Assignment 02 - Select data on per-capita income

Due Feb 10

200 pts

Create a table that matches with the data in `clean-data.csv` .

Load the data in `clean-data.csv` using the `load-bea-data.sql` script.

Load the data in `election_2020.csv` using the `load-election_2020.sql` script.

Construct a select statement that will count the number of counties that each candidate won.

Example Output:

```
$ psql
Pager usage is off.
psql (14.0, server 13.4)
Type "help" for help.

philip=# \o part1.out
philip=# \i ans-count-counties.sql
      winner       | # Of Counties Won
-------------------+-------------------
 Joseph R Biden Jr |               293
 Donald J Trump    |              1798
(2 rows)
philip=# \q
```

Construct a select statement that will take the data and combine it with the county election data in the table `election_2020` (data provided - see the lecture 2 data).

Print out 2 columns of data. The columns should be "Winner" and "Average Per Capita 2020 Income".

Example Output:

```
$ psql
Pager usage is off.
psql (14.0, server 13.4)
Type "help" for help.

philip=# \o part2.out
philip=# \i ans-per-capita-by-candidate.sql
      Winner       | Average Per Capita 2020 Income
-------------------+-------------------------------
 Joseph R Biden Jr |                       65488.43
 Donald J Trump    |                       48000.67
(2 rows)
philip=# \q
```

Note... there is an error in this data - we will discuss it in class on Tuesday.

Formatting: the income number is formatted using a type cast of `::numeric(10,2)` .

To do the query you will need to join to the table that lists the per-county winners, election_2020.

```
1: drop table if exists election_2020 ;
2: create table election_2020 (
3:     id            serial primary key,
4:     county_name  text,
5:     state        text,
6:     winner        text -- either 'Joseph R Biden Jr' or 'Donald...'
7: );
```

Create and populate the table with using insert/update statements.

== Find the errors in the queries. Not all of the counties are represented because the data is missing some stuff. Not all states represent data as "county" data. For example, find "Albemarle + Charlottesville,"74,603","75,885","77,606",7,1.7,2.3,104" line 3004 in lapi1121.csv. This is skiped in the load program. So the data for this combined county and city is not in the

`bea_growth_data` table.

Turn in

1. Your output from the 2 queries.
2. A select from `election_2020` that groups and counts the number of counties that each candidate won. (Hint: Biden won 293 counties)
3. Your select statement that shows the average per capita income by counties that a candidate won. The output from this select. (Hint: should match the example output above).