Reading MySQL

Jeong Jihye

2016년 10월 29일

library(RMySQL)

## Loading required package: DBI

## Connecting and listing databases

ucscDb<-dbConnect(MySQL(),user="genome", host="genome-mysql.cse.ucsc.edu")   
####host : MySQL server가 어디 있는지를 나타냄  
  
result<-dbGetQuery(ucscDb,"show databases;");   
#### "show databases;"는 R명령어가 아닌 MySQL에서 사용하는 명령어임.  
#### MySQL server를 사용한 뒤에는 dbDisconnect 하는 것이 필요함.  
  
result

## Database  
## 1 information\_schema  
## 2 ailMel1  
## 3 allMis1  
## 4 anoCar1  
## 5 anoCar2  
## 6 anoGam1  
## 7 apiMel1  
## 8 apiMel2  
## 9 aplCal1  
## 10 aptMan1  
## 11 balAcu1  
## 12 bosTau2  
## 13 bosTau3  
## 14 bosTau4  
## 15 bosTau5  
## 16 bosTau6  
## 17 bosTau7  
## 18 bosTau8  
## 19 bosTauMd3  
## 20 braFlo1  
## 21 caeJap1  
## 22 caePb1  
## 23 caePb2  
## 24 caeRem2  
## 25 caeRem3  
## 26 calJac1  
## 27 calJac3  
## 28 calMil1  
## 29 canFam1  
## 30 canFam2  
## 31 canFam3  
## 32 cavPor3  
## 33 cb1  
## 34 cb3  
## 35 ce10  
## 36 ce11  
## 37 ce2  
## 38 ce4  
## 39 ce6  
## 40 cerSim1  
## 41 chlSab2  
## 42 choHof1  
## 43 chrPic1  
## 44 ci1  
## 45 ci2  
## 46 criGri1  
## 47 danRer1  
## 48 danRer10  
## 49 danRer2  
## 50 danRer3  
## 51 danRer4  
## 52 danRer5  
## 53 danRer6  
## 54 danRer7  
## 55 dasNov3  
## 56 dipOrd1  
## 57 dm1  
## 58 dm2  
## 59 dm3  
## 60 dm6  
## 61 dp2  
## 62 dp3  
## 63 droAna1  
## 64 droAna2  
## 65 droEre1  
## 66 droGri1  
## 67 droMoj1  
## 68 droMoj2  
## 69 droPer1  
## 70 droSec1  
## 71 droSim1  
## 72 droVir1  
## 73 droVir2  
## 74 droYak1  
## 75 droYak2  
## 76 eboVir3  
## 77 echTel1  
## 78 echTel2  
## 79 equCab1  
## 80 equCab2  
## 81 eriEur1  
## 82 eriEur2  
## 83 felCat3  
## 84 felCat4  
## 85 felCat5  
## 86 felCat8  
## 87 fr1  
## 88 fr2  
## 89 fr3  
## 90 gadMor1  
## 91 galGal2  
## 92 galGal3  
## 93 galGal4  
## 94 galGal5  
## 95 galVar1  
## 96 gasAcu1  
## 97 gbMeta  
## 98 geoFor1  
## 99 go  
## 100 go080130  
## 101 go140213  
## 102 go150121  
## 103 gorGor3  
## 104 gorGor4  
## 105 gorGor5  
## 106 hetGla1  
## 107 hetGla2  
## 108 hg16  
## 109 hg17  
## 110 hg18  
## 111 hg19  
## 112 hg19Patch10  
## 113 hg19Patch2  
## 114 hg19Patch5  
## 115 hg19Patch9  
## 116 hg38  
## 117 hg38Patch2  
## 118 hg38Patch3  
## 119 hg38Patch6  
## 120 hg38Patch7  
## 121 hgFixed  
## 122 hgTemp  
## 123 hgcentral  
## 124 latCha1  
## 125 loxAfr3  
## 126 macEug1  
## 127 macEug2  
## 128 macFas5  
## 129 melGal1  
## 130 melUnd1  
## 131 micMur1  
## 132 micMur2  
## 133 mm10  
## 134 mm10Patch1  
## 135 mm10Patch4  
## 136 mm5  
## 137 mm6  
## 138 mm7  
## 139 mm8  
## 140 mm9  
## 141 monDom1  
## 142 monDom4  
## 143 monDom5  
## 144 musFur1  
## 145 myoLuc2  
## 146 nomLeu1  
## 147 nomLeu2  
## 148 nomLeu3  
## 149 ochPri2  
## 150 ochPri3  
## 151 oreNil1  
## 152 oreNil2  
## 153 ornAna1  
## 154 ornAna2  
## 155 oryCun2  
## 156 oryLat2  
## 157 otoGar3  
## 158 oviAri1  
## 159 oviAri3  
## 160 panPan1  
## 161 panTro1  
## 162 panTro2  
## 163 panTro3  
## 164 panTro4  
## 165 panTro5  
## 166 papAnu2  
## 167 papHam1  
## 168 performance\_schema  
## 169 petMar1  
## 170 petMar2  
## 171 ponAbe2  
## 172 priPac1  
## 173 proCap1  
## 174 proteins120806  
## 175 proteins121210  
## 176 proteins140122  
## 177 proteins150225  
## 178 proteins160229  
## 179 proteome  
## 180 pteVam1  
## 181 rheMac1  
## 182 rheMac2  
## 183 rheMac3  
## 184 rheMac8  
## 185 rn3  
## 186 rn4  
## 187 rn5  
## 188 rn6  
## 189 sacCer1  
## 190 sacCer2  
## 191 sacCer3  
## 192 saiBol1  
## 193 sarHar1  
## 194 sorAra1  
## 195 sorAra2  
## 196 sp120323  
## 197 sp121210  
## 198 sp140122  
## 199 sp150225  
## 200 sp160229  
## 201 speTri2  
## 202 strPur1  
## 203 strPur2  
## 204 susScr2  
## 205 susScr3  
## 206 taeGut1  
## 207 taeGut2  
## 208 tarSyr1  
## 209 tarSyr2  
## 210 test  
## 211 tetNig1  
## 212 tetNig2  
## 213 triMan1  
## 214 tupBel1  
## 215 turTru2  
## 216 uniProt  
## 217 vicPac1  
## 218 vicPac2  
## 219 visiGene  
## 220 xenTro1  
## 221 xenTro2  
## 222 xenTro3  
## 223 xenTro7

## Connecting to hg19 and listing tables

hg19<-dbConnect(MySQL(),user="genome", db="hg19", host="genome-mysql.cse.ucsc.edu")   
#### 여러 database 중 "hg19"만 연결  
  
allTables<-dbListTables(hg19)   
#### 한 datase(ex.hg19)안에 여러개의 data table 존재하므로 목록확인  
  
length(allTables)

## [1] 11047

allTables[1:5]

## [1] "HInv" "HInvGeneMrna" "acembly" "acemblyClass"  
## [5] "acemblyPep"

## Gent dimensions of a specific table

dbListFields(hg19,"affyU133Plus2")

## [1] "bin" "matches" "misMatches" "repMatches" "nCount"   
## [6] "qNumInsert" "qBaseInsert" "tNumInsert" "tBaseInsert" "strand"   
## [11] "qName" "qSize" "qStart" "qEnd" "tName"   
## [16] "tSize" "tStart" "tEnd" "blockCount" "blockSizes"   
## [21] "qStarts" "tStarts"

#### field는 dataframe의 colume개념. "affyU133Plus2"의 colume name 목록을 확인.   
  
dbGetQuery(hg19,"select count(\*) from affyU133Plus2")

## count(\*)  
## 1 58463

#### (table "affyU133Plus2"의 row(\*)개수를 세라)

# Read from the table

affyData<-dbReadTable(hg19,"affyU133Plus2")  
head(affyData)

## bin matches misMatches repMatches nCount qNumInsert qBaseInsert  
## 1 585 530 4 0 23 3 41  
## 2 585 3355 17 0 109 9 67  
## 3 585 4156 14 0 83 16 18  
## 4 585 4667 9 0 68 21 42  
## 5 585 5180 14 0 167 10 38  
## 6 585 468 5 0 14 0 0  
## tNumInsert tBaseInsert strand qName qSize qStart qEnd tName  
## 1 3 898 - 225995\_x\_at 637 5 603 chr1  
## 2 9 11621 - 225035\_x\_at 3635 0 3548 chr1  
## 3 2 93 - 226340\_x\_at 4318 3 4274 chr1  
## 4 3 5743 - 1557034\_s\_at 4834 48 4834 chr1  
## 5 1 29 - 231811\_at 5399 0 5399 chr1  
## 6 0 0 - 236841\_at 487 0 487 chr1  
## tSize tStart tEnd blockCount  
## 1 249250621 14361 15816 5  
## 2 249250621 14381 29483 17  
## 3 249250621 14399 18745 18  
## 4 249250621 14406 24893 23  
## 5 249250621 19688 25078 11  
## 6 249250621 27542 28029 1  
## blockSizes  
## 1 93,144,229,70,21,  
## 2 73,375,71,165,303,360,198,661,201,1,260,250,74,73,98,155,163,  
## 3 690,10,32,33,376,4,5,15,5,11,7,41,277,859,141,51,443,1253,  
## 4 99,352,286,24,49,14,6,5,8,149,14,44,98,12,10,355,837,59,8,1500,133,624,58,  
## 5 131,26,1300,6,4,11,4,7,358,3359,155,  
## 6 487,  
## qStarts  
## 1 34,132,278,541,611,  
## 2 87,165,540,647,818,1123,1484,1682,2343,2545,2546,2808,3058,3133,3206,3317,3472,  
## 3 44,735,746,779,813,1190,1195,1201,1217,1223,1235,1243,1285,1564,2423,2565,2617,3062,  
## 4 0,99,452,739,764,814,829,836,842,851,1001,1016,1061,1160,1173,1184,1540,2381,2441,2450,3951,4103,4728,  
## 5 0,132,159,1460,1467,1472,1484,1489,1497,1856,5244,  
## 6 0,  
## tStarts  
## 1 14361,14454,14599,14968,15795,  
## 2 14381,14454,14969,15075,15240,15543,15903,16104,16853,17054,17232,17492,17914,17988,18267,24736,29320,  
## 3 14399,15089,15099,15131,15164,15540,15544,15549,15564,15569,15580,15587,15628,15906,16857,16998,17049,17492,  
## 4 14406,20227,20579,20865,20889,20938,20952,20958,20963,20971,21120,21134,21178,21276,21288,21298,21653,22492,22551,22559,24059,24211,24835,  
## 5 19688,19819,19845,21145,21151,21155,21166,21170,21177,21535,24923,  
## 6 27542,

# Select a specific subset

query<-dbSendQuery(hg19,"select\*from affyU133Plus2 where misMatches between 1 and 3")   
#### (table "affyU133Plus2"에서 misMatches라는 valiable(1~3의 값을 갖는)의 모든 observations(\*)을 선택하라)  
#### 선택된 항목이 database에 저장  
  
affyMis<-fetch(query); quantile(affyMis$misMatches)

## 0% 25% 50% 75% 100%   
## 1 1 2 2 3

affyMisSmall<-fetch(query,n=10); dbClearResult(query)

## [1] TRUE

####선택된 data 중 10개 열만 추출  
  
dim(affyMisSmall)

## [1] 10 22

dbDisconnect(hg19)

## [1] TRUE

#### 연결을 끝내는 것을 잊지 않도록 주의 !!

#### \*추가적인 commands확인하고자 하면 #### <http://www.pantz.org/software/mysql/mysqlcommands.html> #### <http://www.r-bloggers.com/mysql-and-r/>