LORENZO AUSIELLO

ASSIGNMENT 3

PROBLEM 1

```
-- Import data from banks sec 2002 and banks al 2002. Delete duplicate
-- rows from banks sec 2002

DROP TABLE IF EXISTS banks_sec_2002;
SET datestyle = 'MDY';
CREATE TABLE banks_sec_2002(
  id INTEGER,
  date DATE,
  security INTEGER
);

SELECT * FROM banks_sec_2002;

COPY banks_sec_2002 (id, date, security)
FROM 'C:\Users\Public\banks_sec_2002.csv'
DELIMITER ','
CSV HEADER;

SELECT id, date, security, COUNT(*)
FROM banks_sec_2002
GROUP BY id, date, security
HAVING COUNT(*) > 1;

DELETE FROM banks_sec_2002 a
USING (
   SELECT id, date, security, MIN(ctid) AS min_ctid
   FROM banks_sec_2002
```

```
   GROUP BY id, date, security
   HAVING COUNT(*) > 1
) AS b
WHERE a.id = b.id
  AND a.date = b.date
  AND a.security = b.security
  AND a.ctid <> b.min_ctid;



DROP TABLE IF EXISTS banks_al_2002;
CREATE TABLE banks_al_2002(
  id INTEGER,
  date DATE,
  asset INTEGER,
          liability INTEGER
);

SELECT * FROM banks_al_2002;

COPY banks_al_2002 (id, date, asset,liability)
FROM 'C:\Users\Public\banks_al_2002-1.csv'
DELIMITER ','
CSV HEADER;

SELECT id, date, asset, liability, COUNT(*)
FROM banks_al_2002
GROUP BY id, date, asset, liability
HAVING COUNT(*) > 1;



-- Select proper join manner to join banks sec 2002 and banks al 2002. Make
-- sure that all data from banks sec 2002 are kept in the joint table. Report
```

-- the first 10 observations.

```
SELECT *
FROM banks_sec_2002 AS bs
LEFT JOIN banks_al_2002 AS ba
ON bs.id = ba.id
AND bs.date = ba.date
LIMIT 10;
```

	id integer	date date	security integer	id integer	date date	asset integer	liability integer
1	32307	2002-09-30	0	32307	2002-09-30	53714	49350
2	22598	2002-03-31	0	22598	2002-03-31	57360	53205
3	15879	2002-06-30	5357	15879	2002-06-30	76960	62945
4	35373	2002-06-30	0	35373	2002-06-30	46551	38721
5	5226	2002-09-30	7960	5226	2002-09-30	53873	48146
6	22092	2002-12-31	0	22092	2002-12-31	147828	135596
7	13749	2002-03-31	17476	13749	2002-03-31	214733	189972
8	823	2002-12-31	44972	823	2002-12-31	271961	241138
9	29831	2002-03-31	6505	29831	2002-03-31	58300	43689
10	10203	2002-09-30	3889	10203	2002-09-30	320853	293453

-- Create a new table banks total. Insert the values from previous joint table

-- into this new one. And set a primary key for the table.

```
DROP TABLE IF EXISTS banks_total;
CREATE TABLE banks_total (
   id INTEGER,
   date DATE,
   security INTEGER,
   asset INTEGER,
   liability INTEGER,
         PRIMARY KEY (id, date)
);

SELECT * FROM banks_total;
```

```
INSERT INTO banks_total (id, date, security, asset, liability)
SELECT bs.id, bs.date, bs.security, ba.asset, ba.liability
FROM banks_sec_2002 AS bs
LEFT JOIN banks_al_2002 AS ba
ON bs.id = ba.id
AND bs.date = ba.date;
SELECT * FROM banks_total;
```

	id [PK] integer	date [PK] date	security integer	asset integer	liability integer
1	32307	2002-09-30	0	53714	49350
2	22598	2002-03-31	0	57360	53205
3	15879	2002-06-30	5357	76960	62945
4	35373	2002-06-30	0	46551	38721
5	5226	2002-09-30	7960	53873	48146
6	22092	2002-12-31	0	147828	135596
7	13749	2002-03-31	17476	214733	189972
8	823	2002-12-31	44972	271961	241138
9	29831	2002-03-31	6505	58300	43689
10	10203	2002-09-30	3889	320853	293453

```
-- For each quarter of the year 2002 count how many banks have security
-- over 20% of its' asset.

SELECT
    EXTRACT(quarter FROM date) AS quarter,
    COUNT(*) AS banks_count_over_20_percent
FROM banks_total
WHERE
    EXTRACT(year FROM date) = 2002
    AND security > (0.2 * asset)
GROUP BY quarter
ORDER BY quarter;
```

	quarter numeric	banks_count_over_20_percent bigint
1	1	984
2	2	1023
3	3	1033
4	4	1048

-- How many banks have liability over 90% of assets in first quarter of 2002

-- but goes below 90% in the second quarter of 2002

```
SELECT COUNT(DISTINCT t1.id)
FROM banks_total t1
JOIN banks_total t2 ON t1.id = t2.id
WHERE t1.date = '2002-03-31' AND t2.date = '2002-06-30'
AND t1.liability > 0.9 * t1.asset
AND t2.liability < 0.9 * t2.asset;
```

	count bigint
1	388

-- create csv

```
COPY banks_total
TO 'C:\Users\Public\banks_total_export.csv'
DELIMITER ','
CSV HEADER;
```

PROBLEM 2

```
library("RPostgres")

## Make a connection to your local PostgreSQL database using API.
con <- dbConnect(RPostgres::Postgres(),
```

```
        dbname = "postgres",
        host = "127.0.0.1",
        port = 5432,
        user = "postgres",
        password = '********')

## Import the csv file you got from Problem 1 (banks total)
## into a new table in the database using API.
setwd('C:/Users/Public')
dbWriteTable(con, "new_banks_total", read.csv("banks_total_export.csv"))

## Retrieve the data of table 'banks total' using API. Count how many rows
## in the table.
banks_total <- dbGetQuery(con, "SELECT * FROM banks_total")
result <- dbGetQuery(con, "SELECT COUNT(*) FROM banks_total")
head(banks_total)
```

```
> head(banks_total)
     id       date security  asset liability
1 32307 2002-09-30        0  53714     49350
2 22598 2002-03-31        0  57360     53205
3 15879 2002-06-30     5357  76960     62945
4 35373 2002-06-30        0  46551     38721
5  5226 2002-09-30     7960  53873     48146
6 22092 2002-12-31        0 147828    135596
>
```

```
print(result)
```

```
> print(result)
  count
1 37819
>
```

```
dbDisconnect(con)
```