

Using PostGIS and QGIS for the Management and Geovisualization of American Community Survey Data

Lee Hachadoorian
Visiting Assistant Professor
Department of Geography
Dartmouth College

Lee.Hachadoorian@gmail.com
<http://github.com/leehach/census-postgres>
<http://geospatial.commonsgc.cuny.edu/>

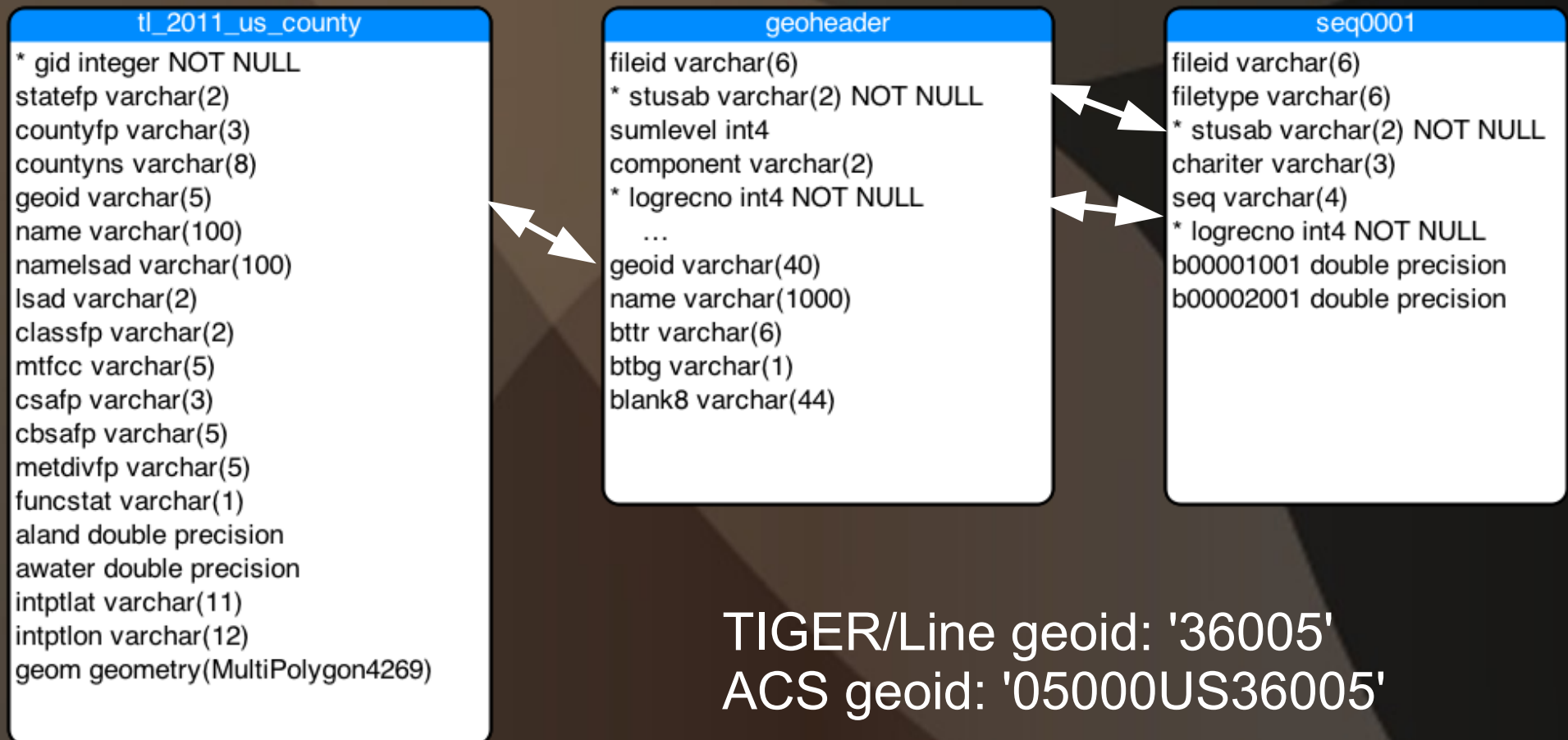
Motivating Concerns

- What are the issues with the ACS data as it is?
- Why use an RDBMS?
- Why use PostgreSQL? 

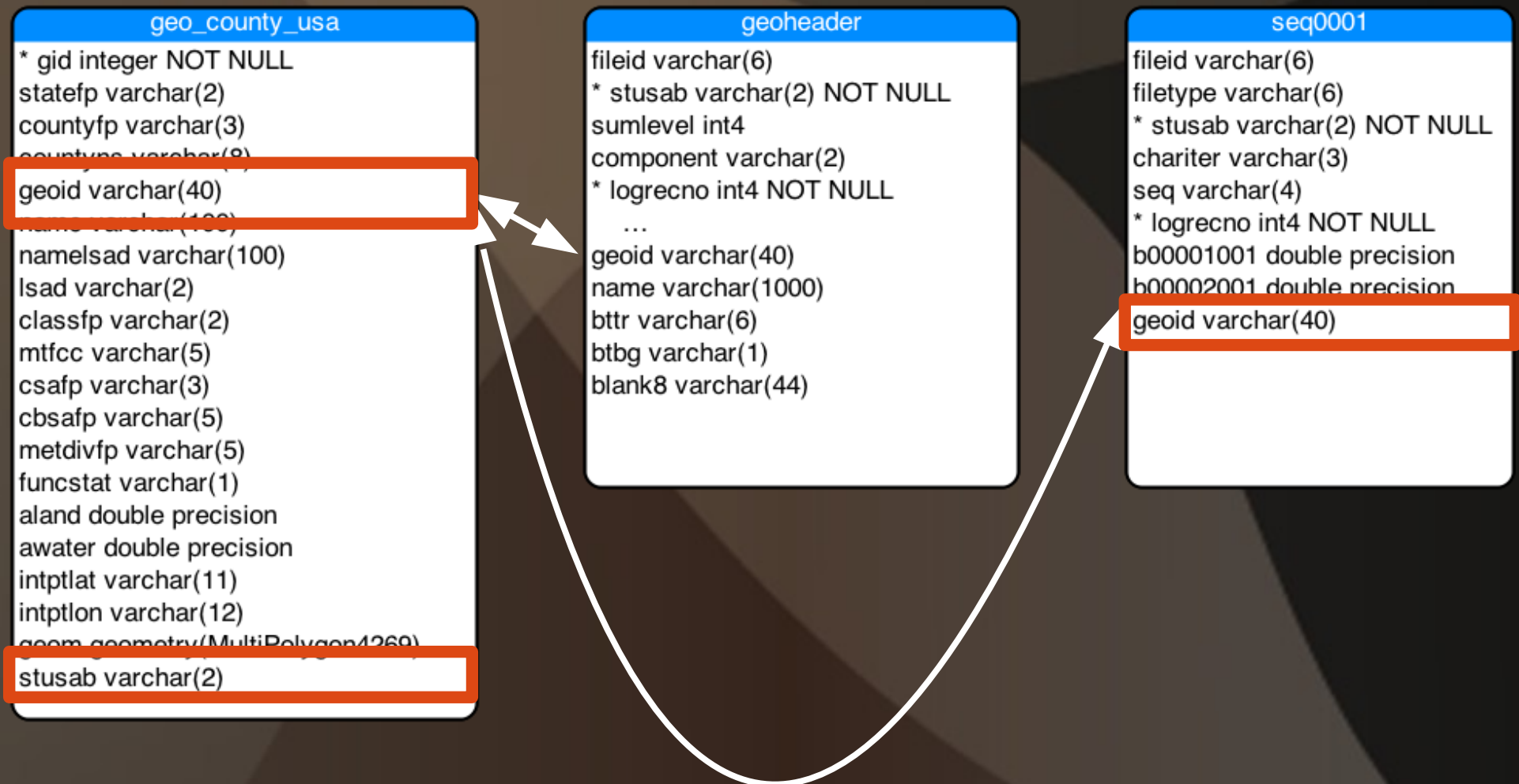
Database Organization

- Each USCB product is in its own schema
 - acs2011_5yr, acs2010_spt, etc...
- Support tables
 - data_dictionary
 - geoheader_schema
- Staging tables (one per sequence + geoheader)
- Data store (one table per sequence + geoheader)
- Views (one per subject table)

Store Sequences in Database Tables



Some Modifications to Facilitate Joins



Subject Tables are Views

```
CREATE OR REPLACE VIEW  
  acs2011_5yr.b01002 AS  
SELECT  
  stusab, logrecno, geoid,  
  b01002001, b01002002, b01002003  
FROM acs2011_5yr.seq0003;
```

Partitioning

Master table (contains no records)

Child table 1 (same structure as master table)

Child table 2 (same structure as master table)

...

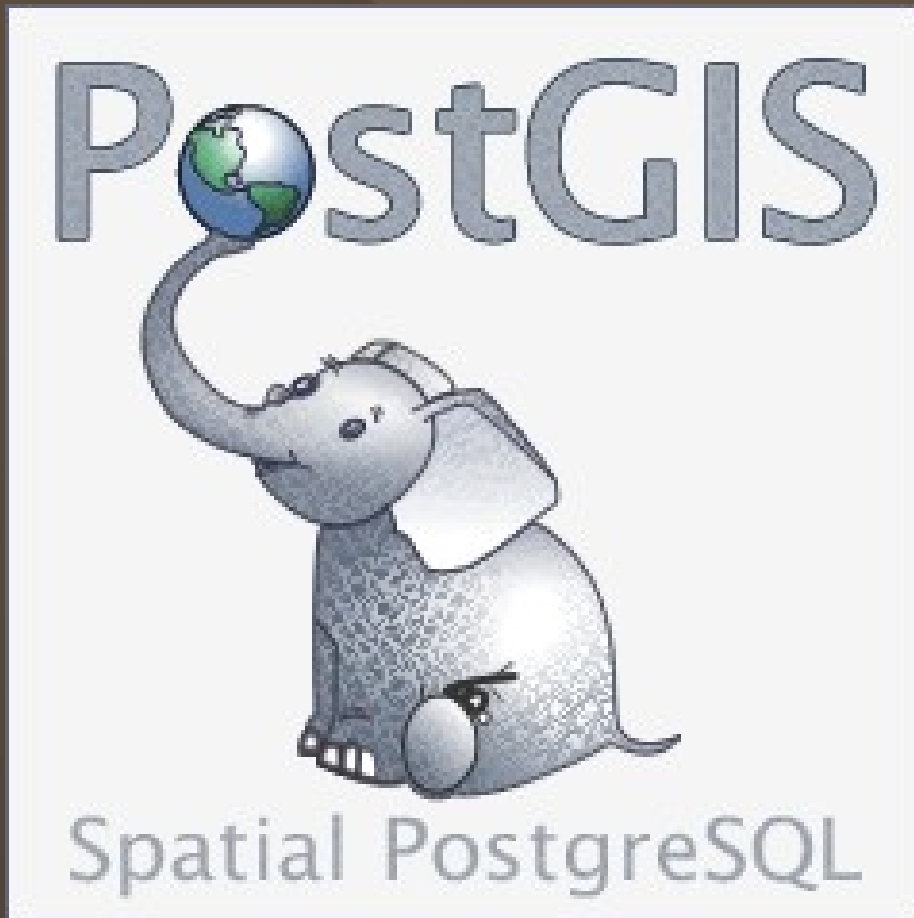
Child table N (same structure as master table)

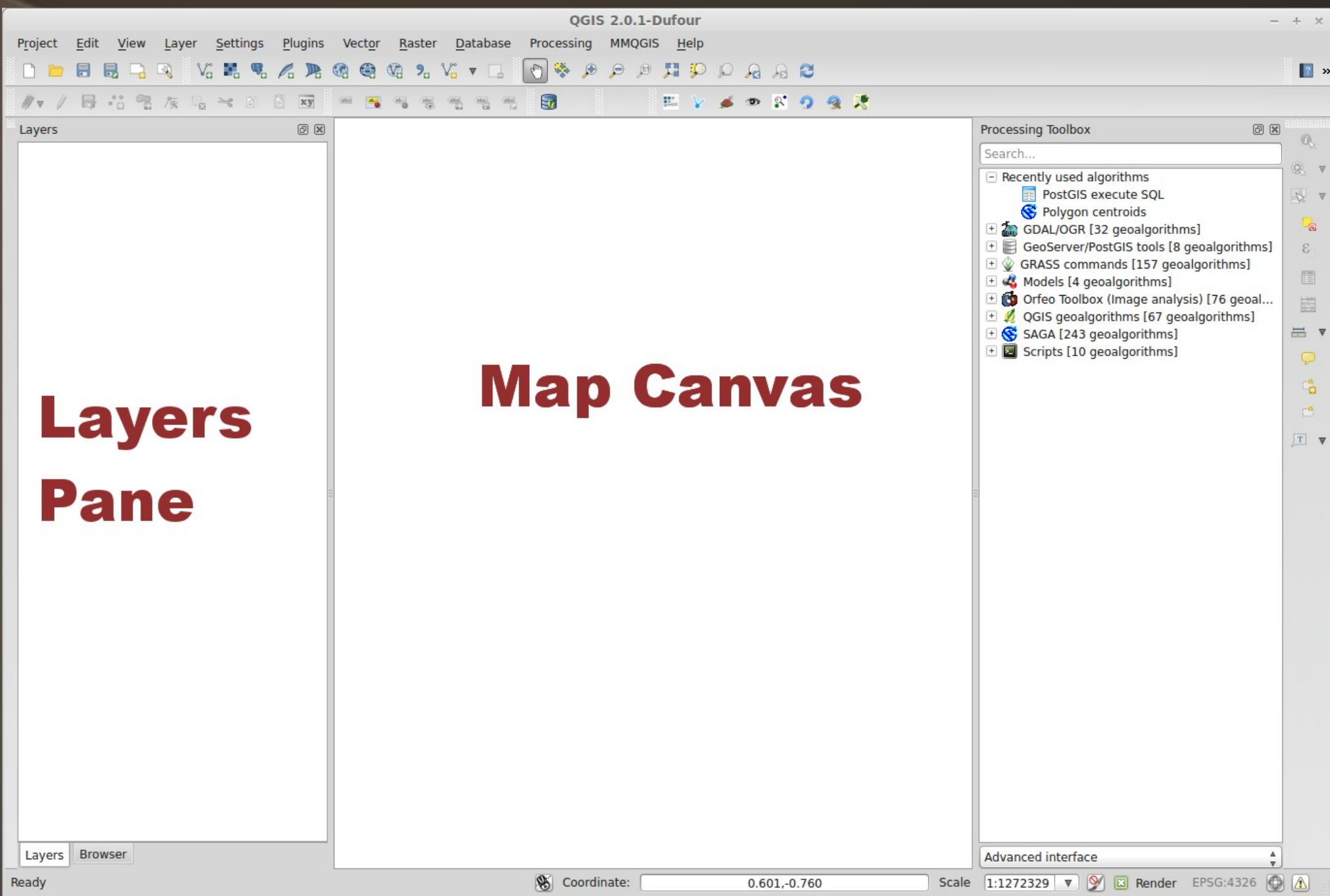
Partitioning Performance

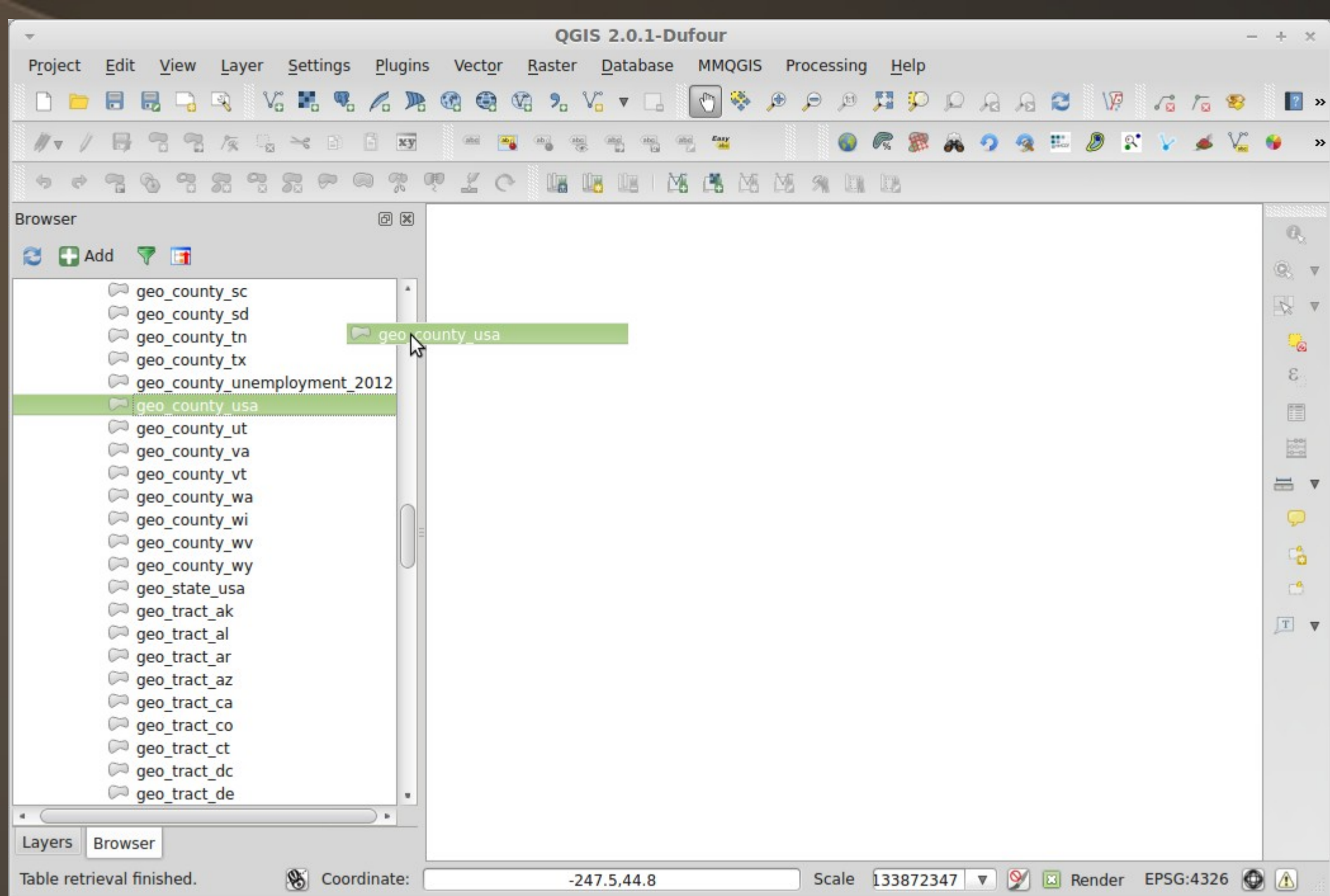
Measure	One table per sequence	Sequence partitioned by state
Black-White Index of Dissimilarity by State	4.9 seconds	4.8 seconds
Black-White Index of Dissimilarity by MSA	1.3 seconds	1.4 seconds
Theil Inequality by State	2.1 seconds	12.7 seconds
Theil Inequality by MSA	1.3 seconds	2.6 seconds

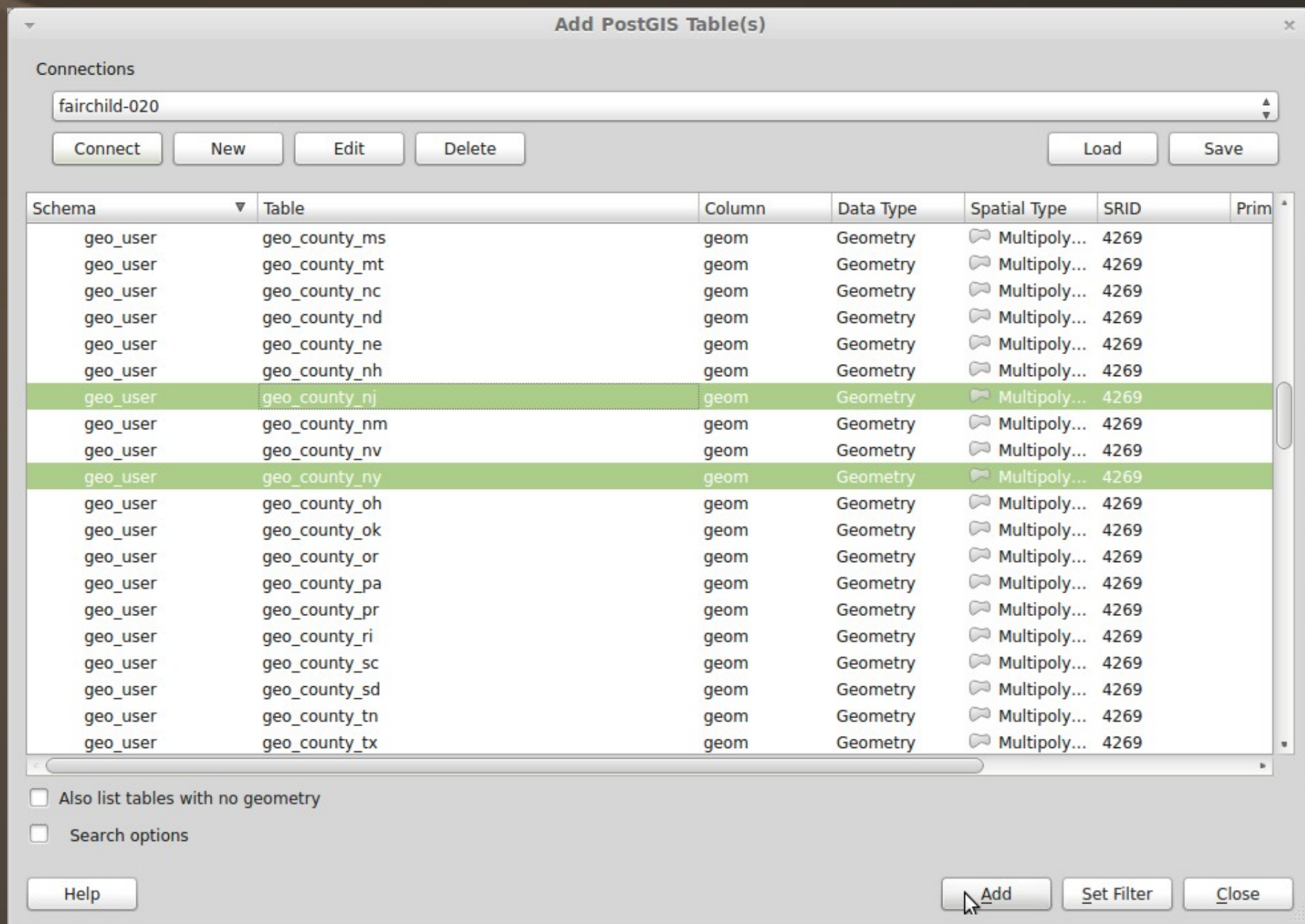
- Surprisingly, partitioning did not help performance (and occasionally hurt)
- Additionally, large number of DB objects slows down some clients when object list first loads

Geovisualization









Comparison of Browser vs. Add Dialog

- Data Browser
 - Loads list of spatial layers faster
 - Only loads list of spatial layers once per session
- Add PostGIS Table Dialog
 - Can add aspatial layers
 - Can query layer prior to adding

*Comparison of Data Loading Speeds— Tracts with Attached Attributes**

Operation	All tracts in one layer	Tracts partitioned by state
Load all tracts in country	5:40 minutes	7:15 minutes
Load all tracts in child table (MA)	NA	6 seconds
Load all tracts in one state by query (MA)	6 seconds	6 seconds
Load all tracts in six child tables (New England)	NA	15 seconds
Load all tracts in six states by query (New England)	12 seconds	1:10 minutes

* Basically, the TIGER/Line file without any additional demographic data

*Comparison of Data Loading Speeds— Tracts Joined to Subject Table B08006**

Operation	All tracts in one layer	Tracts layer partitioned by state
Load all tracts in country	6:42 minutes	5:35 minutes
Load all tracts in one state by query (MA)	8 seconds	9 seconds
Load all tracts in six states by query (New England)	16 seconds	16 seconds

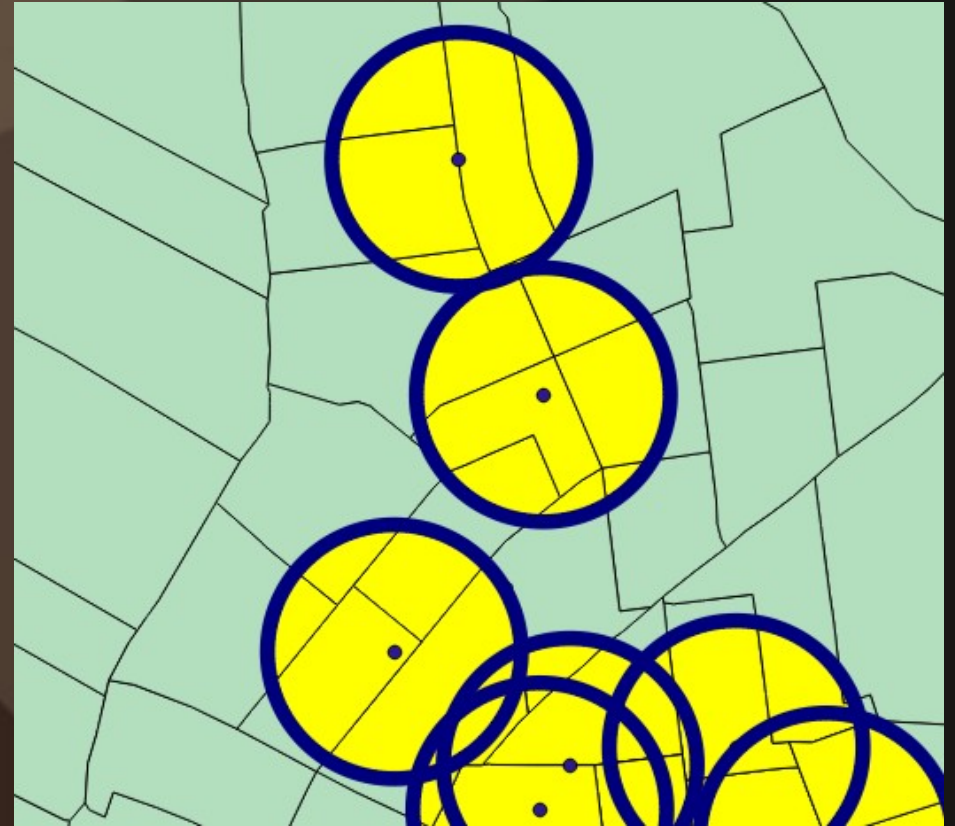
* Sex Of Workers By Means Of Transportation To Work

Approaches to Attribute Joins

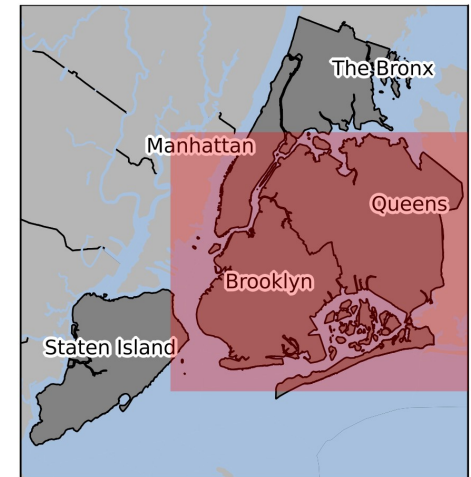
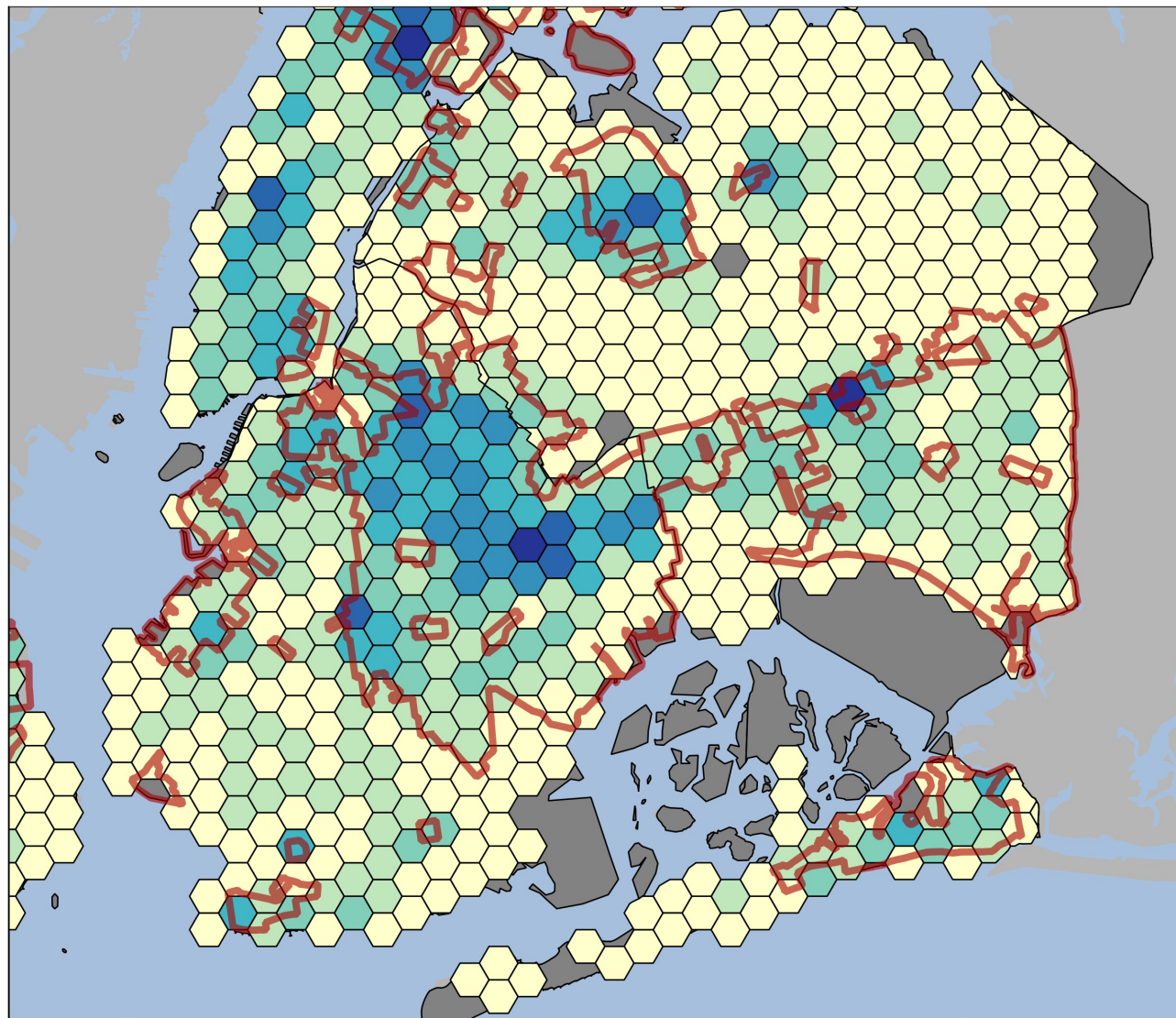
- Permanently join (new columns or new layer)
- Spatial view
- Materialized spatial view
- Ad hoc query in DB Manager

Fun PostGIS Tricks

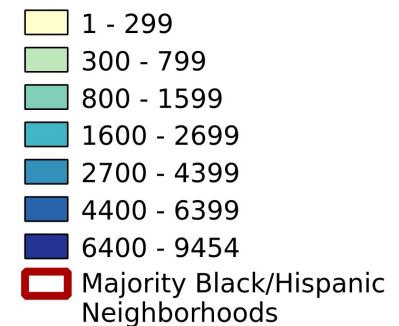
- Store multiple representations of the same geography in a single layer
- Modify geometries without having to materialize the layer (e.g., centroids, buffers)



NYPD Total Annual Stops, Brooklyn/Queens, New York, 2011



Reported Stops, 2011

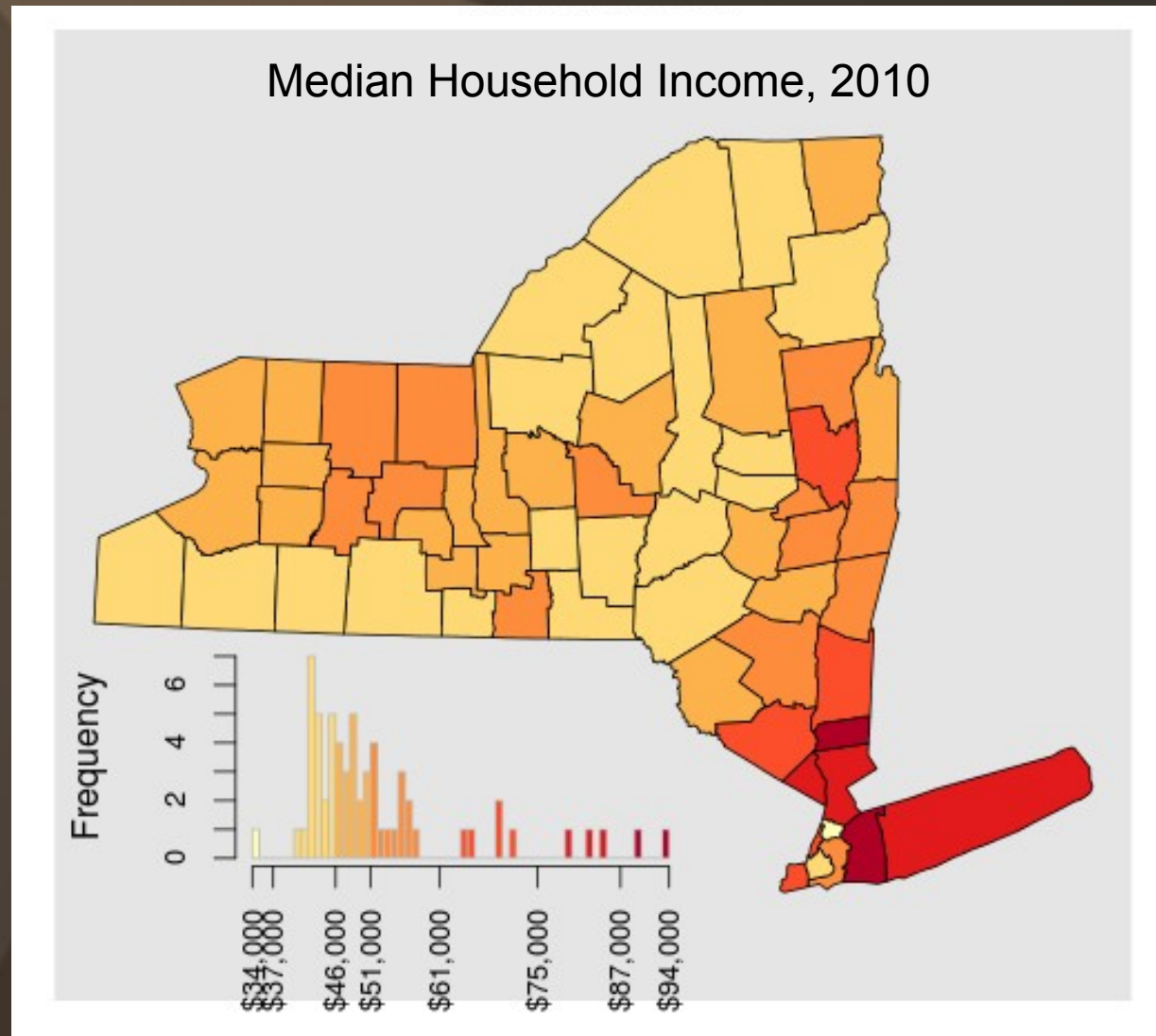


Data: New York Civil Liberties Union,
<http://www.nyclu.org/content/stop-and-frisk-data>



CC BY-ND 2013, Lee Hachadoorian
Some Rights Reserved

Integrating with Other Software



GitHub repository page for **leehach / census-postgres**.

Repository description: PostgreSQL schema and import scripts for recent US Census data — Edit

Repository statistics: 41 commits, 1 branch, 0 releases, 1 contributor.

Branch: master

Commit history:

Commit	Description	Time ago
leehach authored Thursday at 3:15am	Updated acs2011_5yr scripts to create sequences with geoid field and ...	latest commit 11328be6fc
acs2010_5yr	Added single view which returns all tracts in MSAs with associated MS...	9 days ago
acs2010_spt	Added state-specific import scripts to acs2010_5yr; Cleanup of delete...	2 years ago
acs2011_5yr	Updated acs2011_5yr scripts to create sequences with geoid field and ...	2 days ago
acs2012_1yr	Added data dictionary of ACS 2012 1-year, will allow meta-script import.	7 months ago
meta-scripts	sql_insert_into_tables now adds geoid value on table load. This appro...	2 days ago
postgis_support	sql_insert_into_tables now adds geoid value on table load. This appro...	2 days ago
.gitignore	Minor changes to experimental (array and hstore) storage methods	7 months ago

Right sidebar navigation: Code, Issues (0), Pull Requests (0), Wiki, Pulse, Graphs, Network, Settings.

HTTPS clone URL:

Thank you!

Lee.Hachadoorian@gmail.com

<http://github.com/leehach/census-postgres>

<http://geospatial.commonsgc.cuny.edu/>