Week 2

Tomás

18 de noviembre de 2019

# SQL

library(DBI)

## Warning: package 'DBI' was built under R version 3.6.1

library(RMySQL)  
ucscDB <- dbConnect(MySQL(),user = 'genome', host = 'genome-mysql.cse.ucsc.edu')  
result <- dbGetQuery(ucscDB,'show databases'); dbDisconnect(ucscDB)

## [1] TRUE

result ## Entrega una lista de todas las bases de datos disponibles en el host

## Database  
## 1 acaChl1  
## 2 ailMel1  
## 3 allMis1  
## 4 allSin1  
## 5 amaVit1  
## 6 anaPla1  
## 7 ancCey1  
## 8 angJap1  
## 9 anoCar1  
## 10 anoCar2  
## 11 anoGam1  
## 12 anoGam3  
## 13 apaSpi1  
## 14 apaVit1  
## 15 apiMel1  
## 16 apiMel2  
## 17 aplCal1  
## 18 aptFor1  
## 19 aptMan1  
## 20 aquChr2  
## 21 araMac1  
## 22 ascSuu1  
## 23 balAcu1  
## 24 balPav1  
## 25 bisBis1  
## 26 bosTau2  
## 27 bosTau3  
## 28 bosTau4  
## 29 bosTau5  
## 30 bosTau6  
## 31 bosTau7  
## 32 bosTau8  
## 33 bosTau9  
## 34 bosTauMd3  
## 35 braFlo1  
## 36 bruMal2  
## 37 bucRhi1  
## 38 burXyl1  
## 39 caeAng2  
## 40 caeJap1  
## 41 caeJap4  
## 42 caePb1  
## 43 caePb2  
## 44 caePb3  
## 45 caeRem2  
## 46 caeRem3  
## 47 caeRem4  
## 48 caeSp111  
## 49 caeSp51  
## 50 calAnn1  
## 51 calJac1  
## 52 calJac3  
## 53 calMil1  
## 54 canFam1  
## 55 canFam2  
## 56 canFam3  
## 57 capCar1  
## 58 carCri1  
## 59 cavPor3  
## 60 cb1  
## 61 cb3  
## 62 cb4  
## 63 ce10  
## 64 ce11  
## 65 ce2  
## 66 ce4  
## 67 ce6  
## 68 cerSim1  
## 69 chaVoc2  
## 70 cheMyd1  
## 71 chlSab2  
## 72 chlUnd1  
## 73 choHof1  
## 74 chrPic1  
## 75 chrPic2  
## 76 ci1  
## 77 ci2  
## 78 ci3  
## 79 colLiv1  
## 80 colStr1  
## 81 corBra1  
## 82 corCor1  
## 83 cotJap2  
## 84 criGri1  
## 85 criGriChoV1  
## 86 criGriChoV2  
## 87 cucCan1  
## 88 danRer1  
## 89 danRer10  
## 90 danRer11  
## 91 danRer2  
## 92 danRer3  
## 93 danRer4  
## 94 danRer5  
## 95 danRer6  
## 96 danRer7  
## 97 dasNov3  
## 98 dipOrd1  
## 99 dirImm1  
## 100 dm1  
## 101 dm2  
## 102 dm3  
## 103 dm6  
## 104 dp2  
## 105 dp3  
## 106 droAna1  
## 107 droAna2  
## 108 droEre1  
## 109 droGri1  
## 110 droMoj1  
## 111 droMoj2  
## 112 droPer1  
## 113 droSec1  
## 114 droSim1  
## 115 droSim2  
## 116 droVir1  
## 117 droVir2  
## 118 droYak1  
## 119 droYak2  
## 120 eboVir3  
## 121 echTel1  
## 122 echTel2  
## 123 egrGar1  
## 124 equCab1  
## 125 equCab2  
## 126 equCab3  
## 127 eriEur1  
## 128 eriEur2  
## 129 eurHel1  
## 130 falChe1  
## 131 falPer1  
## 132 felCat3  
## 133 felCat4  
## 134 felCat5  
## 135 felCat8  
## 136 felCat9  
## 137 ficAlb2  
## 138 fr1  
## 139 fr2  
## 140 fr3  
## 141 fulGla1  
## 142 gadMor1  
## 143 galGal2  
## 144 galGal3  
## 145 galGal4  
## 146 galGal5  
## 147 galGal6  
## 148 galVar1  
## 149 gasAcu1  
## 150 gavSte1  
## 151 gbMeta  
## 152 geoFor1  
## 153 go  
## 154 go080130  
## 155 go140213  
## 156 go150121  
## 157 go180426  
## 158 gorGor3  
## 159 gorGor4  
## 160 gorGor5  
## 161 haeCon2  
## 162 halAlb1  
## 163 halLeu1  
## 164 hetBac1  
## 165 hetGla1  
## 166 hetGla2  
## 167 hg16  
## 168 hg17  
## 169 hg18  
## 170 hg19  
## 171 hg19Patch10  
## 172 hg19Patch13  
## 173 hg38  
## 174 hg38Patch11  
## 175 hgFixed  
## 176 hgcentral  
## 177 information\_schema  
## 178 latCha1  
## 179 lepDis1  
## 180 letCam1  
## 181 loaLoa1  
## 182 loxAfr3  
## 183 macEug1  
## 184 macEug2  
## 185 macFas5  
## 186 manPen1  
## 187 melGal1  
## 188 melGal5  
## 189 melHap1  
## 190 melInc2  
## 191 melUnd1  
## 192 merNub1  
## 193 mesUni1  
## 194 micMur1  
## 195 micMur2  
## 196 mm10  
## 197 mm10Patch4  
## 198 mm5  
## 199 mm6  
## 200 mm7  
## 201 mm8  
## 202 mm9  
## 203 monDom1  
## 204 monDom4  
## 205 monDom5  
## 206 musFur1  
## 207 myoLuc2  
## 208 nanPar1  
## 209 nasLar1  
## 210 necAme1  
## 211 nipNip1  
## 212 nomLeu1  
## 213 nomLeu2  
## 214 nomLeu3  
## 215 ochPri2  
## 216 ochPri3  
## 217 oncVol1  
## 218 opiHoa1  
## 219 oreNil1  
## 220 oreNil2  
## 221 oreNil3  
## 222 ornAna1  
## 223 ornAna2  
## 224 oryCun2  
## 225 oryLat2  
## 226 otoGar3  
## 227 oviAri1  
## 228 oviAri3  
## 229 oviAri4  
## 230 panPan1  
## 231 panPan2  
## 232 panRed1  
## 233 panTro1  
## 234 panTro2  
## 235 panTro3  
## 236 panTro4  
## 237 panTro5  
## 238 panTro6  
## 239 papAnu2  
## 240 papAnu4  
## 241 papHam1  
## 242 pelCri1  
## 243 pelSin1  
## 244 performance\_schema  
## 245 petMar1  
## 246 petMar2  
## 247 petMar3  
## 248 phaCar1  
## 249 phaLep1  
## 250 phoRub1  
## 251 picPub1  
## 252 ponAbe2  
## 253 ponAbe3  
## 254 priExs1  
## 255 priPac1  
## 256 priPac3  
## 257 proCap1  
## 258 proteins120806  
## 259 proteins121210  
## 260 proteins140122  
## 261 proteins150225  
## 262 proteins160229  
## 263 proteins180404  
## 264 proteome  
## 265 pteGut1  
## 266 pteVam1  
## 267 pygAde1  
## 268 pytBiv1  
## 269 rheMac1  
## 270 rheMac10  
## 271 rheMac2  
## 272 rheMac3  
## 273 rheMac8  
## 274 rhiRox1  
## 275 rn3  
## 276 rn4  
## 277 rn5  
## 278 rn6  
## 279 sacCer1  
## 280 sacCer2  
## 281 sacCer3  
## 282 saiBol1  
## 283 sarHar1  
## 284 serCan1  
## 285 sorAra1  
## 286 sorAra2  
## 287 sp120323  
## 288 sp121210  
## 289 sp140122  
## 290 sp150225  
## 291 sp160229  
## 292 sp180404  
## 293 speTri2  
## 294 strCam1  
## 295 strPur1  
## 296 strPur2  
## 297 strRat2  
## 298 susScr11  
## 299 susScr2  
## 300 susScr3  
## 301 taeGut1  
## 302 taeGut2  
## 303 tarSyr1  
## 304 tarSyr2  
## 305 tauEry1  
## 306 tetNig1  
## 307 tetNig2  
## 308 thaSir1  
## 309 tinGut2  
## 310 triMan1  
## 311 triSpi1  
## 312 triSui1  
## 313 tupBel1  
## 314 turTru2  
## 315 tytAlb1  
## 316 uniProt  
## 317 vicPac1  
## 318 vicPac2  
## 319 visiGene  
## 320 xenLae2  
## 321 xenTro1  
## 322 xenTro2  
## 323 xenTro3  
## 324 xenTro7  
## 325 xenTro9  
## 326 zonAlb1

hg19 <- dbConnect(MySQL(),user = 'genome', host = 'genome-mysql.cse.ucsc.edu', db = 'hg19')  
allTables <- dbListTables(hg19)  
length(allTables)

## [1] 12425

allTables[1:5] # Ejemplo especifico de una database que tiene mas de 10.000 data frames dentro

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

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"

dbGetQuery(hg19,'select count(\*) from affyU133Plus2') ## Cuenta todas las filas que existen en la tabla affy....

## count(\*)  
## 1 58463

affydata <- dbReadTable(hg19, 'affyU133Plus2') # Retorna la dataframe affy... con esto puedo trabajarla

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 0 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 1 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 2 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 3 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 4 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 5 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 6 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 7 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 8 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 11  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 12  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 13  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 15  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 16  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 17  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 18  
## imported as numeric

class(affydata)

## [1] "data.frame"

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,

query <- dbSendQuery(hg19, 'select \* from affyU133Plus2 where misMatches between 1 and 3')

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 0 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 1 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 2 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 3 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 4 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 5 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 6 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 7 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 8 imported  
## as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 11  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 12  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 13  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 15  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 16  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 17  
## imported as numeric

## Warning in .local(conn, statement, ...): Unsigned INTEGER in col 18  
## imported as numeric

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

dbDisconnect(hg19)

## [1] TRUE

# Read from web