PSQL

Jim Harner

1/7/2020

The Data Expo data set consists of seven atmospheric measurements at locations on a 24 by 24 grid averaged over each month for six years (72 time points). The elevation (height above sea level) at each location is also included in the data set.

The table schema for `dataexpo` is defined as follows.

```
date_table ( ID [PK], date, month, year )

location_table ( ID [PK], longitude, latitude, elevation )

measure_table ( date [PK] [FK date_table.ID],
                location [PK] [FK location_table.ID],
                cloudhigh, cloudlow, cloudmid, ozone,
                pressure, surftemp, temperature )
```

The `dataexpo` database can be invoked using `psql` in RStudio's `bash` shell as follows:

```
psql -h postgres dataexpo
```

The `-w` option causes a prompt for your password, but is not needed for the Dockerized version of this course. `psql` is in `/usr/bin`, which is in the `PATH` environmental variable, i.e., it is not necessary to invoke by `/usr/bin/psql`.

The `-h` option specifies the host, which in this case is `postgres`. It is not needed if the Postgres is on the same machine as RStudio, but in the Dockerized version Postgres is in a separate container called `postgres`.

Databases typically are only setup by the database administrator (DBA). Once established you can populate it with tables if you have write permissions. Tables could be added to the `dataexpo` database by the following command if they are not already there. But don't since the database is populated.

```
# Do not run!
psql -h postgres dataexpo < dataexpo.sql
```

`dataexpo.sql` is in your working directory and it contains code for constructing tables (and their schema) and inserting the data into these tables. The order of creating tables (`CREATE TABLE`) is important since a table must be present before it can be referenced.

If you have not done so, enter interactive mode in a terminal by:

```
psql -h postgres dataexpo
```

Try it in RStudio's shell.

Once in interactive mode, the `psql` commands for listing the tables in the database are `\d` and for specific information about a specific table `\d table`. At the `dataexpo` prompt type:

```
\d
```

```
            List of relations
 Schema |      Name      | Type  |  Owner
--------+----------------+-------+---------
 public | date_table     | table | rstudio
 public | location_table | table | rstudio
 public | measure_table  | table | rstudio
```

```
(3 rows)

\d date_table

         Table "public.date_table"
 Column |         Type          | Modifiers
--------+-----------------------+-----------
 id     | integer               | not null
 date   | date                  |
 month  | character varying(10) |
 year   | integer               |
Indexes:
    "date_table_pkey" PRIMARY KEY, btree (id)
Referenced by:
    TABLE "measure_table" CONSTRAINT "measure_date_table_fk" FOREIGN KEY (date) REFERENCES date_table(id

\q
```

The last command quits `psql`.

To get help use:

- `\h` to list SQL commands;
- `\h command` to show the syntax for `command`;
- `\?` to list psql commands

You can run batch commands in `psql` by putting a SQL `--command` in quotes.

```
psql -h postgres  dataexpo --command "select * from location_table limit 5"
```

```
##  id | longitude | latitude | elevation
## ----+-----------+----------+-----------
##   1 |   -113.75 |    36.25 |   1526.25
##   2 |   -111.25 |    36.25 |   1759.56
##   3 |   -108.75 |    36.25 |   1948.38
##   4 |   -106.25 |    36.25 |   2241.31
##   5 |   -103.75 |    36.25 |   1692.75
## (5 rows)
```