Analysis of Big Spatial Data with PostgreSQL/PostGIS and R – Case Studies in OpenStreetMap and Interactive Web Mapping from R

Dennis Zielstra, Ph. D. – GIS Administrator
Francesco Tonini, Ph. D. – Post-doctorate Researcher

North Carolina State University
Geospatial Analytics Forum
02/12/2015
PostgreSQL / PostGIS

- Open source, freely available, “object-relational database management system” (ORDBMS) since 1995
- Supports Windows, Mac OS X and Linux systems
PostgreSQL / PostGIS

PostgreSQL (short: Postgres)

– Open source, freely available, "object-relational database management system" (ORDBMS) since 1995

– Supports Windows, Mac OS X and Linux systems
PostgreSQL / PostGIS

PostgreSQL (short: Postgres) ([http://www.postgresql.org](http://www.postgresql.org)) – Open source, freely available, "object-relational database management system" (ORDBMS) since 1995 – Supports Windows, Mac OS X and Linux systems
PostgreSQL / PostGIS

- PostGIS (http://postgis.net/)
- Open Source, freely available, and fairly OGC compliant spatial database extender for Postgres
- Similar in functionality to SQL Server 2008 Spatial, ESRI ArcSDE, Oracle Spatial, and DB2 spatial extender.
- http://postgis.refractions.net/docs/reference.html
PostgreSQL / PostGIS

- PostGIS (http://postgis.net)
- Open Source, freely available, and fairly OGC compliant spatial database extender for Postgres
- Similar in functionality to SQL Server 2008 Spatial, ESRI ArcSDE, Oracle Spatial, and DB2 spatial extender.
- http://postgis.refractions.net/docs/reference.html
OpenStreetMap

User Generated Content

Volunteered Geographic Information
OpenStreetMap (OSM)

1.9 million members

3.5 billion Track points
OpenStreetMap (OSM)
OpenStreetMap vs. Proprietary Data
OpenStreetMap (OSM)

- Example: Cycling
- [http://www.opencyclemap.org/](http://www.opencyclemap.org/)
OpenStreetMap (OSM)

- Example: Cycling
- http://www.opencyclemap.org/
OpenStreetMap (OSM)

- Example: Winter sports
- http://openpistemap.org/
OpenStreetMap (OSM)

- Example: Winter sports
- http://openpistemap.org/
OpenStreetMap

- OSM and Emergency response
- HOT (Humanitarian OpenStreetMap Team)

Before earthquake

After earthquake
OpenStreetMap

- OSM and Emergency response
- HOT (Humanitarian OpenStreetMap Team)
How do we analyze these large OSM (or other) datasets without visualizing them in traditional GIS applications?

1. Import data into PostgreSQL: Shp2pgsql and PSQL

```
shp2pgsql -s 2237 TOWNS_POLY towns > towns.sql
```

```
psql -d gisdb -h localhost -U postgres -f towns.sql
```

Function: Shp2pgsql and PSQL

<SRID> Spatial reference identifier

Output file name

Shapefile

Table name

Function: PSQL

Database

Server/Host

Username

File created in previous step
Processing Data in PostgreSQL

– How do we analyze these large OSM (or other) datasets without visualizing them in traditional GIS applications?
– 1. Import data into PostgreSQL: Shp2pgsql and PSQL
Processing Data in PostgreSQL

2. Run spatial queries to analyze data
   – Union to create one polygon from individual States
     SELECT ST_Union(the_geom) FROM USMap;

   – Length of roads in km
     SELECT sum(ST_Length(the_geom))/1000 AS km_roads FROM bc_roads;

   – Length of roads fully contained within each municipality
     SELECT m.name,
            sum(ST_Length(r.the_geom))/1000 as roads_km
     FROM bc_roads AS r,
            bc_municipality AS m
     WHERE ST_Contains(m.the_geom,r.the_geom)
     GROUP BY m.name
     ORDER BY roads_km;
Processing Data in PostgreSQL

2. Run spatial queries to analyze data
Processing Data in PostgreSQL

3. Create Shapefile from Table or direct connect with QGIS:

- `pgsql2shp`

  `pgsql2shp -f "/path" -h server -u user -P password mygisdb public.streets`

- Combine with SQL Query:

  `pgsql2shp -f "/path" -h server -u user -P password mygisdb "SELECT neigh_name, the_geom FROM neighborhoods WHERE neigh_name = 'Jamaica Plain'"`
Processing Data in PostgreSQL

3. Create Shapefile from Table or direct connect with QGIS:
Pedestrian Road Network Comparison (Completeness)

- Comparison of proprietary and OSM data

Zielstra & Hochmair (2012)
OSM Completeness after TIGER Import

- Strong pedestrian data collection efforts after TIGER import
- Network for motorized traffic incomplete and erroneous

Zielstra & Hochmair (2013)
The R software for Statistical Computing

www.r-project.org
The R software for Statistical Computing

Companies using R:

...and many more!
The R software for Statistical Computing

Some PROS:
- Free, open source ([GNU General Public License](http://www.gnu.org/licenses/gpl.html)), cross-platform
- Flexible & extensible
- Most powerful statistical programming language
- Huge community, brilliant developers (1800+ hosted projects on R-Forge, 6000+ packages on CRAN)
- Improved geospatial data handling

Some CONS:
- Not as efficient and fast compared to lower-level languages (C/C++ 100-1000x faster, Python 2-10x faster)
- Syntax sometimes not as “user-friendly”
The R software for Statistical Computing

R’s popularity is growing fast!:
blog.revolutionanalytics.com/popularity
The R software for Statistical Computing

http://www.revolutionanalytics.com

The Revolution R product suite:

- **Revolution R Open**
  - Free and open source R distribution
  - Enhanced and distributed by Revolution Analytics

- **Revolution R Plus**
  - Open-source distribution of R, packages, and other components
  - Enhanced, supported and indemnified by Revolution Analytics

- **Revolution R Enterprise**
  - Secure, Scalable and Supported Distribution of R
  - With proprietary components created by Revolution Analytics
The R software for Statistical Computing

Open Source at Revolution Analytics:

- **Revolution R Open**: Revolution R Open is an enhanced distribution of open source R from Revolution Analytics.
- **DeployR Open**: DeployR offers simple, secure R integration for application developers.
- **Reproducible R Toolkit**: RRT provides tools to ensure that the results of R code are repeatable over time by anyone.
- **RHadoop**: RHadoop is a collection of packages for connecting R to Hadoop and running R on Hadoop nodes.
- **ParallelR**: ParallelR is a collection of R packages for parallel and distributed programming.

Revolution R Open (RRO) is the enhanced distribution of R from Revolution Analytics:

- Multi-threaded math libraries that brings multi-threaded computations to R.
- A high-performance default CRAN repository
- The Reproducible R Toolkit that make it easy to share R code and replicate results
The R software for Statistical Computing

Revolution R Enterprise (RRE)
The All-Inclusive Big Data Big Analytics Platform

Free Revolution R Enterprise Academic Edition!
Is it possible to transfer geographic data from PostGIS to R?

(a) Use RODBC + rgdal R packages to move data to/from PostGIS

(b) Use QGIS as a visualization tool for PostGIS

OR Use R package ‘RPostgreSQL’

```
# Load data from the PostGIS server
conn = dbConnect(
    dbDriver("PostgreSQL"), dbname=dbname, host=host, port=5432,
    user=user, password=password)

strSQL = "SELECT gid, ST_AsText(geom) AS wkt_geometry, attr1,
          attr2[, ..] FROM geo_layer"

dfTemp = dbGetQuery(conn, strSQL)
row.names(dfTemp) = dfTemp$gid
```
Using R with GIS: useful resources

**WEB:**
- CRAN TASK VIEW: [http://cran.r-project.org/web/views/](http://cran.r-project.org/web/views/)

**BOOKS:**

![Image of books]
Spatial Classes in R

- Spatial Points Diagram (class name = blue, slots = white, inheritance = arrows)

- Spatial Polygons & Lines Classes
Spatial Classes in R

- Spatial Grid and Pixels Classes/Slots

```
SpatialGridDataFrame
  SpatialGrid
    data
    grid
      grid.index
      SpatialPoints

SpatialPixelsDataFrame
  SpatialPixels
    data
    grid
      grid.index
      SpatialPoints

GridTopology
  cellcentre.offset
  cellsize
  cells.dim

Spatial
  bbox
  proj4string

SpatialPoints
  coords
  Spatial

```

- Recap

```
data type  class                  attributes      extends
points      SpatialPoints         none            Spatial
points      SpatialPointsDataFrame data.frame     SpatialPoints
pixels      SpatialPixels         none            SpatialPoints
pixels      SpatialPixelsDataFrame data.frame     SpatialPixels
full grid   SpatialGrid          none            Spatial
full grid   SpatialGridDataFrame data.frame     SpatialGrid
line        Line                  none            Line list
lines       Lines                 none            Line list
lines       SpatialLines         none            Spatial, Lines list
lines       SpatialLinesDataFrame data.frame     SpatialLines
polygon     Polygon               none            Polygon list
polygons    Polylines            none            Polygon, Polylines list
polygons    SpatialPolygons      none            Spatial, Polylines list
polygons    SpatialPolygonsDataFrame data.frame SpatialPolygons
```
Using R with GIS

R + GIS analysis + Geo-visualization

- SAGA GIS: Use ‘RSAGA’ R package
- GRASS GIS: Use ‘spgrass6’ R package
- GOOGLE EARTH: Use ‘plotKML’ R package
Using R with (Arc)GIS

☞ R + ArcGIS:

[Image of Geospatial Modelling Environment]

http://www.spatialecology.com/gme/

☞ R + ArcGIS + Python:

[Image of R tool]

https://github.com/Esri/R-toolbox-py
R as a GIS

Some Basic Tools:

- library(maps) #for creating geographical maps
- library(maptools) #tools for handling spatial objects
- library(mapproj) #for creating projected maps
- library(raster) #tools to deal with raster maps
- library(rasterVis) #visualization of raster layers
- library(ggplot2) #to create maps
- library(gpclib) #general polygon clipper
- library(rgdal) #read/write GDAL raster/OGR vectors

...more
R as a GIS

(Very) Simple Maps:

Credits: http://www.nyu.edu/projects/politicsdatalab/workshops/GISwR.pdf
R as a GIS

More Involved Maps:

Credits: http://www.nyu.edu/projects/politicsdatalab/workshops/GISwR.pdf
R as a GIS

Google Maps: R package ‘RgoogleMaps’ to map your data onto Google Map tiles

Credits: http://spatial.ly

Credits: http://www.blogto.com/city/
Google Maps: R package `plotGoogleMaps`

Creating web map of point data (click hyperlink to .html files)
Google Maps: R package ‘plotGoogleMaps’

Creating web map of spatial polygons (click hyperlink to .html files)
Google Maps: R package ‘plotGoogleMaps’

Creating web map of spatial grid/pixels (click hyperlink to .html files)
Google Maps: R package ‘plotGoogleMaps’

- Combine several layers (click hyperlink to .html files)
R as a GIS

**OpenStreetMap**: R package `osmar` to work with OSM data

```
> library("osmar")
Loading required package: XML
Loading required package: RCurl
Loading required package: bitops
Loading required package: gtools
Loading required package: geosphere
Loading required package: sp
```
R as a GIS

Getting the data

- Retrieve elements by IDs:

  ```
  src <- osmsource_api()
  > get_osm(node(18961430), source = src)
  ```

- Retrieve elements by bounding box:

  ```
  > bb <- center_bbox(174.76778, -36.85056, 700, 700)
  > ua <- get_osm(bb, source = src)
  > ua
  osmar object
  2427 nodes, 428 ways, 70 relations
  ```

Working with the data

```
> summary(ua$nodes)
```

---

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>addr:city</td>
<td>Auckland</td>
<td>101</td>
</tr>
<tr>
<td>addr:street</td>
<td>Queen Street</td>
<td>61</td>
</tr>
<tr>
<td>addr:country</td>
<td>NZ</td>
<td>40</td>
</tr>
<tr>
<td>addr:postcode</td>
<td>1010</td>
<td>39</td>
</tr>
<tr>
<td>comment</td>
<td>Determined via Keypa...</td>
<td>29</td>
</tr>
<tr>
<td>addr:street</td>
<td>Symonds Street</td>
<td>27</td>
</tr>
<tr>
<td>highway</td>
<td>traffic_signals</td>
<td>23</td>
</tr>
<tr>
<td>addr:street</td>
<td>Lorne Street</td>
<td>19</td>
</tr>
<tr>
<td>highway</td>
<td>bus_stop</td>
<td>15</td>
</tr>
<tr>
<td>amenity</td>
<td>cafe</td>
<td>11</td>
</tr>
</tbody>
</table>
R as a GIS

Working with the data

Find specific elements:

```r
> ts_ids <- find(ua, node(tags(v == "traffic_signals")))
> hw_ids <- find(ua, way(tags(k == "highway")))
> bs_ids <- find(ua, node(tags(v %agrep% "busstop")))
```

Subset:

```r
> ts_ids <- find(ua, node(tags(v == "traffic_signals")))
> ts <- subset(ua, node_ids = ts_ids)
```

Plot:

```r
> plot(ua)
> plot_ways(hw, add = TRUE, col = "green")
> plot_nodes(ts, add = TRUE, col = "red")
> plot_nodes(bs, add = TRUE, col = "blue")
```
R as a GIS

Example

1) Getting the data

```r
src <- osamape::osmsource_api()
bb <- center_bbox(-1.53492, 53.81934, 1000, 1000)
ctown <- get_osm(bb, source = src)
plot(ctown)
points(-1.53492, 53.81934, col = "red", lwd = 5)
```

2) Working with the data

```r
res <- find(ctown, way(tags(k = "highway" & v == "residential")))
res <- find_down(ctown, way(res))
res <- subset(ctown, ids = res)
plot(ctown)
plot_ways(res, add = T, col = "green", lwd = 3)
```

3) Converting the data

```r
> hw_line <- as_sp(hw, "lines")
> bs_points <- as_sp(bs, "points")
```
R as a GIS

OpenStreetMap as simple background maps
R as a GIS

**ggplot2**: R package ‘ggplot2’

**London Cycle Hire Journeys**
Thicker, yellower lines mean more journeys

Credits: http://spatial.ly

Credits: http://spatialanalysis.co.uk/2013/02/mapped-twitter-languages-york/

New York’s Twitter Languages
500,000 Non-English tweets Jan 2010 - Feb 2013

Credits: http://spatialanalysis.co.uk/2013/02/mapped-twitter-languages-york/
R as a GIS

Stunning Maps: world topography by Robin Edwards at UCL CASA (London, UK)

Credits: http://spatial.ly
R as a GIS

Stunning Maps: world’s biggest airlines by James Cheshire at UCL CASA (London, UK)

Credits: http://spatial.ly
R as a GIS

Stunning Maps: visualizing FB friends by Paul Butler

Credits: http://paulbutler.org/archives/visualizing-facebook-friends/
R as a GIS

Stunning Maps: from home to work in the UK by James Cheshire & Oliver Uberti

From Home to Work
High paying jobs draw workers from far, far away

In this depiction of daily commutes, London shines like the Sun in the constellation of Southern England. Like all stars, it has an immense gravitational pull. Whether by car, train or tube, thousands travel into the capital each day from all directions, including this "commuter belt" beyond the Greater London Authority boundary. The capital city of the largest metropolis areas in the EU, with a population of more than 13 million.

Half of London's workforce make their journey by public transport, compared with only 9% in the rest of the country. Some travel thirty minutes or more to get to work. Elsewhere in the UK, only 20% have commutes that long. Why do so many live so far away?

For one, London salaries go further in satellite towns like Banbury. As of May 2014, a five-bedroom detached home there was going for the price of two-bedroom flats along the Underground's Central Line (see pp. 66-7). It's only a matter of time before faster trains propel commuters into ever wider orbits.

Credits: http://theinformationcapital.com/
(Interactive) Web Mapping
(Interactive) Web Mapping with R

**googleVis**: R package for visualization of data in a web browser (Google Visualization API)

[GoogleMap Example](http://cran.r-project.org/web/packages/googleVis/vignettes/googleVis_examples.html)
(Interactive) Web Mapping with R

**Leaflet Maps**: R package ‘leafletR’ R package

rMaps (still under development, can be installed from github using devtools) makes it easy to create, customize and share interactive maps from R, with a few lines of code. It supports several JavaScript based mapping libraries like Leaflet, DataMaps and Crosslet, with many more to be added.

Leaflet Heat Maps


Animated Choropleths

http://rmaps.github.io/blog/posts/animated-choropleths/index.html