 신청 사유를 선택한 후 확인 버튼을 클릭합니다.
- 03 step** 관리자에 의해 분석결과 파일 검토 후 승인이 완료되면 분석결과파일을 [분석결과 다운로드 현황] 화면에서 다운로드 받습니다.

■ 유의사항

- 다운로드 대상은 분석결과에 한하여 가능합니다.
- 다운로드 기간은 승인 완료일부터 일주일까지이며 해당 기간 이후에는 자동 삭제됩니다.

다운로드 가능 데이터	<input checked="" type="checkbox"/> test.r	01 다운로드 할 분류 체크
분석결과 다운로드 사유	선택	02 다운로드 사유 선택

03 분석결과 다운로드 신청 클릭

분석결과 다운로드 신청