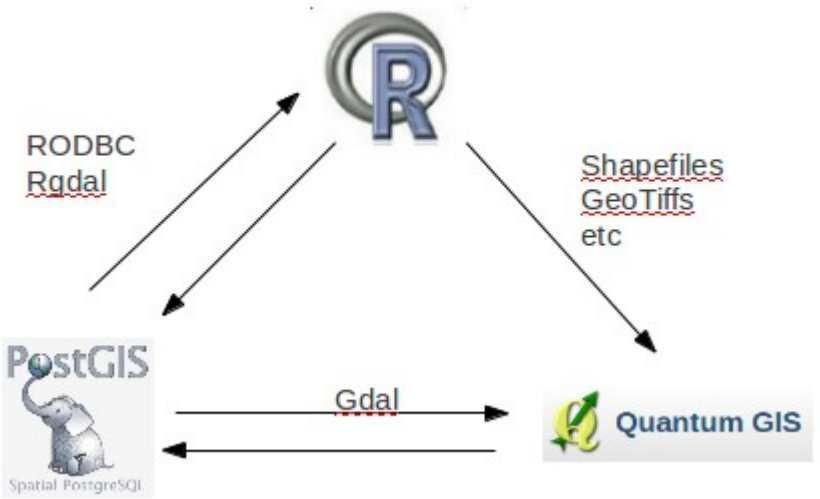
Linking PostGIS and R

Duncan Golicher



ECOSUR
El Colegio de la Frontera Sur
Bournemouth University

What is PostGIS?

- Set of spatial functions added as an extension to Postgresql
- Used for storing and processing very large complex data sets involving multiple users eg
  - Canadian Forest Resources Inventories

    Vectorborne Diseases, UC Davis
- Originally only for vector data

Strengths of spatial data bases

- Structured data storage
- Data sharing
- Slicing, dicing and combining data
- Efficient use of networked resources
- Real time space-time analysis (for example animal tracking etc)
- Online mapping applications

PostGIS and R

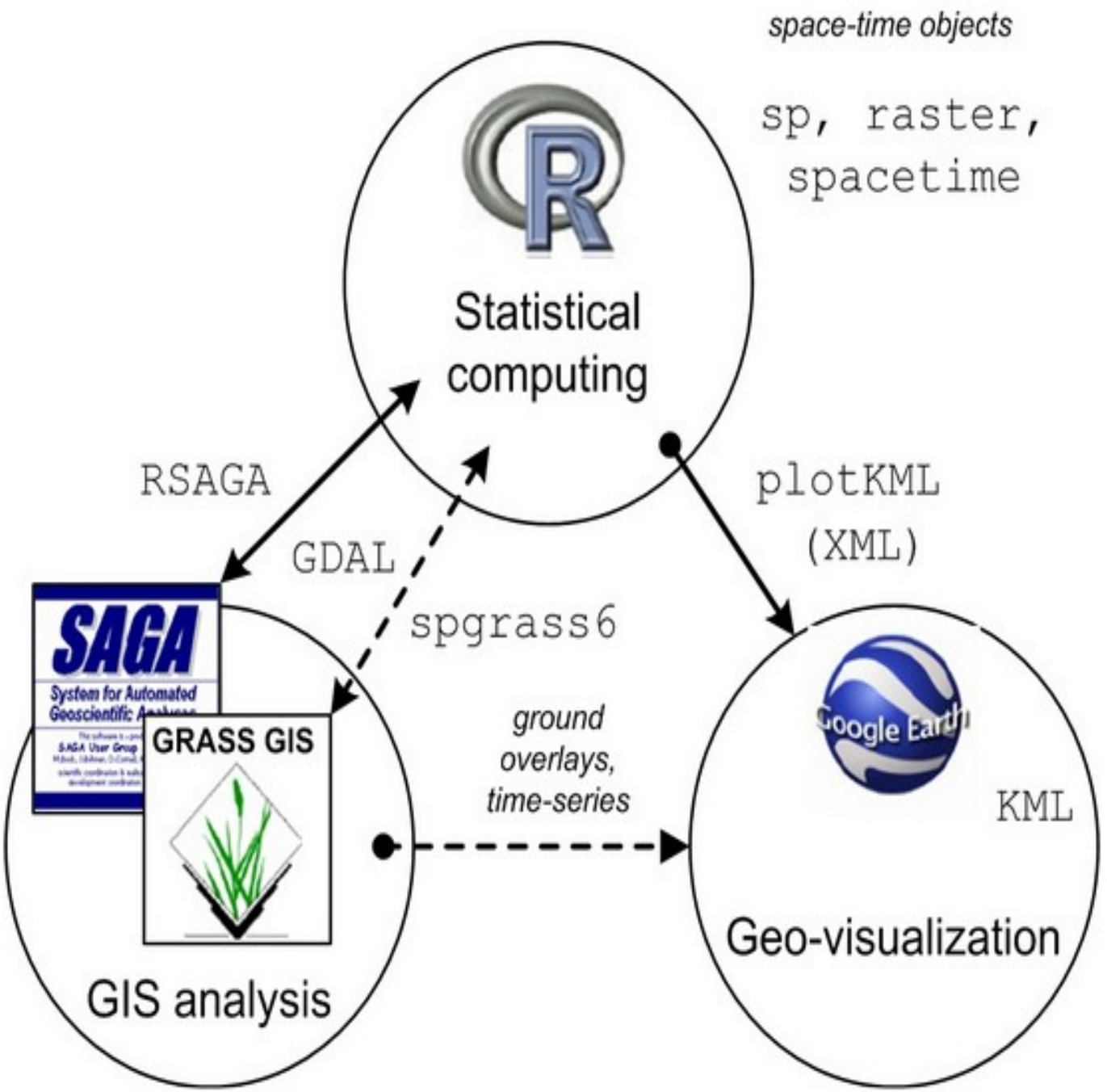
- PostGIS can be installed locally in order to allow R to use spatial SQL
- Particularly useful for querying vector data
- PostGIS has gained raster support relatively recently, thanks mainly to Pierre Racine and Bborie Park (UC Davis)
- Raster processing now integrated within PostGIS installs for all platforms

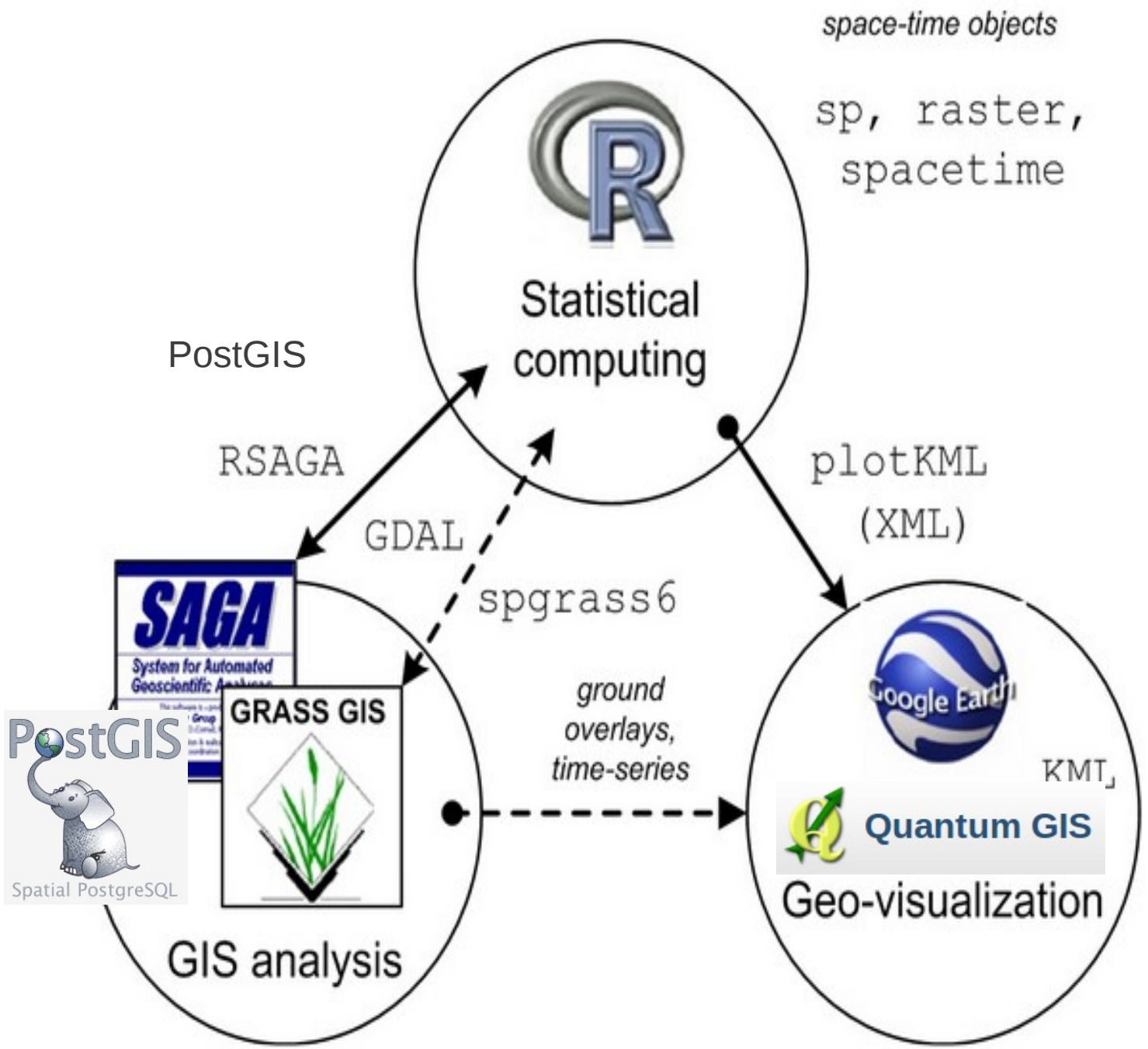
PostGIS in our research

- We are currently modelling the distribution of over 2000 species of Mexican trees for Conabio
- PostGIS allows us to combine diverse sources of spatial and non spatial information
- Useful for “slicing and dicing” data before import to R
- Spatial queries in PostGIS involve vector data.
- Most of our raster operations use dismo and the raster package
- However … We have experimented with PostGIS raster, particularly for overlays involving polygons

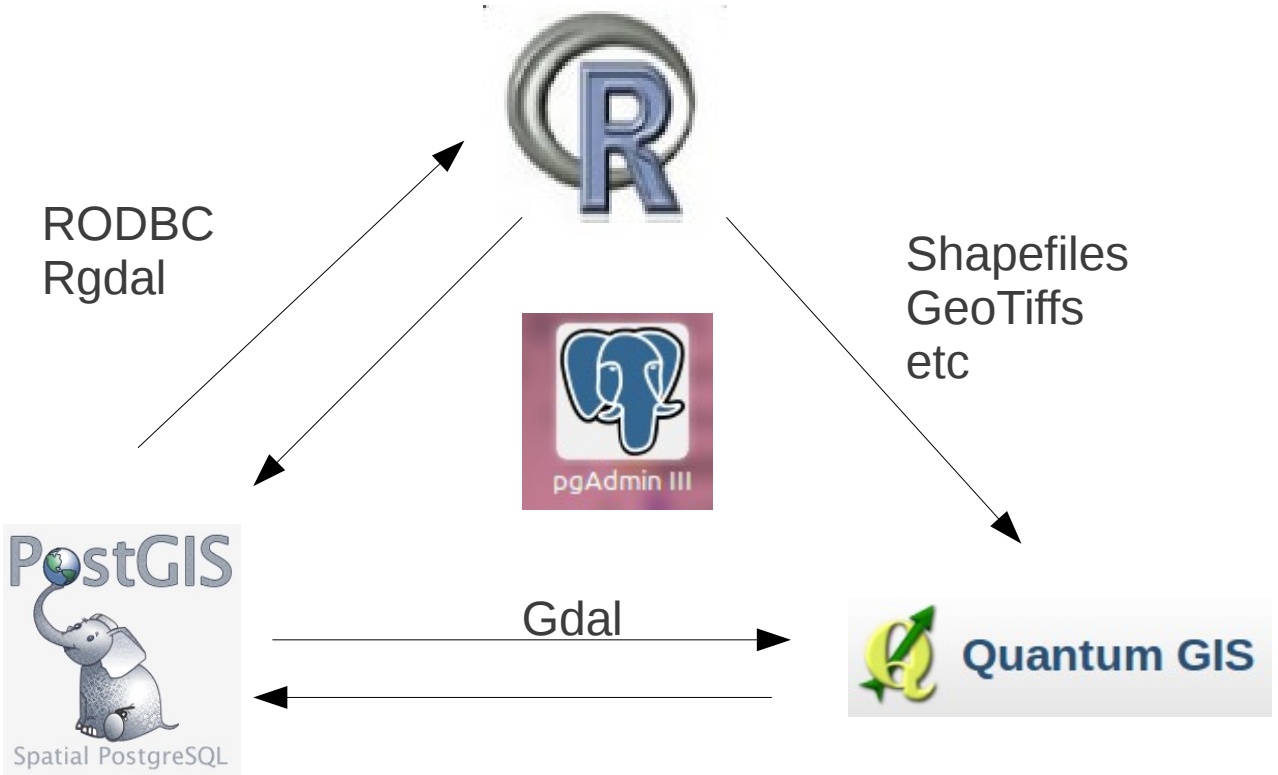
Tutorial

- The tutorial for this course includes a small trial data set.
- http://tinyurl.com/QuebecPostGIS
- Can be run with only minor modification (setting paths) on Ubuntu
- Contains code to install PostGIS from source on Ubuntu 12.04
- Demonstrates some useful features of PostGIS as a complement to other GIS functionality in R


space-time objects
sp, raster,
spacetime
Statistical
computing
RSAGA
GDAL
spgrass6
plotKML
(XML)
SAGA
System for Automated
Geoscientific
GRASS GIS
ground
overlays,
time-series
Google Earth
KML
GIS analysis
Geo-visualization


space-time objects
sp, raster,
spacetime
Statistical
computing
PostGIS
RSAGA
GDAL
spgrass6
plotKML
(XML)
SAGA
System for Automated
GRASS GIS
PostGIS
Spatial PostgreSQL
ground
overlays,
time-series
Google Earth
KML
Quantum GIS
GIS analysis
Geo-visualization

Communicating between R and PostGIS



Installing PostGIS locally

- Install is very easy on either Windows and Ubuntu
- Windows users follow a simple graphical online setup procedure
- A basic setup on Ubuntu in four easy steps
  - Install from repository
  - Set up user passwords and privileges
  - Create a database
  - Add the PostGIS extension

Installing PostGIS on Ubuntu

```
sudo apt-add-repository ppa:sharpie/postgis-nightly
sudo apt-get update
sudo apt-get install postgresql-9.1-postgis
sudo apt-get install postgresql-9.1-plr

sudo passwd postgres
sudo -u postgres psql -c "alter user postgres with password
'postgres';"
sudo -u postgres createuser [duncan] --superuser

 createdb geostats

 psql -d geostats -c 'create extension postgis
 psql -d geostats -c 'create extension plr''
```

Set passwords for the OS user and within the database

Building from source

```
sudo apt-get install libgdal.dev libproj.dev libxml2-dev libgeos-dev
sudo apt-get install libarmadillo-dev libpoppler-dev libepsilon-dev libexpat-dev liblzma-dev

wget http://postgis.net/stuff/postgis-2.1.0beta3dev.tar.gz
tar xvzf postgis-2.1*
cd postgis*
./configure
make
sudo make install

wget https://github.com/jconway/plr/archive/master.zip
unzip master.zip
 cd plr-master
 USE_PGXS=1 make
 sudo USE_PGXS=1 make install
```

RODBC

- Open data base connectivity
- Install the postgresql driver
- Windows

  http://ftp.postgresql.org/pub/odbc/versions/msi/psql
- Ubuntu

```
sudo apt-get install unixODBC unixODBC-dev
sudo apt-get install odbc-postgresql
```

Getting started

- Create a database
- Create the PostGIS extension

- system("createdb geostats")
  - CREATE EXTENSION POSTGIS
  - CREATE EXTENSION PLR

ODBC

- Allows R to “speak” PostGIS
- In Linux connections saved in etc/odbc.ini

```
[geostats]
Driver = /usr/lib/i386-linux-gnu/odbc/psqlodbcw.so
Database = geostats
Servername = localhost
Username = postgres
Password = postgres
Protocol = 8.2.5
ReadOnly = 0
```

ODBC

```
con <- odbcConnect("geostats")
query<-”select some stuff from ….”
RDataFrame<-sqlQuery(con,query)
```

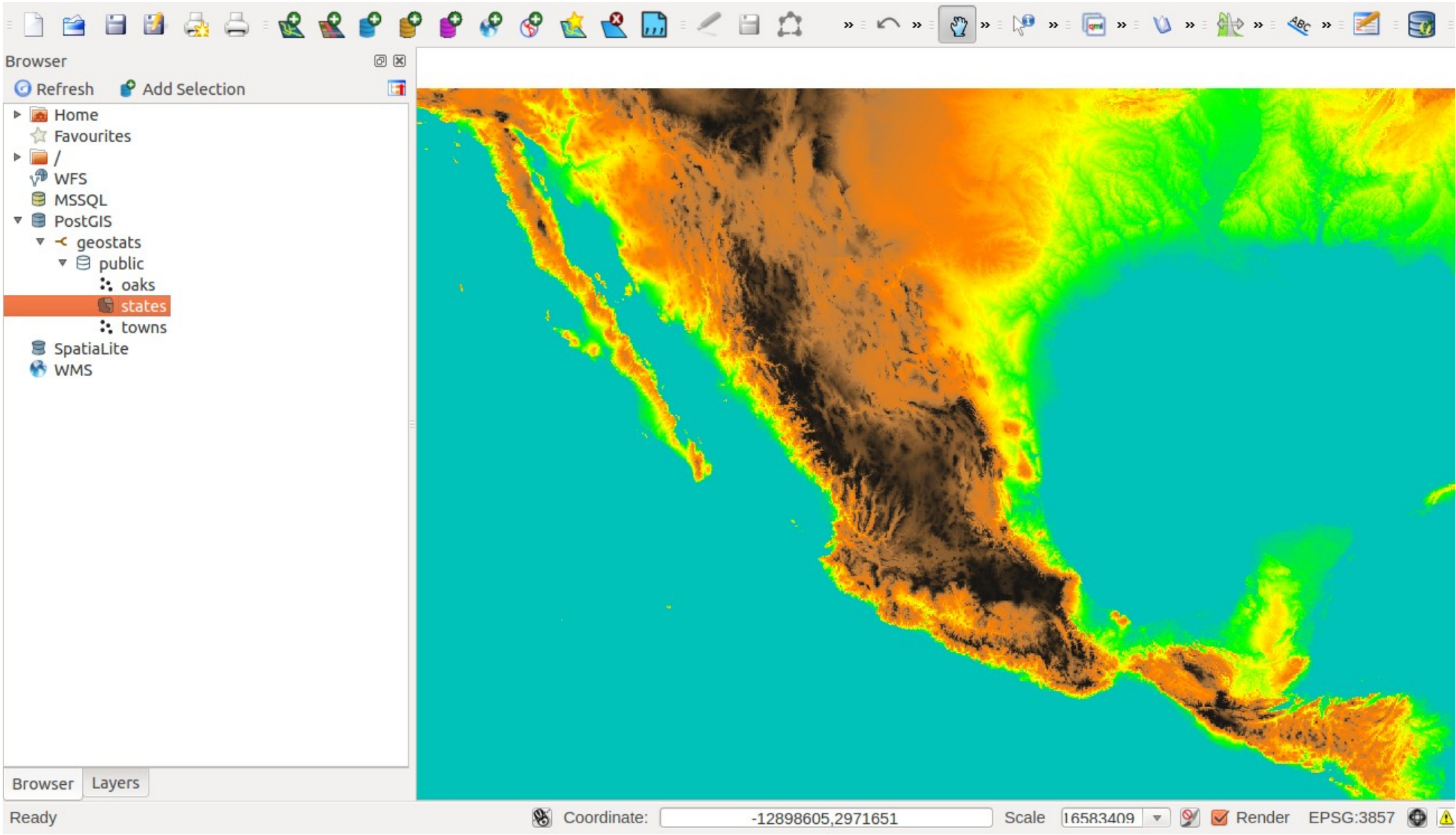
Rgdal

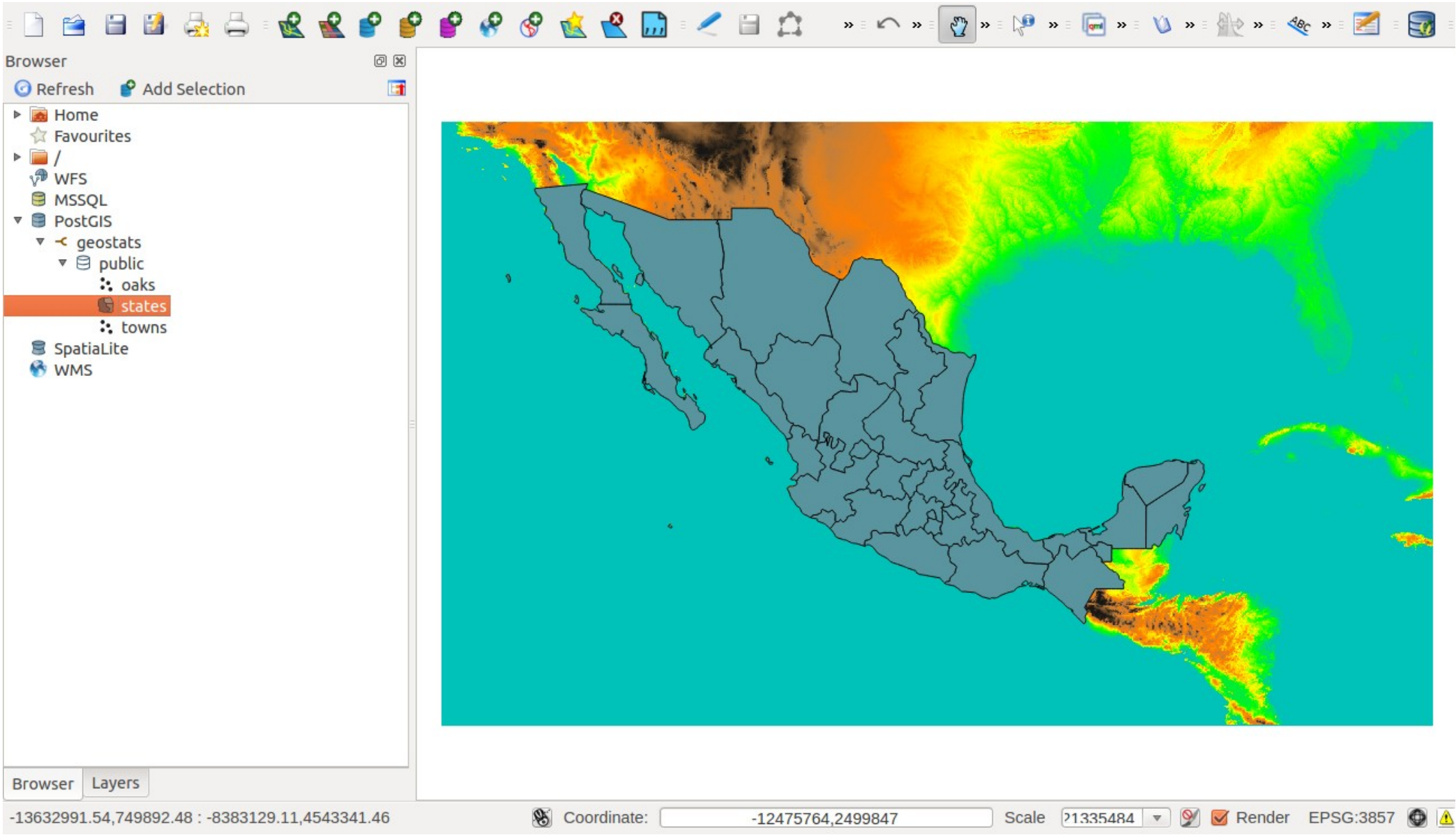
- Allows R to read and write spatial data using gdal drivers
- Can read and write from shapefiles and PostGIS
- Complements RODBC

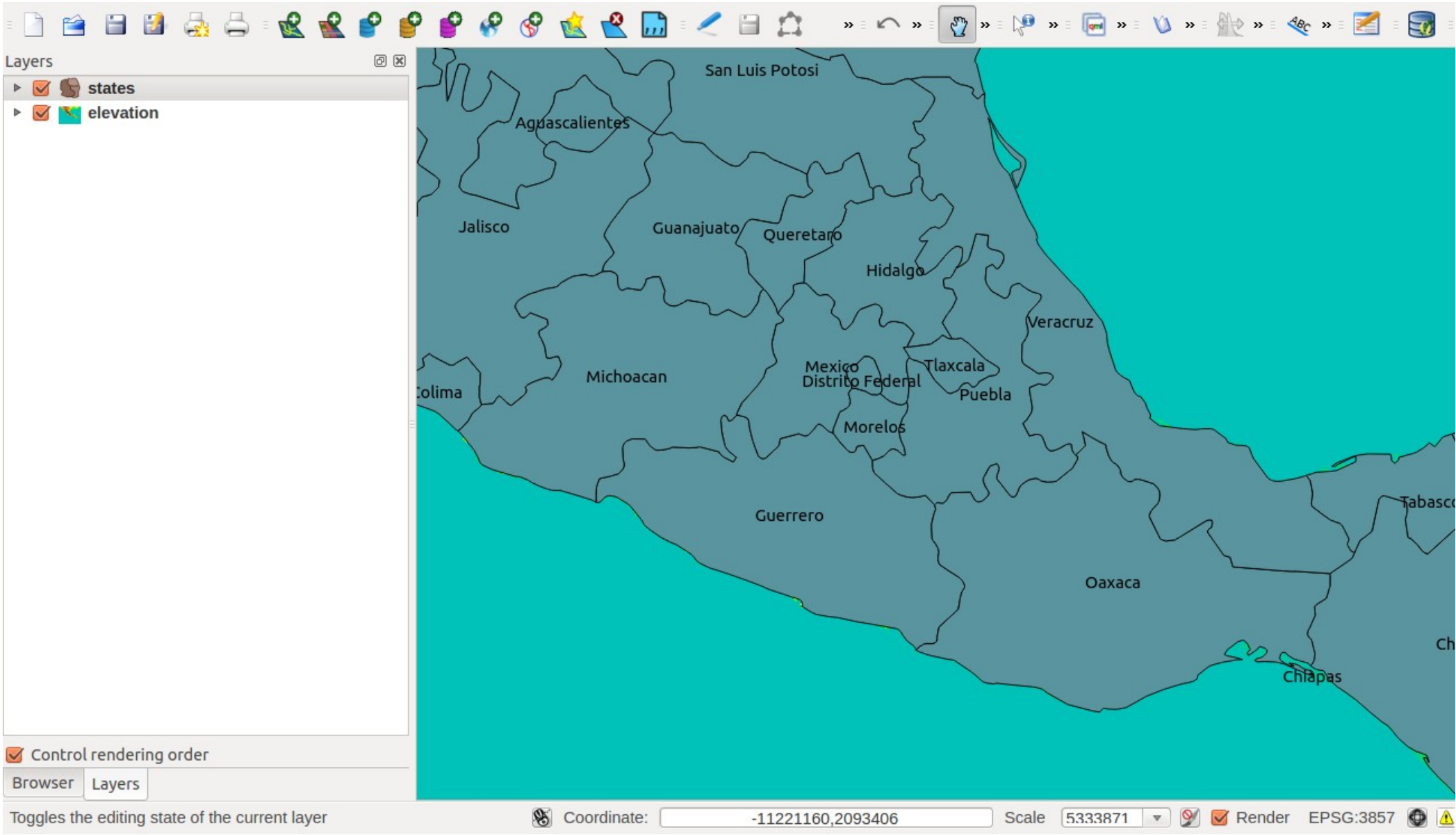
```
>states<-readOGR("shapefiles","states")

>writeOGR(states,"PG:dbname=geostats",
layer_options="geometry_name=geom","states",
"PostgreSQL")

>readOGR("PG:dbname=geostats", "states")
```

QGIS

- Visualisation tool and GUI based GIS
- PostGIS layers can be added simply by dragging and dropping


Browser
Refresh
Add Selection
Home
Favourites
/
WFS
MSSQL
PostGIS
geostats
public
oaks
states
towns
SpatiaLite
WMS
Browser
Layers
Ready
Coordinate:
-12898605,2971651
Scale
16583409
Render
EPSG:3857

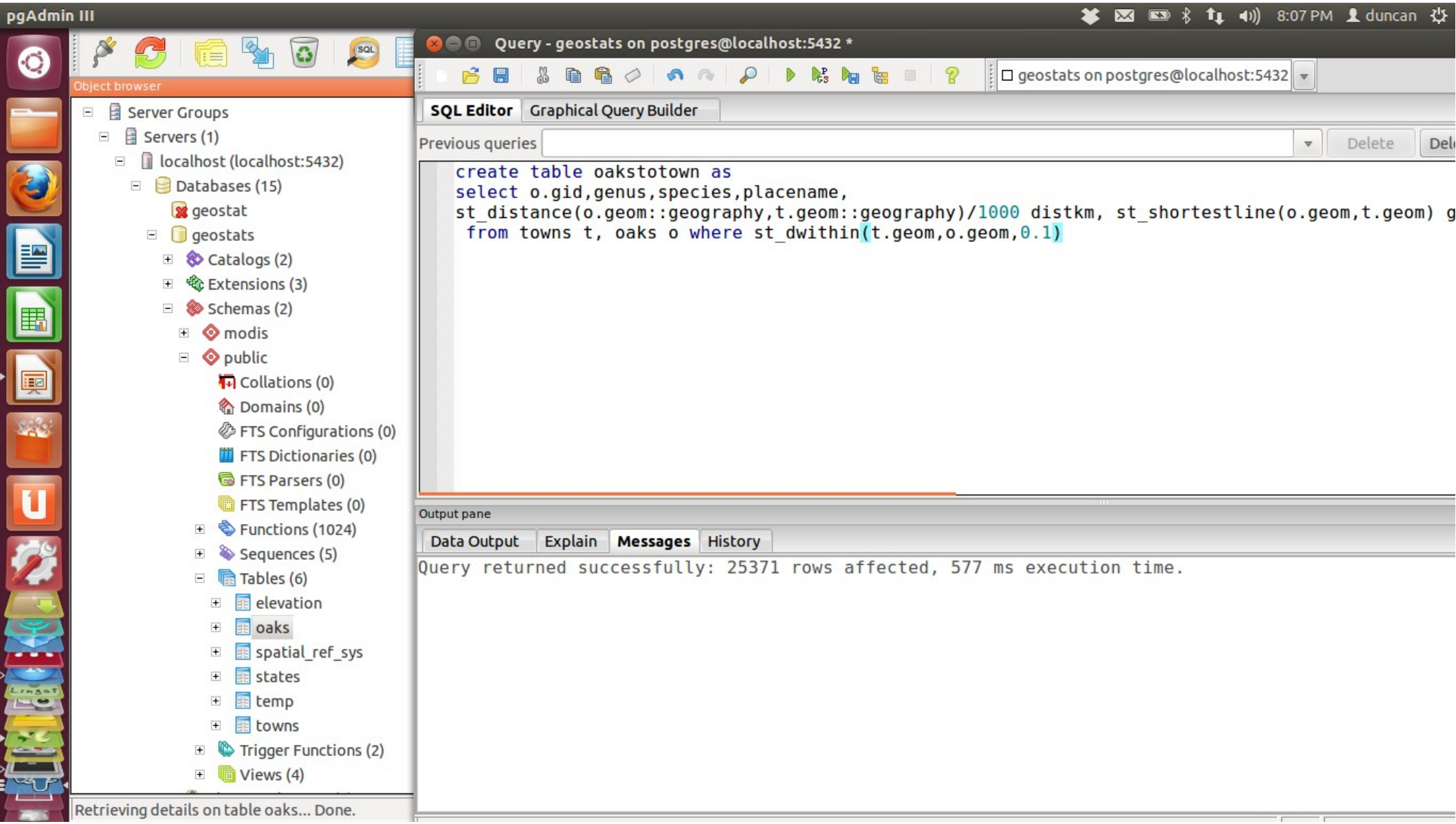

Browser
Refresh
Add Selection
Home
Favourites
/
WFS
MSSQL
PostGIS
geostats
public
oaks
states
towns
SpatiaLite
WMS
Browser
Layers
-13632991.54,749892.48 : -8383129.11,4543341.46
Coordinate:
-12475764,2499847
Scale
?1335484
Render
EPSG:3857


Layers
states
elevation
Control rendering order
Browser
Layers
San Luis Potosi
Aguascalientes
Jalisco
Guanajuato
Queretaro
Hidalgo
Veracruz
Michoacan
Colima
Mexico
Distrito Federal
Tlaxcala
Puebla
Morelos
Guerrero
Oaxaca
Chiapas
Toggles the editing state of the current layer
Coordinate:
-11221160,2093406
Scale
5333871
Render
EPSG:3857

PGAdmin

- Powerful graphical admin tool for Postgresql
- Can be used to design and test queries before integration into R scripts

PgAdmin

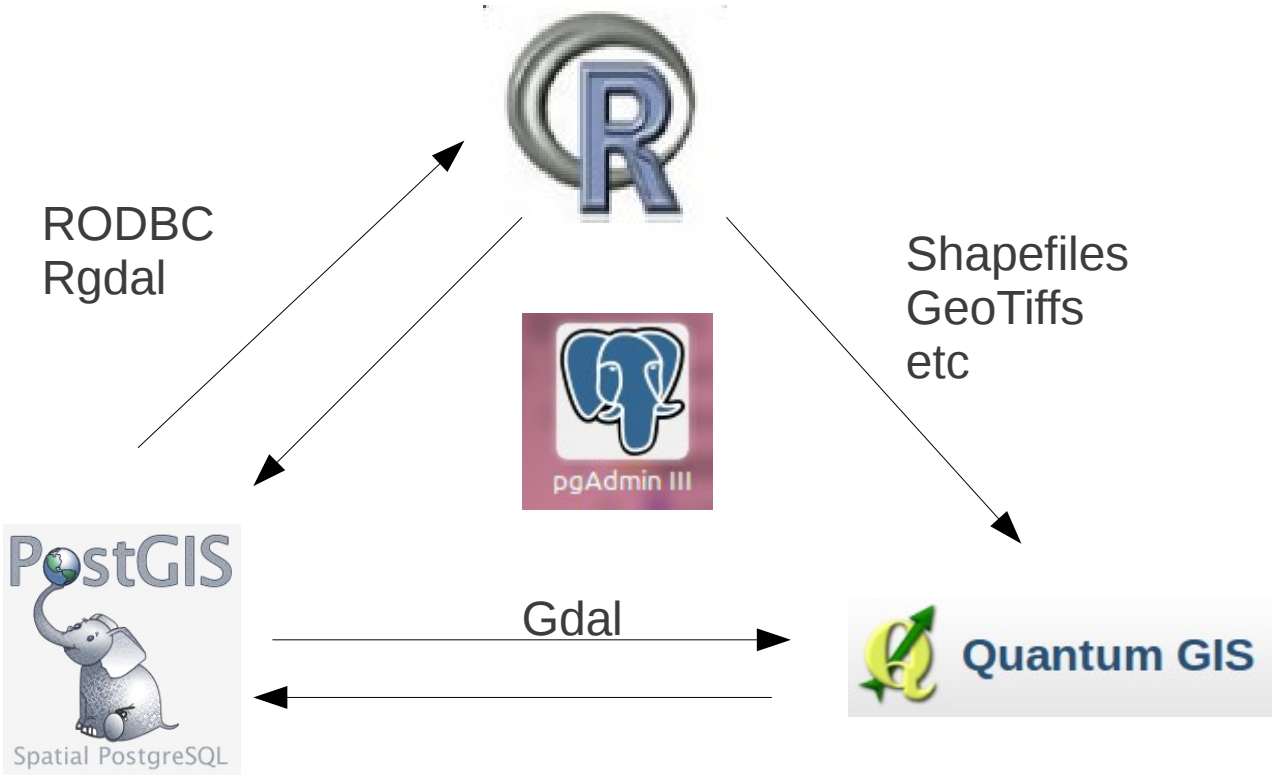

pgAdmin III
8:07 PM duncan
Object browser
Server Groups
Servers (1)
localhost (localhost:5432)
Databases (15)
geostat
geostats
Catalogs (2)
Extensions (3)
Schemas (2)
modis
public
Collations (0)
Domains (0)
FTS Configurations (0)
FTS Dictionaries (0)
FTS Parsers (0)
FTS Templates (0)
Functions (1024)
Sequences (5)
Tables (6)
elevation
oaks
spatial_ref_sys
states
temp
towns
Trigger Functions (2)
Views (4)
Retrieving details on table oaks... Done.
Query - geostats on postgres@localhost:5432 *
geostats on postgres@localhost:5432
SQL Editor
Graphical Query Builder
Previous queries
Delete
create table oakstotown as
select o.gid,genus,species,placename,
st_distance(o.geom::geography,t.geom::geography)/1000 distkm, st_shortestline(o.geom,t.geom) g
from towns t, oaks o where st_dwithin(t.geom,o.geom,0.1)
Output pane
Data Output
Explain
Messages
History
Query returned successfully: 25371 rows affected, 577 ms execution time.

System commands

- Vector and raster data can be loaded from the command line
- R scripts can be used to control the process

```
 command<-paste("raster2pgsql -s 4326 -d
-M -R ",f," -F -t 100x100 temp|psql -d
geostats",sep="")
  print(command)
```

Putting it all together



Example in RStudio