Connect to a database

INTERMEDIATE IMPORTING DATA IN R

Filip Schouwenaars
Instructor, DataCamp

datacamp

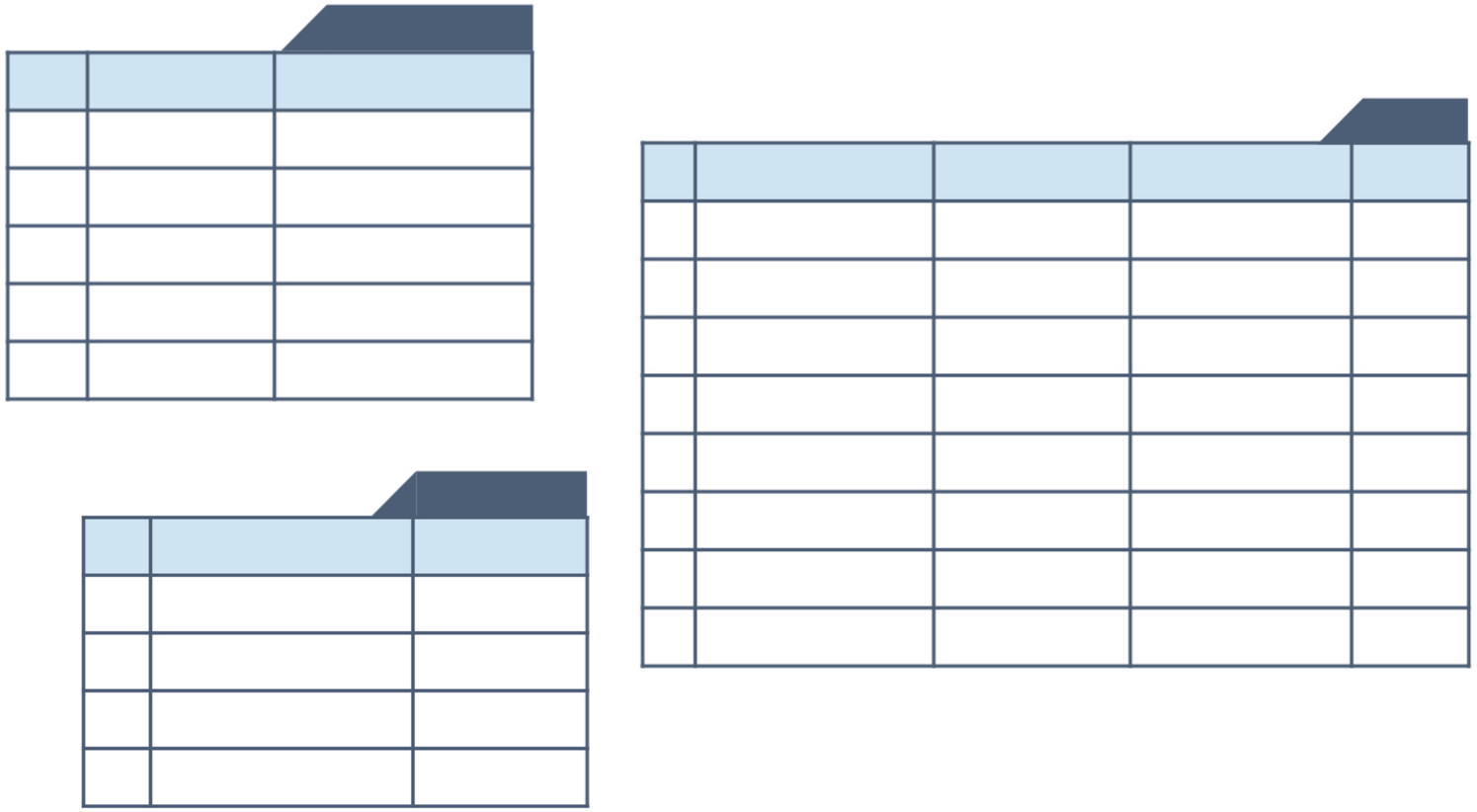
Up to now

- Flat files
- Excel files

Relational Databases

- What is a relational database?
- How to connect?
- How to read table?

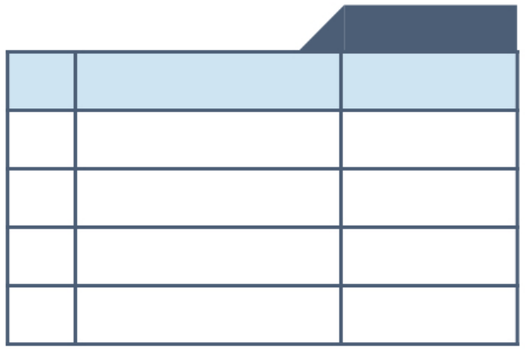
company

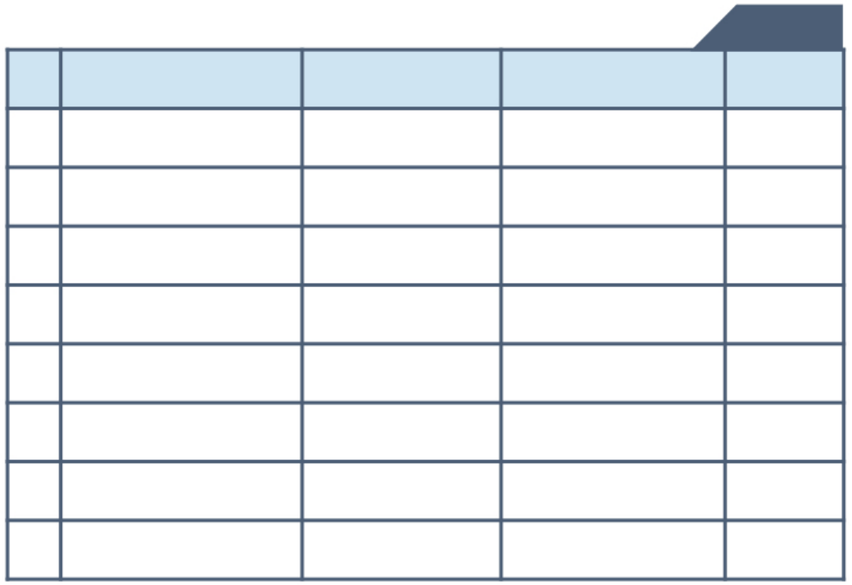


company

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





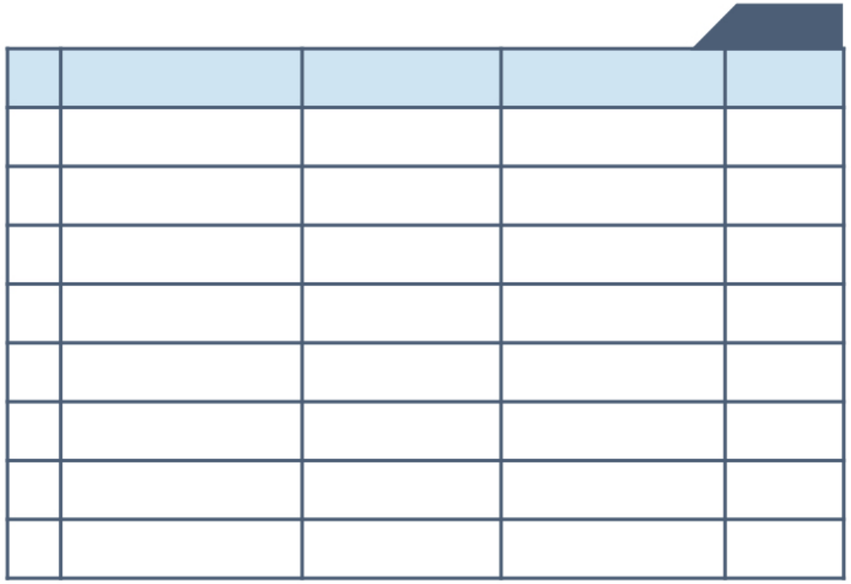
company

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



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

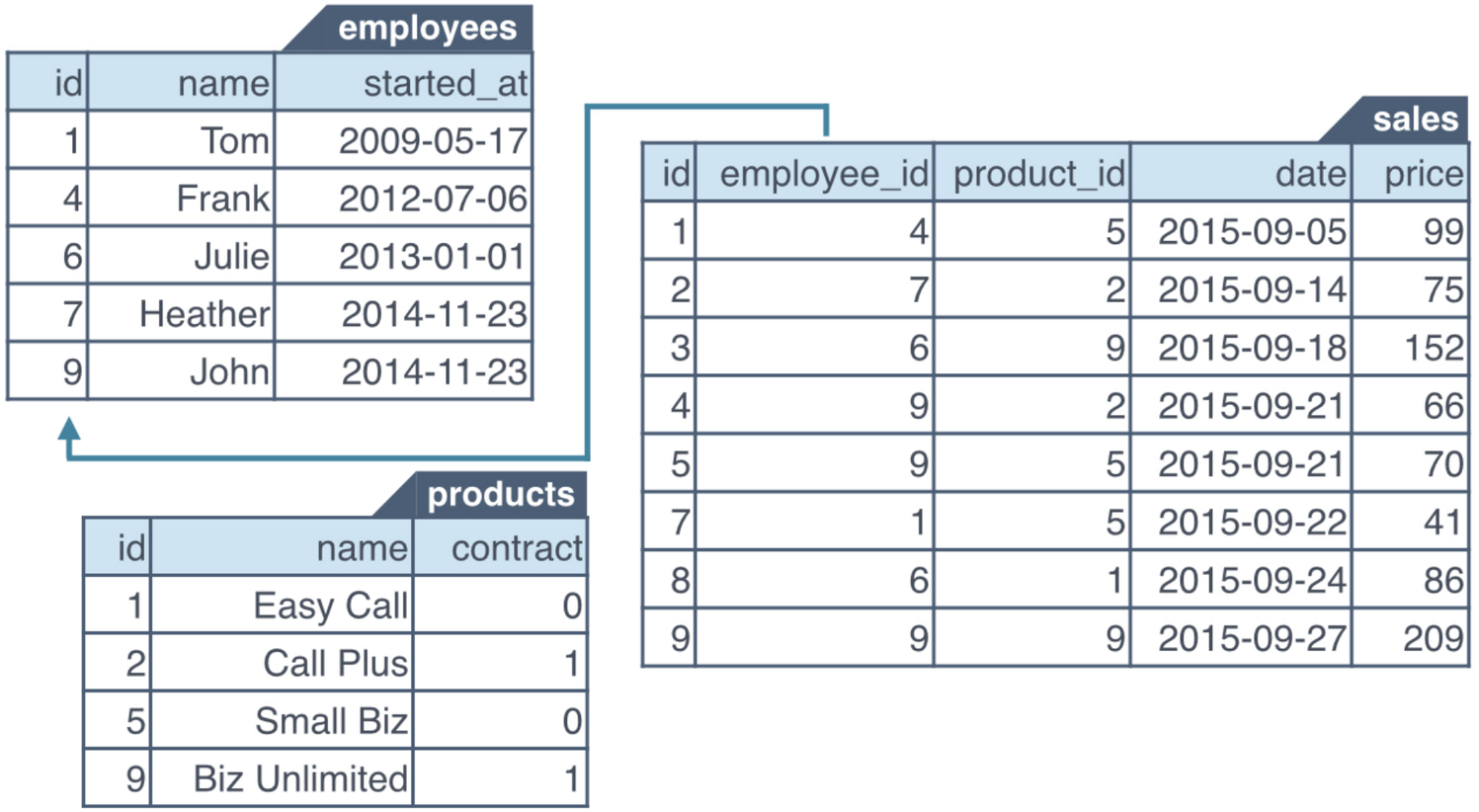
products

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

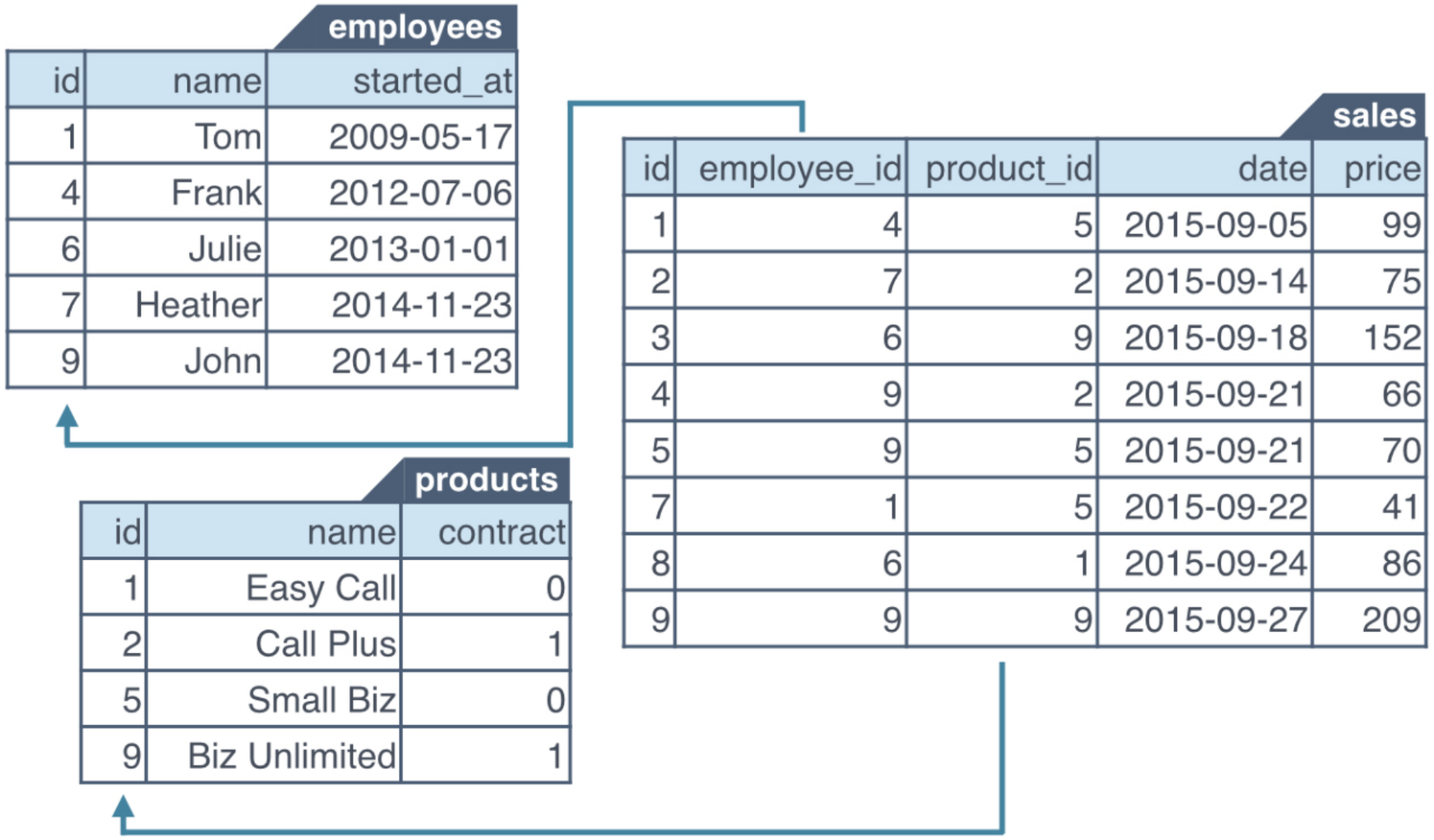
products

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

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

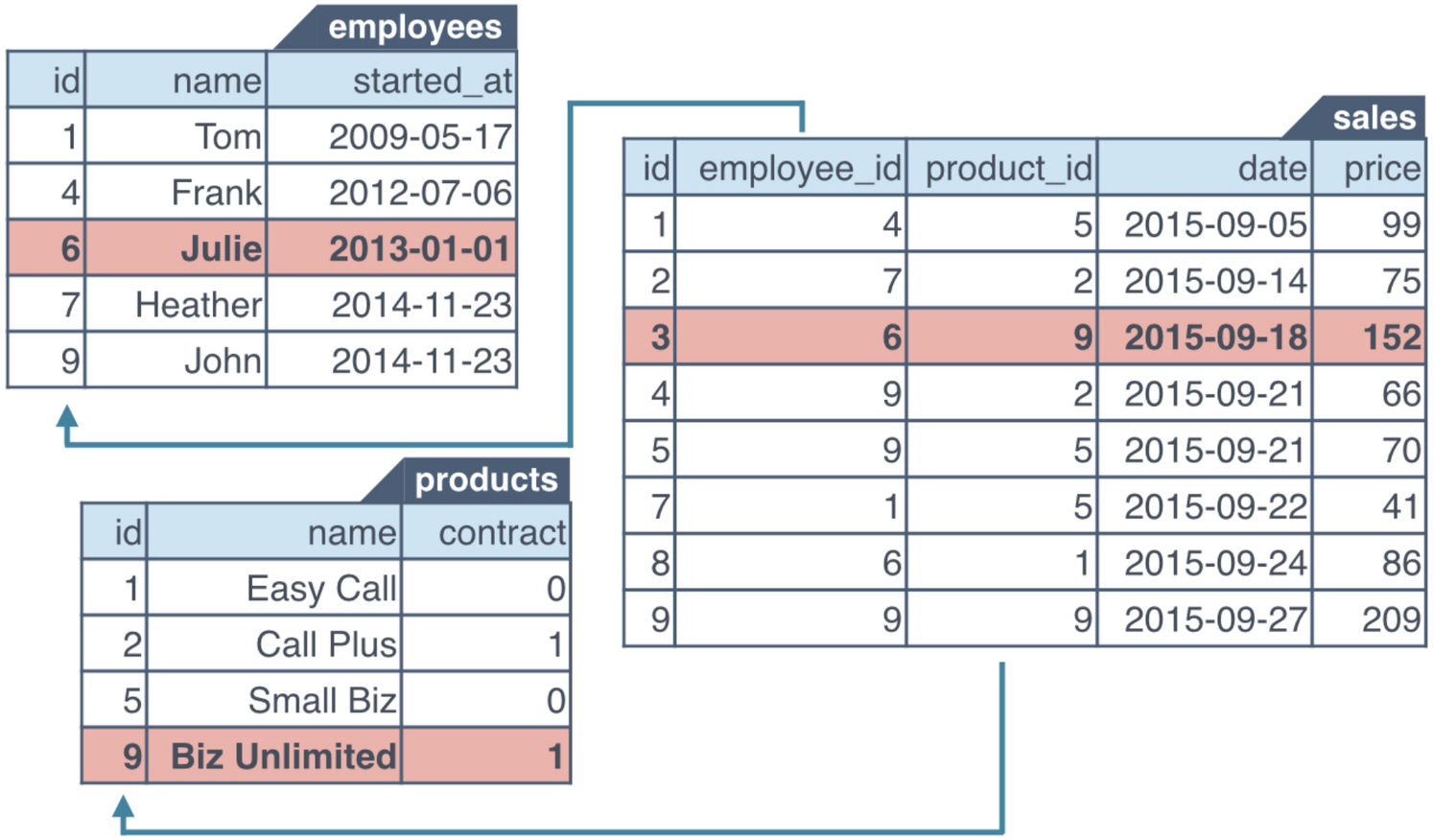
sales

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

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

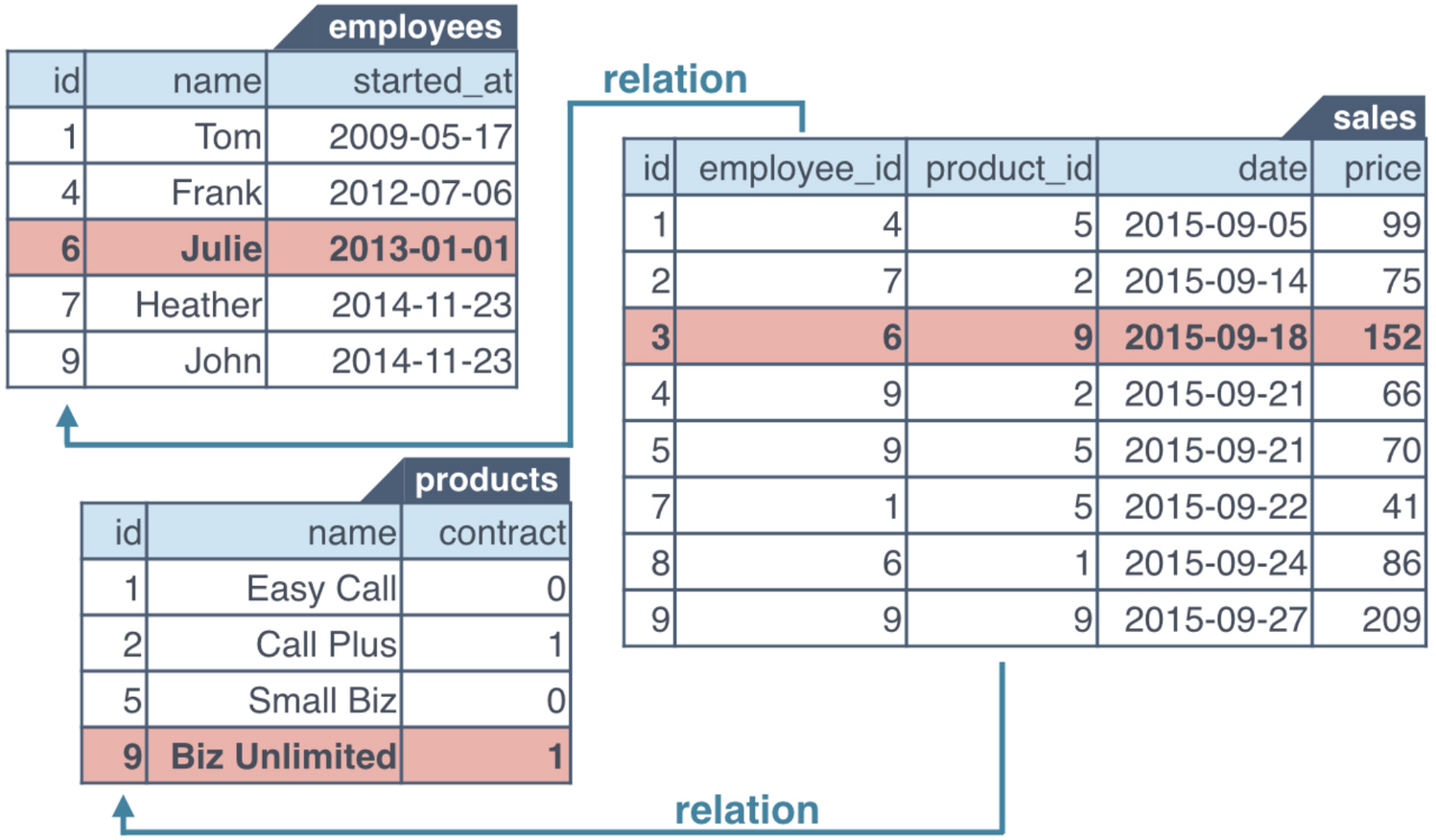
products

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

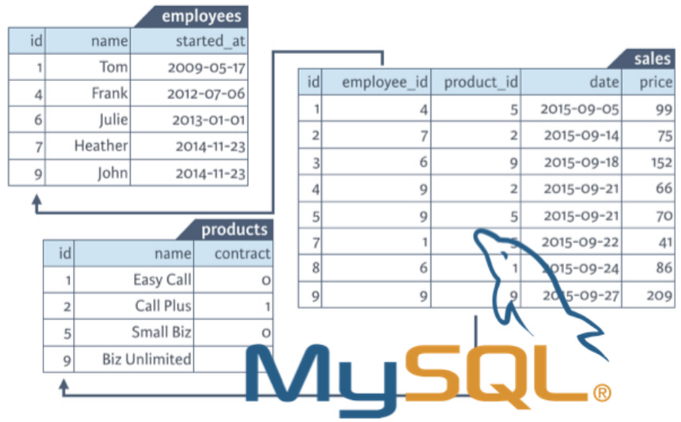

employees
id name started_at
1 Tom 2009-05-17
4 Frank 2012-07-06
6 Julie 2013-01-01
7 Heather 2014-11-23
9 John 2014-11-23
relation
sales
id employee_id product_id date price
1 4 5 2015-09-05 99
2 7 2 2015-09-14 75
3 6 9 2015-09-18 152
4 9 2 2015-09-21 66
5 9 5 2015-09-21 70
7 1 5 2015-09-22 41
8 6 1 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
relation

Database Management System

- DBMS
- Open source
  - MySQL, PostgreSQL, SQLite
- Proprietary
  - Oracle Database, Microsoft SQL Server
- SQL = Structured Query Language

Databases in R

- Different R packages
- MySQL: RMySQL
- PostgresSQL: RPostgresSQL
- Oracle Database: ROracle
- Conventions specified in DBI




```
install.packages("RMySQL")
library(DBI)
# library(RMySQL) not required
```

Connect to database

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

- con is DBIConnection object

Import table data

INTERMEDIATE IMPORTING DATA IN R

Filip Schouwenaars
Instructor, DataCamp

datacamp

con

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

List and import tables

```
dbListTables(con)
```

```
"employees" "products"  "sales"
```

	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

```
dbReadTable(con, "employees")
```

```
  id    name  started_at
1  1     Tom  2009-05-17
2  4   Frank  2012-07-06
3  6   Julie  2013-01-01
4  7 Heather  2014-11-23
5  9    John  2015-05-12
```

```
dbDisconnect(con)
```

```
TRUE
```

```
con
```

```
Error in .local(dbObj, ...) :
  internal error in RS_DBI_getConnection: ...
```

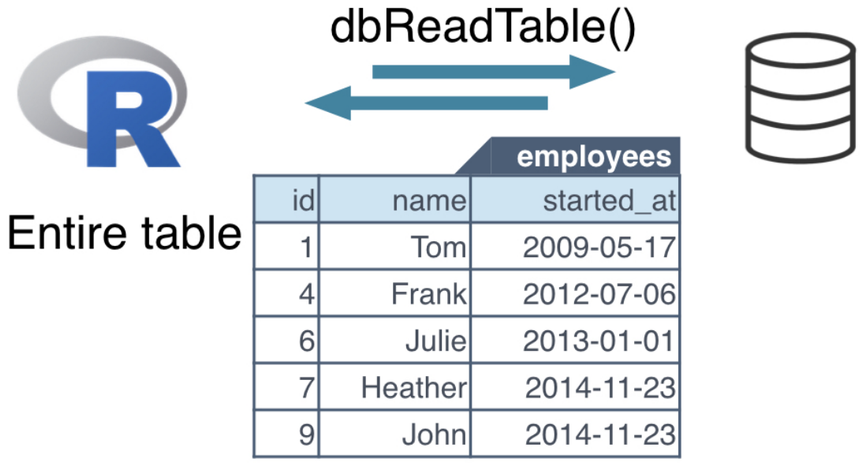
SQL Queries from inside R

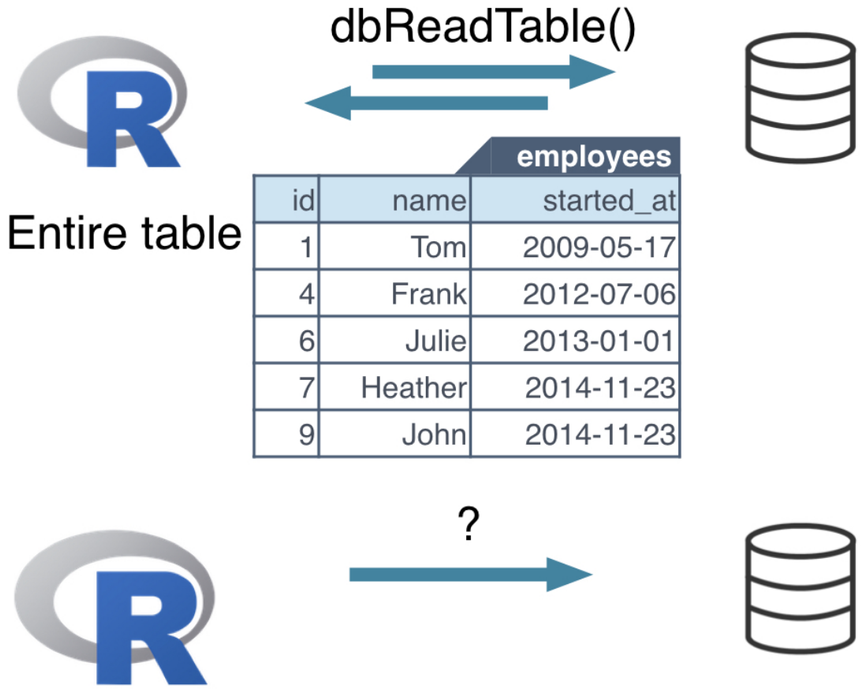
INTERMEDIATE IMPORTING DATA IN R

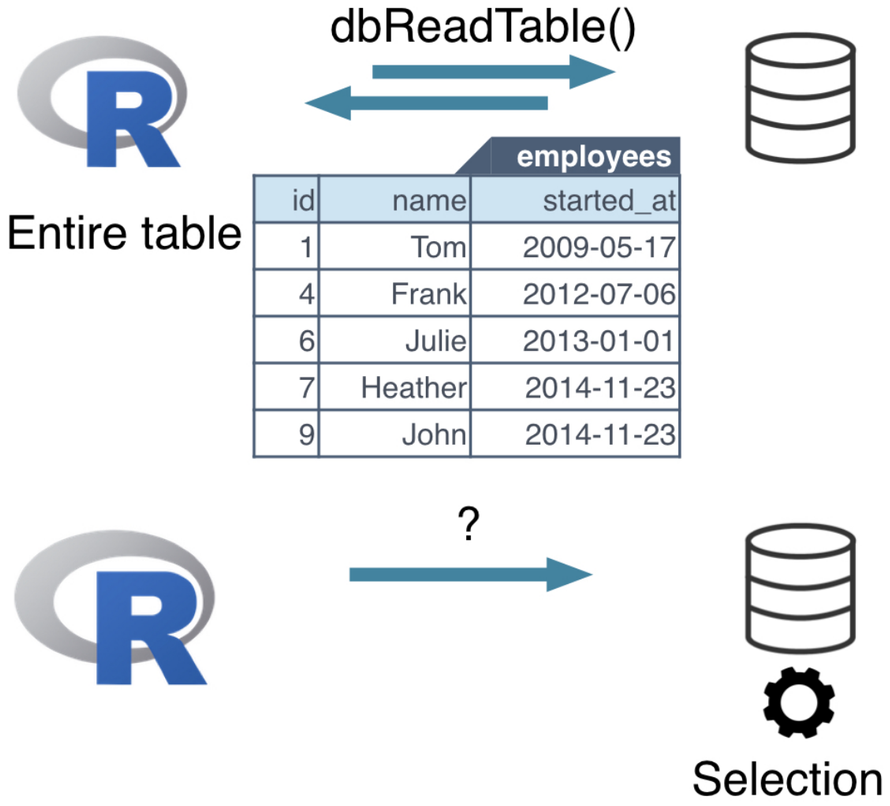
Filip Schouwenaars
Instructor, DataCamp

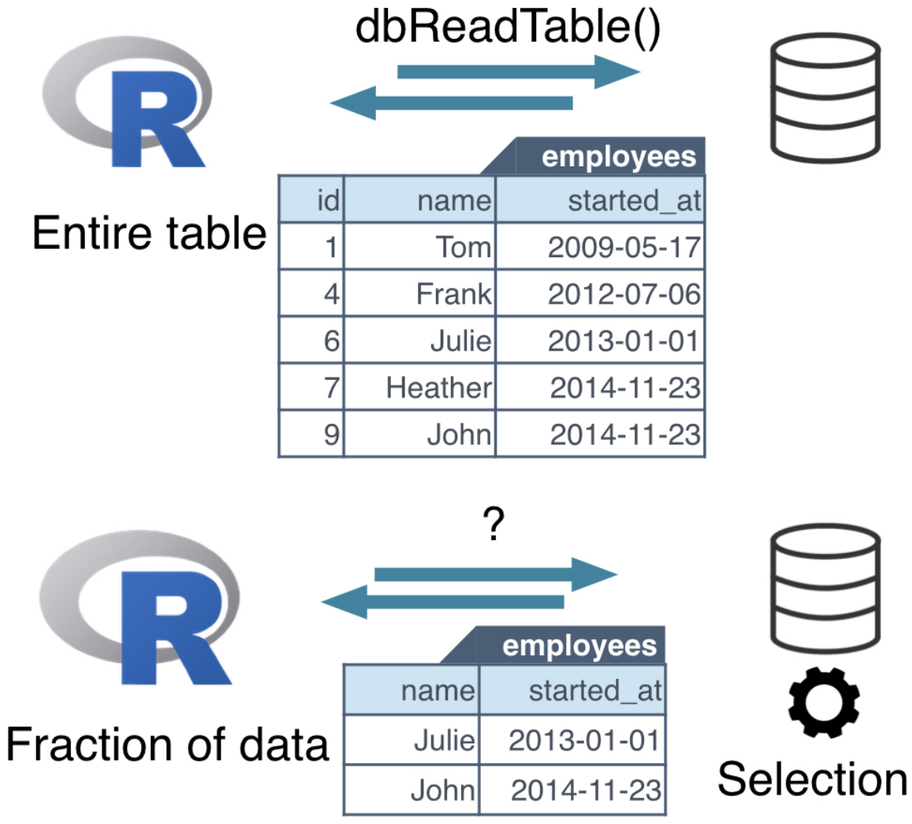
datacamp


dbReadTable()


dbReadTable()
Entire table
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


dbReadTable()
Entire table
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
?

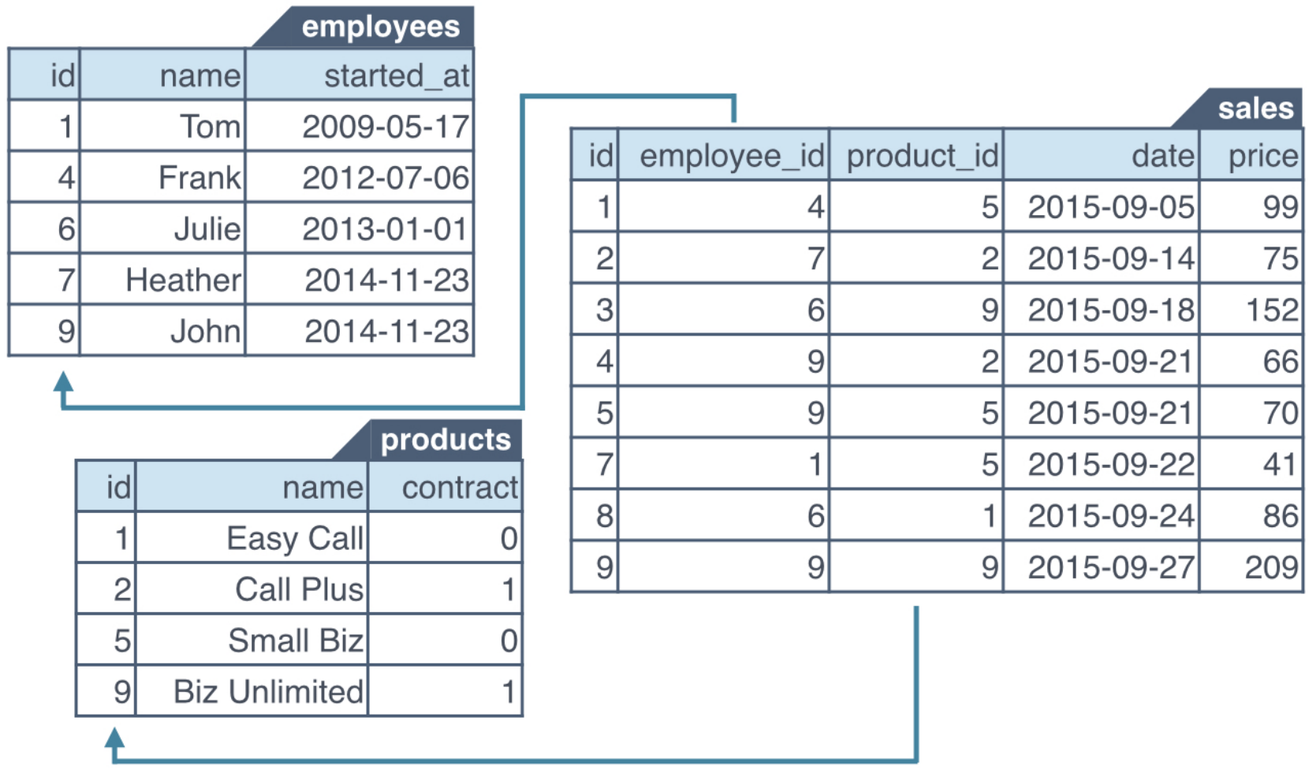

dbReadTable()
Entire table
employees
id name started_at
1 Tom 2009-05-17
4 Frank 2012-07-06
6 Julie 2013-01-01
7 Heather 2014-11-23
9 John 2014-11-23
?
Selection


dbReadTable()
Entire table
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

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

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
```

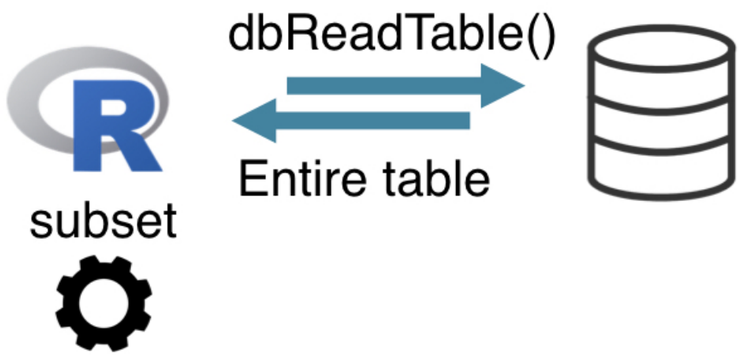
```
dbGetQuery(con, "SELECT name FROM employees
                      WHERE started_at > "2012-09-01"")
```

```
     name
1   Julie
2 Heather
3    John
```

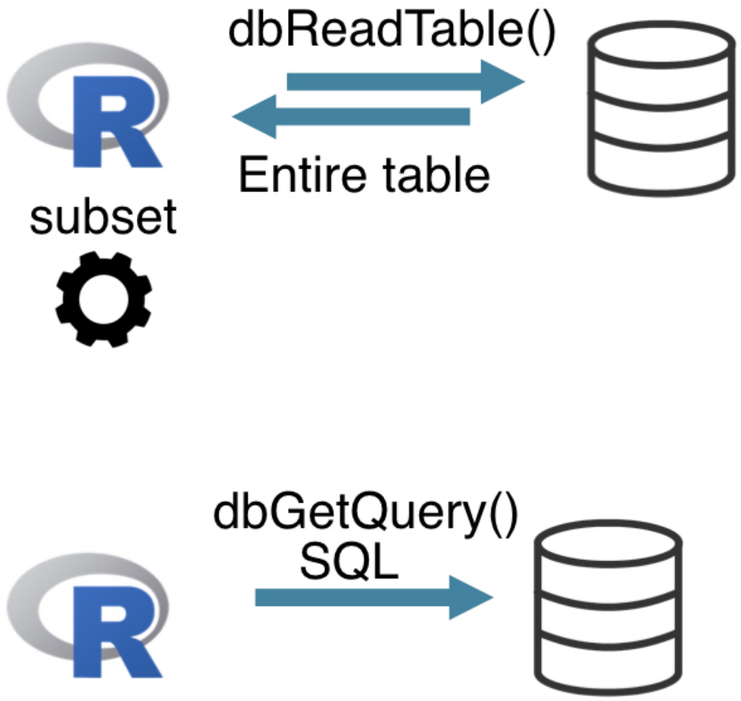
Example 1

Example 1

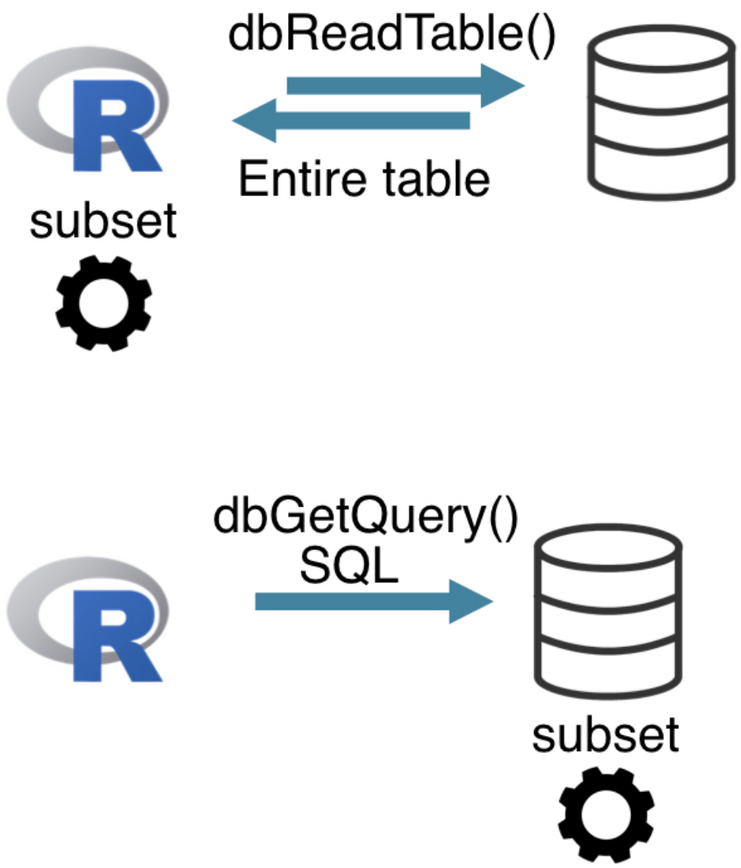
Example 1



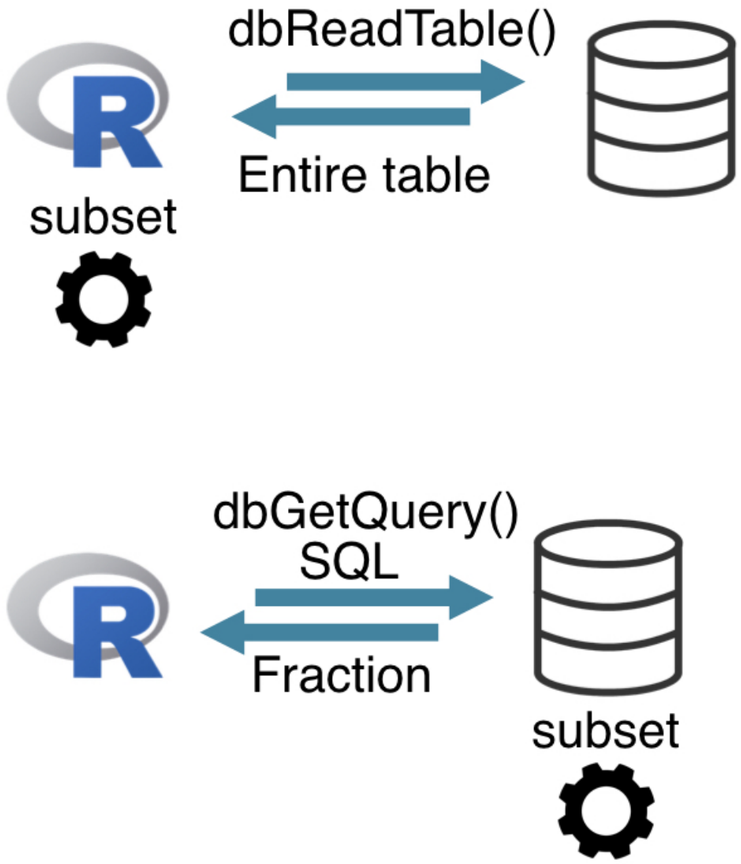
Example 1



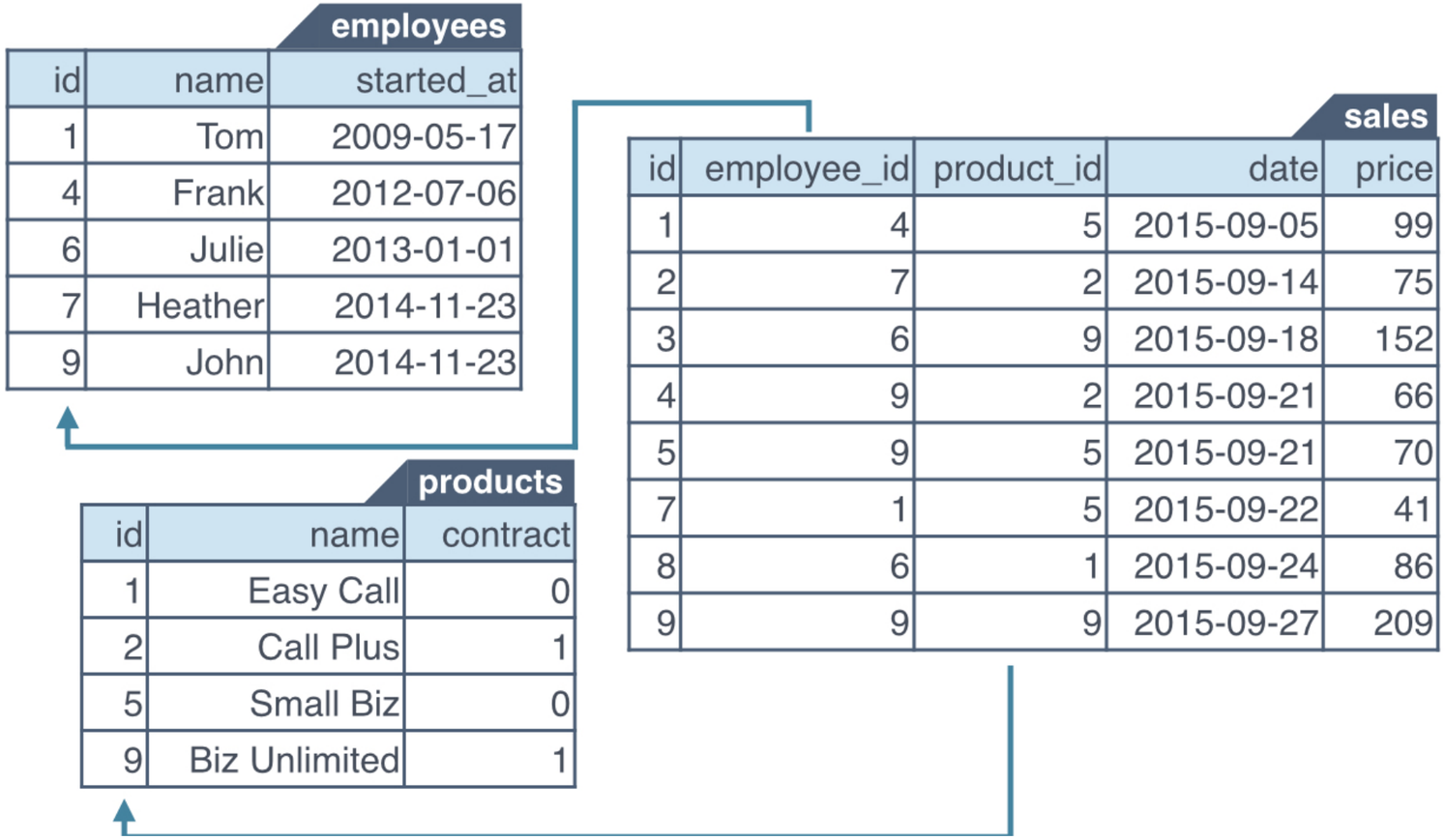
Example 1



Example 1


dbReadTable()
Entire table
subset
dbGetQuery()
SQL
Fraction
subset

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

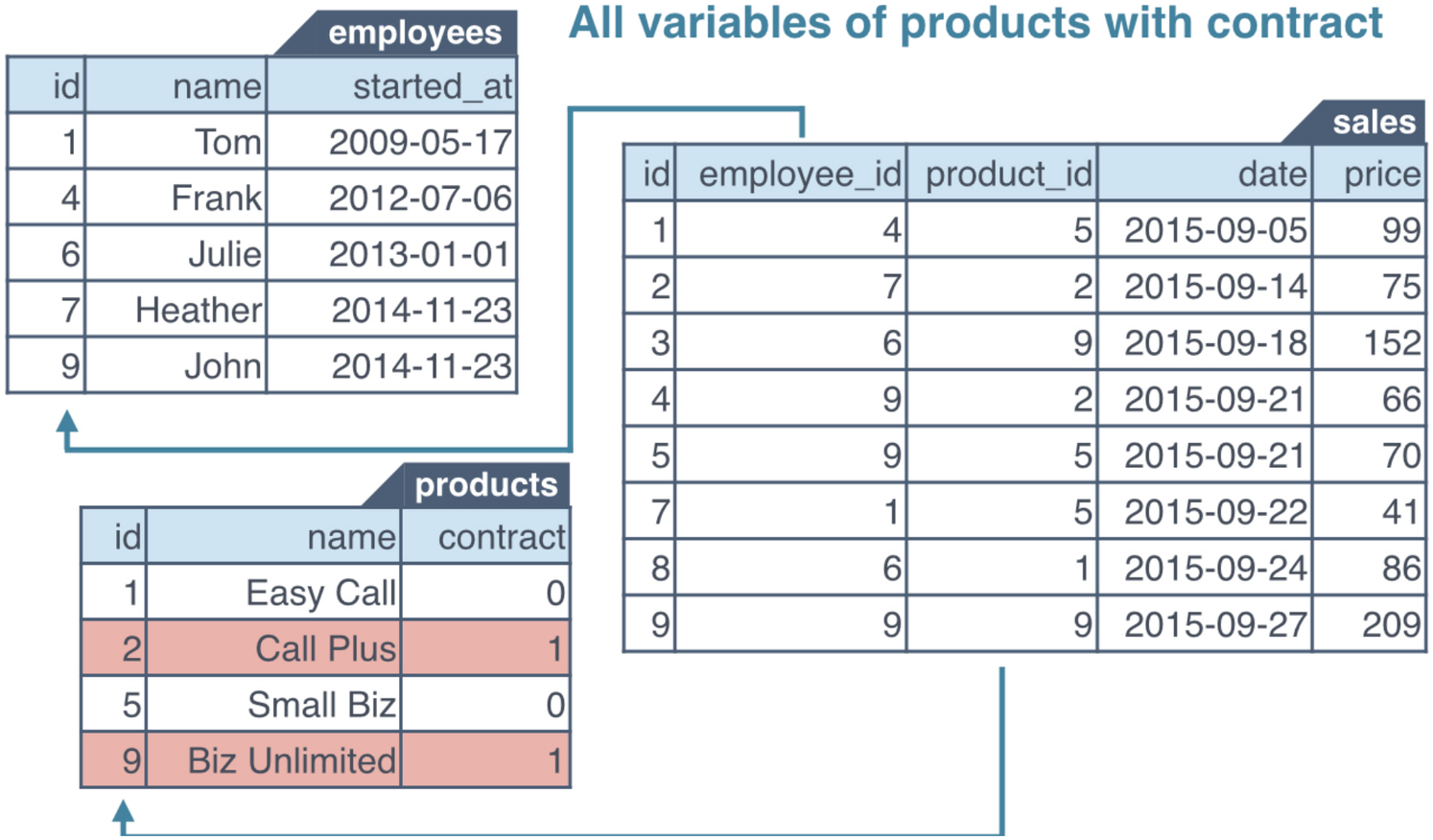
sales

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

company


All variables of products with contract
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

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
```

DBI internals

INTERMEDIATE IMPORTING DATA IN R

Filip Schouwenaars
Instructor, DataCamp

datacamp

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)
```

```
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
id       name     contract
<0 rows> (or 0-length row.names)
```

```
dbClearResult(res)
```

```
TRUE
```

Disconnect

```
dbDisconnect(con)
```

```
TRUE
```