LEDE PROGRAM: DATA AND DATABASES DAY 4

CSV to SQL

We will download two CSV files from http://data.un.org/Explorer.aspx

In the Energy Statistics Database download Solar Electricity and Wind Electricity by clicking on "View Data" and then on that page click "download", and under "value separated" choose "comma".

Here is the process for "solar" (please follow the same process for "wind")

The file you download will be called something like: UNdata_Export_20170530_183137975.zip

Unzip the file, and rename it something readable, like "UN_solar.csv"

Put that file in a folder on your computer (not in a remote/cloud drive, but in the computer itself). Make sure there are no spaces in the name of the folders it is in.

Maybe open the file in Excel in order to read the column names and see what kind of data is in each column. (Make sure you never save the file while in Excel--even if you made no changes, Excel may still add invisible and additional characters that will make the file unreadable in the terminal.)

Based on the columns write a CREATE TABLE statement that contains columns and datatypes for each column. For example:

```
CREATE TABLE solar (
  country varchar(40),
  type varchar(40),
  year int,
  unit varchar(40),
  usage double precision,
  notes varchar(40)
);
```

Once you successfully created the table in PostgreSQL, you need to **prep this CSV** for processing.

Again, make sure that you have not saved the document in Excel. Open the document in a text editor that does not add invisible characters. These include sublimetext, textwrangler, notepad++, and others -- don't use the text editor that is preinstalled on your computer. Make sure it is made for development/programming.

Once you have found the proper text editor, open up the CSV file and delete the header, then scroll down to the bottom and delete any footers (lines that come after the rows). So your entire document should consistently look like this:

```
"Argentina","Electricity - total solar production","2010","Kilowatt-hours, million","0.09",""
"Argentina","Electricity - total solar production","2009","Kilowatt-hours, million","0.08",""
"Australia","Electricity - total solar production","2014","Kilowatt-hours, million","4858",""
"Australia","Electricity - total solar production","2013","Kilowatt-hours, million","3826",""
```

Once that is prepped, save the file and close it.

Now you can run the copy command:

```
\copy tablename from pathname delimiter ',' csv
```

or, more specifically:

```
\copy solar from /Users/Jon/Documents/columbia_syllabus/undata-class/testing.csv delimiter ',' csv
```

Make sure that the full path you use has no spaces in it, and that it ends with .csv

If this worked you will get a message like this

```
COPY 1105
```

In which 1105 is the number of rows that were copied into the table!