

Loading Data from a Database

making a connection to the database as follows

```
In [3]: library(RMySQL)
```

```
Warning message:
"package 'RMySQL' was built under R version 3.6.3"Loading required package: DBI
Warning message:
"package 'DBI' was built under R version 3.6.3"
```

```
In [7]: con <- dbConnect(RMySQL::MySQL(), dbname="tweater",
                        host="courses.csrrinzqubik.us-east-1.rds.amazonaws.com", port=
                        3306,
                        user="student", password="datacamp"
                        )
```

```
In [8]: tables = dbListTables(con)
        str(tables)

chr [1:3] "comments" "tweets" "users"
```

read a specific table called users

```
In [9]: users <- dbReadTable(con, "users")
        print(users)
```

	id	name	login
1	1	elisabeth	elismith
2	2	mike	mikey
3	3	thea	teatime
4	4	thomas	tomatotom
5	5	oliver	olivander
6	6	kate	katebenn
7	7	anjali	lianja

Exercises

1. Import the data from the other tables and print the data

```
In [11]: # import data from "tweats" table
tweats_table = dbReadTable(con, "tweats")
print(tweats_table)
```

	id	user_id	post	date
1	75	3	break egg. bake egg. eat egg.	2015-09-05
2	88	4	wash strawberries. add ice. blend. enjoy.	2015-09-14
3	77	6	2 slices of bread. add cheese. grill. heaven.	2015-09-21
4	87	5	open and crush avocado. add shrimps. perfect starter.	2015-09-22
5	49	1	nachos. add tomato sauce, minced meat and cheese. oven for 10 mins.	2015-09-22
6	24	7	just eat an apple. simply and healthy.	2015-09-24

```
In [13]: summary(tweats_table)
```

	id	user_id	post	date
Min.	:24.00	Min. :1.000	Length:6	Length:6
1st Qu.	:55.50	1st Qu.:3.250	Class :character	Class :character
Median	:76.00	Median :4.500	Mode :character	Mode :character
Mean	:66.67	Mean :4.333		
3rd Qu.	:84.50	3rd Qu.:5.750		
Max.	:88.00	Max. :7.000		

2. Determine the relations among them

There are 3 columns on the tweats table: `id` , `user_id` , `post` . The `id` is the number assigned by the database for identifying the data generated by a particular `user_id` . In this case, every `user_id` generate two categories of information : `date` and `post` . The `post` column contains tweet post generated by each `user_id` whereas the `date` column gives the date of each `post`

```
In [20]: # import the data from comments table
comments_table <- dbReadTable(con, "comments")
print(comments_table)
```

	id	tweet_id	user_id	message
1	1022	87	7	nice!
2	1000	77	7	great!
3	1011	49	5	love it
4	1012	87	1	awesome! thanks!
5	1010	88	6	yuck!
6	1026	77	4	not my thing!
7	1004	49	1	this is fabulous!
8	1030	75	6	so easy!
9	1025	88	2	oh yes
10	1007	49	3	serious?
11	1020	77	1	couldn't be better
12	1014	77	1	saved my day

3. Comment 1012: what is the tweet message? who post the tweet? who post the comment?

comment 1012 was posted by `user_id 1` which is elisabeth. It was a comment on a `tweet_id 87` posted by `user_id 5` (oliver). The tweet message was: *"nachos. add tomato sauce, minced meat and cheese. oven for 10 mins."*