**Getting and Cleaning Data – Class Notes**

**Week 2: MySQL**

* “RMySQL” package allows pulling data from MySQL databases
* Once the package is loaded, you can connect to the database and send queries directly to the SQL databases
* Example 1:
  + ucscDb <- dbConnect(MySQL(), user = “genome”, host = “genome-mysql.cse.ucsc.edu”)
    - this opens a connection with the SQL database
  + result <- dbGetQuery(ucscDb, “show databases;”); dbDisconnect(uscsDb);
    - here we establish the connection with the SQL database and then we send a SQL query command – notice SQL syntax of the request. The TRUE response will come from the Disconnect request.
* Example 2:
  + > hg19 <- dbConnect(MySQL(), user ="genome", db="hg19", host ="genome-mysql.cse.ucsc.edu")
  + > allTables <- dbListTables(hg19)
  + > length(allTables)
  + Assume we want to access field within a specific database
  + dbListFields(hg19,"affyU133Plus2")
  + Count records in the table
  + dbGetQuery(hg19, "select count(\*) from affyU133Plus2")
* You can also send query to the database:
  + query <- dbSendQuery(hg19, "select \* from affyU133Plus2 where misMatches between 1 and 3")
  + affyMis <- fetch(query); quantile(affyMis$misMatches)
* With this package, you can GetQuery right away or you can SendQuery and then you can extract just the subset from the dataset.
* Remember to close the connection as soon as you don’t’ need it any more!