607 Meetup #2
Feb 11, 2016

**CUNY School of
Professional Studies**

M.S. in Data Analytics

CU
NY

Tonight:

R and SQL

- Where should the work be done?
- How should I connect my (R) client to the database (native drivers or ODBC or ORM)
- Demo: sample RMySQL beach head

Review your assignment 2 solutions!

Where should the data processing work be done?
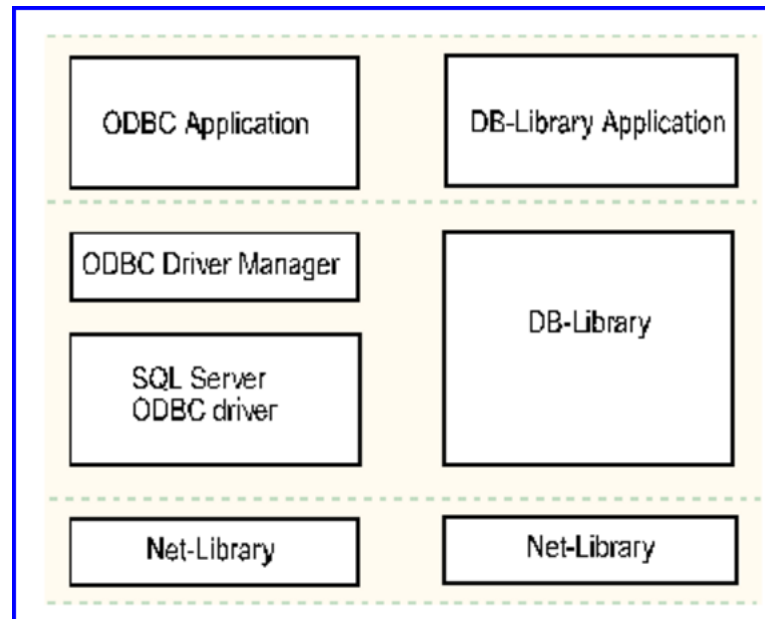
Can it / should it be done all in SQL?

Can it / should it be done all in R?

If you are going to use both SQL and R… how should you connect your SQL to your R?

- File handoff
- Programmatic connection

Often, the data scientist will not have direct access to the SQL data for security reasons.

Native Drivers vs. ODBC

RMySQL Beachhead

```
library(RMySQL)

drv <- dbDriver("MySQL")

con <- dbConnect(drv, user="root", password =
"Kasparov", dbname = "flights",

                 host = "localhost")

dbListTables(con)

dbListFields(con, "weather")

query <- "SELECT year, month, MAX(wind_speed)
    AS maxwind FROM weather GROUP BY year, month;"

max_wind <- dbGetQuery(con, query)
```

Your Assignment 2 solutions