IMPORTING DATA IN R

SQL Queries from inside R

Entire table

dbReadTable()

employees

id	name	started_at
1	Tom	2009-05-17
4	Frank	2012-07-06
6	Julie	2013-01-01
7	Heather	2014-11-23
9	John	2014-11-23

Fraction of data

?

employees

name	started_at
Julie	2013-01-01
John	2014-11-23

Selection

Selective importing

- SQL Queries
- DBI -> RMySQL, RPostgreSQL, ...
- Just the basics of SQL

company

employees

id	name	started_at
1	Tom	2009-05-17
4	Frank	2012-07-06
6	Julie	2013-01-01
7	Heather	2014-11-23
9	John	2014-11-23

products

id	name	contract
1	Easy Call	0
2	Call Plus	1
5	Small Biz	0
9	Biz Unlimited	1

sales

id	employee_id	product_id	date	price
1	4	5	2015-09-05	99
2	7	2	2015-09-14	75
3	6	9	2015-09-18	152
4	9	2	2015-09-21	66
5	9	5	2015-09-21	70
7	1	5	2015-09-22	41
8	6	1	2015-09-24	86
9	9	9	2015-09-27	209

company

Names of employees that started after 2012-09-01?

employees

id	name	started_at
1	Tom	2009-05-17
4	Frank	2012-07-06
6	Julie	2013-01-01
7	Heather	2014-11-23
9	John	2014-11-23

sales

id	employee_id	product_id	date	price
1	4	5	2015-09-05	99
2	7	2	2015-09-14	75
3	6	9	2015-09-18	152
4	9	2	2015-09-21	66
5	9	5	2015-09-21	70
7	1	5	2015-09-22	41
8	6	1	2015-09-24	86
9	9	9	2015-09-27	209

products

id	name	contract
1	Easy Call	0
2	Call Plus	1
5	Small Biz	0
9	Biz Unlimited	1

Load package and connect

```
> library(DBI)
> con <- dbConnect(RMySQL::MySQL(),
           dbname = "company",
           host = "courses.csrrinzqubik.us-
                              east-1.rds.amazonaws.com",
           port = 3306,
           user = "student",
           password = "datacamp")
```

Example 1

```
> employees <- dbReadTable(con, "employees")

> subset(employees,
         subset = started_at > "2012-09-01",
         select = name)
     name
3   Julie
4 Heather
5    John

> dbGetQuery(con, "SELECT name FROM employees
                   WHERE started_at > \"2012-09-01\"")
     name
1   Julie
2 Heather
3    John
```

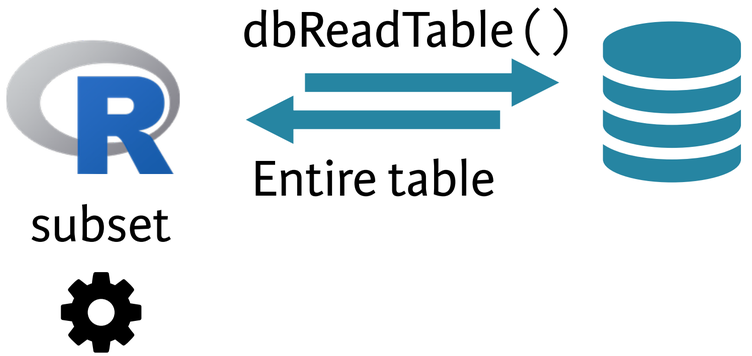
Example 1

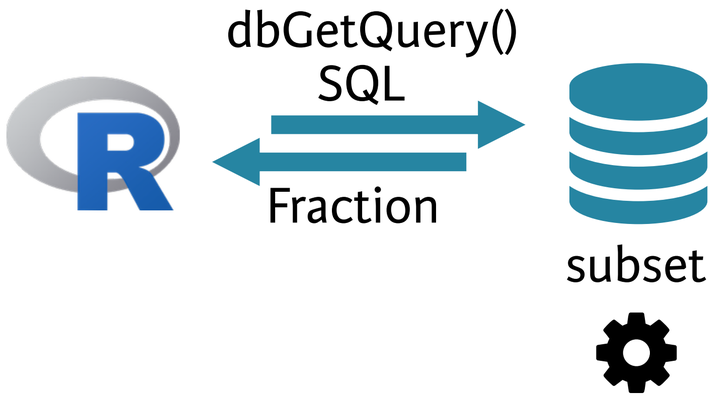
```
> employees <- dbReadTable(con, "employees")

> subset(employees,
         subset = started_at > "2012-09-01",
         select = name)
     name
3   Julie
4 Heather
5    John

> dbGetQuery(con, "SELECT name FROM employees
                      WHERE started_at > \"2012-09-01\"")
     name
1   Julie
2 Heather
3    John
```

Way more efficient for big databases!





company

employees

id	name	started_at
1	Tom	2009-05-17
4	Frank	2012-07-06
6	Julie	2013-01-01
7	Heather	2014-11-23
9	John	2014-11-23

products

id	name	contract
1	Easy Call	0
2	Call Plus	1
5	Small Biz	0
9	Biz Unlimited	1

sales

id	employee_id	product_id	date	price
1	4	5	2015-09-05	99
2	7	2	2015-09-14	75
3	6	9	2015-09-18	152
4	9	2	2015-09-21	66
5	9	5	2015-09-21	70
7	1	5	2015-09-22	41
8	6	1	2015-09-24	86
9	9	9	2015-09-27	209

company

employees

id	name	started_at
1	Tom	2009-05-17
4	Frank	2012-07-06
6	Julie	2013-01-01
7	Heather	2014-11-23
9	John	2014-11-23

products

id	name	contract
1	Easy Call	0
2	Call Plus	1
5	Small Biz	0
9	Biz Unlimited	1

All variables of products with contract

sales

id	employee_id	product_id	date	price
1	4	5	2015-09-05	99
2	7	2	2015-09-14	75
3	6	9	2015-09-18	152
4	9	2	2015-09-21	66
5	9	5	2015-09-21	70
7	1	5	2015-09-22	41
8	6	1	2015-09-24	86
9	9	9	2015-09-27	209

Example 2

```
> products <- dbReadTable(con, "products")


> subset(products, subset = contract == 1)
  id           name contract
2  2     Call Plus        1
4  9 Biz Unlimited        1


> dbGetQuery(con, "SELECT * FROM products
                     WHERE contract = 1")
  id           name contract
1  2     Call Plus        1
2  9 Biz Unlimited        1
```

keep all columns

single equals sign

Example 2

```
> dbGetQuery(con, "SELECT * FROM products
                                    WHERE contract = 1")
  id           name contract
1  2      Call Plus        1
2  9 Biz Unlimited        1
```

dbGetQuery()

```
> dbGetQuery(con, "SELECT * FROM products
                                  WHERE contract = 1")
  id           name contract
1  2     Call Plus        1
2  9 Biz Unlimited        1


> res <- dbSendQuery(con, "SELECT * FROM products
                                  WHERE contract = 1")

> dbFetch(res)
  id           name contract
1  2     Call Plus        1
2  9 Biz Unlimited        1

> dbClearResult(res)
[1] TRUE
```

dbFetch() one by one

```
> res <- dbSendQuery(con, "SELECT * FROM products
                           WHERE contract = 1")

> while(!dbHasCompleted(res)) {
+   chunk <- dbFetch(res, n = 1)
+   print(chunk)
+ }

  id      name contract
1  2 Call Plus        1
  id          name contract
1  9 Biz Unlimited        1
[1] id       name     contract
<0 rows> (or 0-length row.names)

> dbClearResult(res)
[1] TRUE
```

Disconnect

```
> dbDisconnect(con)
[1] TRUE
```

IMPORTING DATA IN R

Let’s practice!

IMPORTING DATA IN R

DBI internals

dbGetQuery()

```
> dbGetQuery(con, "SELECT * FROM products
                           WHERE contract = 1")
  id           name contract
1  2     Call Plus        1
2  9 Biz Unlimited        1


> res <- dbSendQuery(con, "SELECT * FROM products
                           WHERE contract = 1")

> dbFetch(res)
  id           name contract
1  2     Call Plus        1
2  9 Biz Unlimited        1

> dbClearResult(res)
[1] TRUE
```

dbFetch() one by one

```
> res <- dbSendQuery(con, "SELECT * FROM products
                           WHERE contract = 1")

> while(!dbHasCompleted(res)) {
+   chunk <- dbFetch(res, n = 1)
+   print(chunk)
+ }

  id      name contract
1  2 Call Plus        1
  id          name contract
1  9 Biz Unlimited        1
[1] id       name     contract
<0 rows> (or 0-length row.names)

> dbClearResult(res)
[1] TRUE
```

Disconnect

```
> dbDisconnect(con)
[1] TRUE
```

IMPORTING DATA IN R

Let’s practice!