R FOR SQL USERS

Pre-Requisites

- Understanding of basic R programming
- Understanding SQL Scripting

SQL-Operations

DATA VIEWING

dataset<-read.csv(file.choose())

VIEWING DATA IN A TABLE

Converts data to tbl class. tbl's are easier to examine than data frames.

library(dplyr)

tbl_df(*dataset*)

VIEWING DATA IN A SPREADSHEET

View data set in spreadsheet-like display

View(*dataset*)

SUMMARISE DATA

COUNT

Count number of rows with each unique value of variable

library(dplyr)

count(*dataset, column*)

DISTINCT VALUES IN A COLUMN

number of distinct values in a column vector

library(dplyr)

n_distinct(*dataset,column*)

GROUP DATA

GROUP DATA

Group data into rows with the same value of column specified

library(dplyr)

group_by(*dataset,column*)

SUBSET DATA

FILTER DATA

Extract rows that meet logical criteria.

library(dplyr)

filter(dataset, condition)

DISTINCT/UNIQUE VALUES

Remove duplicate rows.

library(dplyr)

distinct(dataset)

RESHAPING DATA

MAKING NEW COLUMNS

Compute and append one or more new columns.

library(dplyr)

mutate(dataset, columns condition)

COMBINING DATASET

LEFT JOIN

Join matching rows from dataset2 to dataset1

library(dplyr)

left_join(dataset1,dataset2,by="condition")

RIGHT JOIN

Join matching rows from dataset1 to dataset2

library(dplyr)

right_join(dataset1,dataset2,by="condition")

INNER JOIN

Join data. Retain only rows in both sets.

library(dplyr)

inner_join(dataset1,dataset2,by="condition")

FULL JOIN

Join data. Retain all values, all rows

library(dplyr)

full_join(dataset1,dataset2,by="condition")

DATA DIFFERENCE AND COMPARISON

INTERSECT

Rows that appear in both dataset1 and dataset2

intersect(dataset1,dataset2)

UNION

Rows that appear in either or both dataset1 and dataset2

union(dataset1,dataset2)

DATA COMPARISON

Rows that appear in dataset1 but not dataset2

setdiff(dataset1,dataset2)