1 Data Base Drivers

The database driver package of your choice

```
install.packages("RPostgreSQL")
install.packages("RMySQL")
install.packages("RMongo")
install.packages("RODBC")
install.packages("RSQLite")
```

Everyone does it when they first start (myself included). You've just written an awesome query in your preferred SQL editor. Everything is perfect - the column names are all snake case, the dates have the right datatype, you finally debugged the "must appear in the GROUP BY clause or be used in an aggregate function" issue. You're ready to do some analysis in R, so you run the query in your SQL editor, copy the results to a csv (or...God forbid... .xlsx) and read into R. You don't have to do this !

`R` has great drivers for nearly every conceivable database. On the off chance you're using a database which doesn't have a standalone driver (SQL Server), you can always use RODBC.

```
library(RPostgreSQL)

drv <- dbDriver("PostgreSQL")
db <- dbConnect(drv, dbname="ncaa",
                user="YOUR USER NAME", password="YOUR PASSWORD")


q <- "SELECT
        *
      FROM
        game_scores;"

data <- dbGetQuery(db, q)
```

```
head(data)
#id          school  game_date spread school_score    opponent opp_score was_home
#1 45111 Boston College 1985-11-16    6.0          21   Syracuse        41     False
#2 45112 Boston College 1985-11-02   13.5          12 Penn State        16     False
#3 45113 Boston College 1985-10-26  -11.0          17 Cincinnati        24     False
#4 45114 Boston College 1985-10-12   -2.0          14       Army        45     False
#5 45115 Boston College 1985-09-28    5.0          10      Miami        45      True
#6 45116 Boston College 1985-09-21    6.5          29 Pittsburgh        22     False
nrow(data)
#[1] 30932
ncol(data)
#[1] 8
```

Next time you've got that perfect query written, just paste it into R and execute it using RPostgreSQL, RMySQL, RMongo, SQLite, or RODBC. In addition to preventing you from having tens of hundreds of CSV files sitting arround, running the query in R saves you time both in I/O but also in converting datatypes. Dates, times, and datetimes will be automatically set to their R equivalent. It also makes your R script reproducible, so you or someone else on your team can easily produce the same results.