Spatial Analysis Using R

Chieko Maene

University of Chicago

October 5, 2015

Who we are

- http://gis.uchicago.edu/
- Two-person GIS team at Social Sciences Division
- We (not the university IT) provide GIS software licenses and services
- We also offer GIS courses and workshop
- We like to talk about GIS
- but nobody "owns" GIS - it's for everybody!

ArcGIS - the main GIS software we provide

- NOT using this today but..
- The most popular desktop GIS
  - Windows OS only
  - for creating, analyzing and managing spatial data
- Large installation files, complicated software
  - known to crash often
  - but it is the most comprehensive GIS
- Available from ITS Virtual Lab (vlab.uchicago.edu)

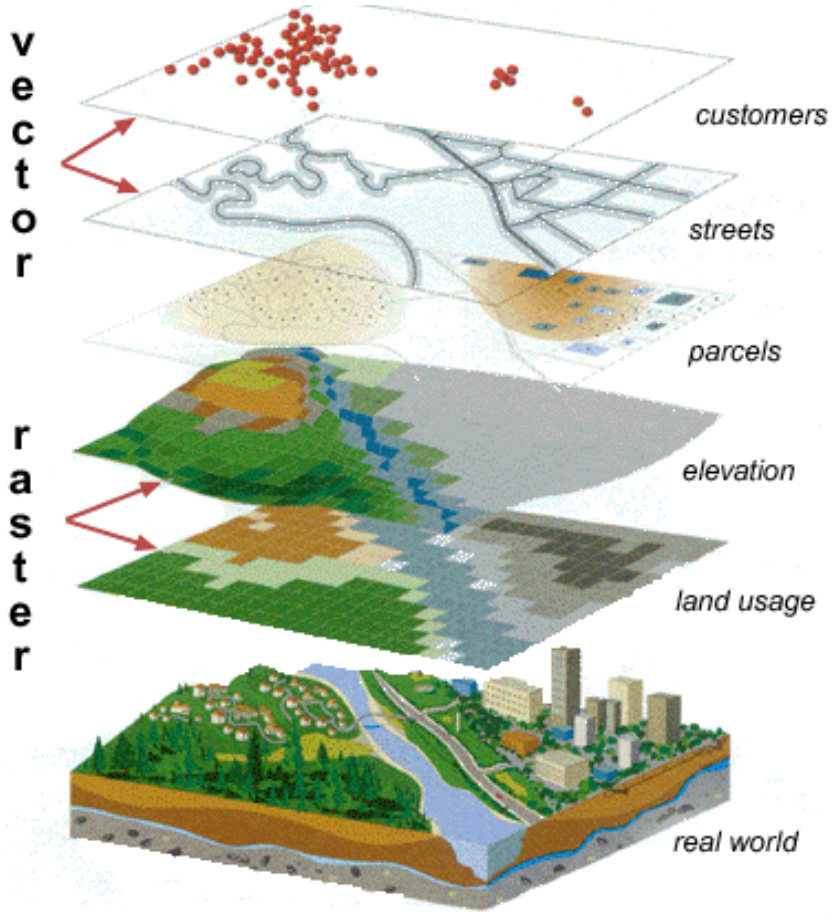
Open Source GIS

- OSGEO (Open Source Geospatial) Foundation projects
  - Supports and maintains many open source spatial projects
  - Ex. QGIS desktop
    - Alternative to ArcGIS
  - Host for geospatial libraries
    - GDAL & GEOS (core spatial libraries used by R, Python, QGIS and more)
- Other projects
  - Spatialite for SQLite

What are spatial datasets?

Two data models

vector	Geometry (points/lines/polygons) not 1-dimension array
raster	Matrix (cells/pixels)

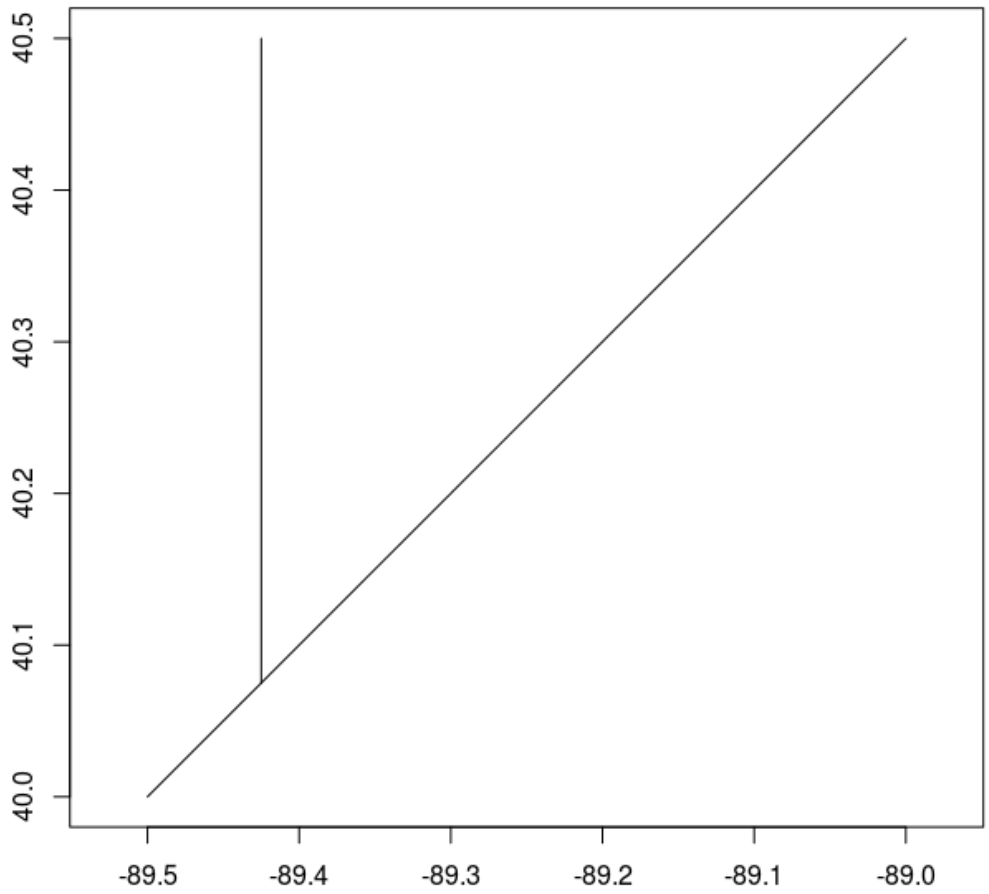


Spatial Data Types and Formats - Vector

Often accompanied by tabular descriptive attributes
- Data type choice for discrete objects

Lines example (right)
..but could be:

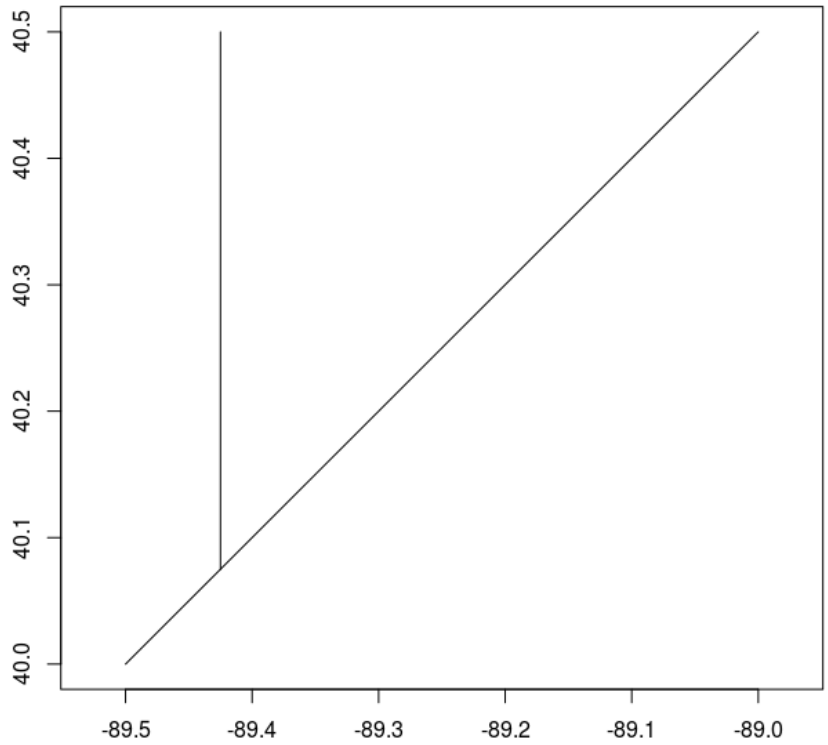
- Points
- Lines (polylines)
- Polygons (areas)



Spatial Data Types and Formats - Vector

Example (xy coordinates recorded at each vertex)

```
1   0
2    -89.00, 40.50
3    -89.05, 40.45
4    -89.10, 40.40
5    -89.15, 40.35
6    -89.20, 40.30
7    -89.25, 40.25
8    -89.30, 40.20
9    -89.35, 40.15
10   -89.40, 40.10
11   -89.45, 40.05
12   -89.50, 40.00
13  end
14  1
15   -89.425, 40.50
16   -89.425, 40.45
17   -89.425, 40.40
18   -89.425, 40.35
19   -89.425, 40.30
20   -89.425, 40.25
21   -89.425, 40.20
22   -89.425, 40.15
23   -89.425, 40.10
24   -89.425, 40.075
25  end
26  end
```



Spatial Data Types and Formats - Vector

Binary

- SHP/SHX/DBF (Shapefiles)
  - ArcGIS (ESRI)
- TAB
  - MapInfo
- DWG
  - AutoCAD

Database

- ArcGIS Geodatabase
- PostgreSQL/PostGIS
- SQLite/Spatialite

Text (Ascii/Unicode)

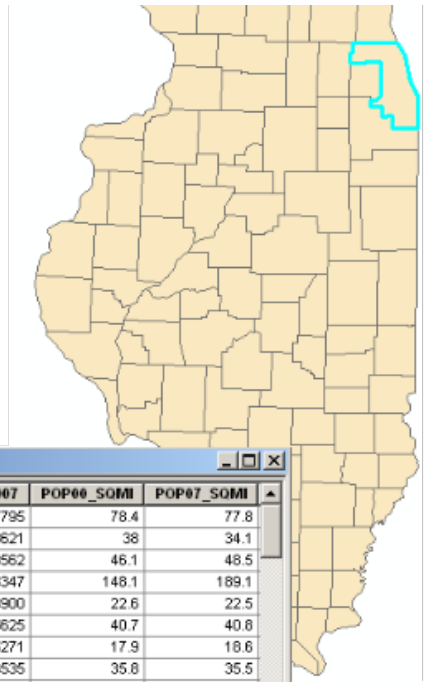
- DXF
  - AutoCAD
- KML
  - Google Earth
- OSM
  - Openstreetmap

Most likely your data is in the shapefiles format, but databases and text files that follow OGC's simple features standard are becoming more common.

Spatial Data Types and Formats - Vector

Shapefiles - requires minimum 3 extensions (optional: prj,sbn,sbx,xml)

- SHP - shape/geometry
- DBF - attributes/table
- SHX - index to bind above two



Attributes of IL_Counties

FID	Shape *	ObjectID	NAME	STATE_NAME	STATE_FIPS	CNTY_FIPS	FIPS	POP2000	POP2007	POP00_SQMI	POP07_SQMI
33	Polygon	973	Adams	Illinois	17	001	17001	68277	67795	78.4	77.8
87	Polygon	1665	Alexander	Illinois	17	003	17003	9590	8621	38	34.1
59	Polygon	1255	Bond	Illinois	17	005	17005	17633	18562	46.1	48.5
92	Polygon	3065	Boone	Illinois	17	007	17007	41786	53347	148.1	189.1
36	Polygon	993	Brown	Illinois	17	009	17009	6950	6900	22.6	22.5
5	Polygon	666	Bureau	Illinois	17	011	17011	35503	35625	40.7	40.8
52	Polygon	1156	Calhoun	Illinois	17	013	17013	5084	5271	17.9	18.6
95	Polygon	3079	Carroll	Illinois	17	015	17015	16674	16535	35.8	35.5
35	Polygon	990	Cass	Illinois	17	017	17017	13695	14181	35.7	37
28	Polygon	929	Champaign	Illinois	17	019	17019	179669	191869	180.1	192.3
43	Polygon	1051	Christian	Illinois	17	021	17021	35372	35283	49.4	49.3
51	Polygon	1142	Clark	Illinois	17	023	17023	17008	16848	33.7	33.4
61	Polygon	1278	Clay	Illinois	17	025	17025	14560	14396	31	30.6
65	Polygon	1313	Clinton	Illinois	17	027	17027	35535	37510	70.6	74.5
46	Polygon	1087	Coles	Illinois	17	029	17029	53196	51742	104.3	101.4
96	Polygon	3080	Cook	Illinois	17	031	17031	5376741	5407427	5613.9	5646.3
57	Polygon	1212	Crawford	Illinois	17	033	17033	20452	19840	45.9	44.5

Spatial Data Types and Formats - Vector

KML (kind of like a database)

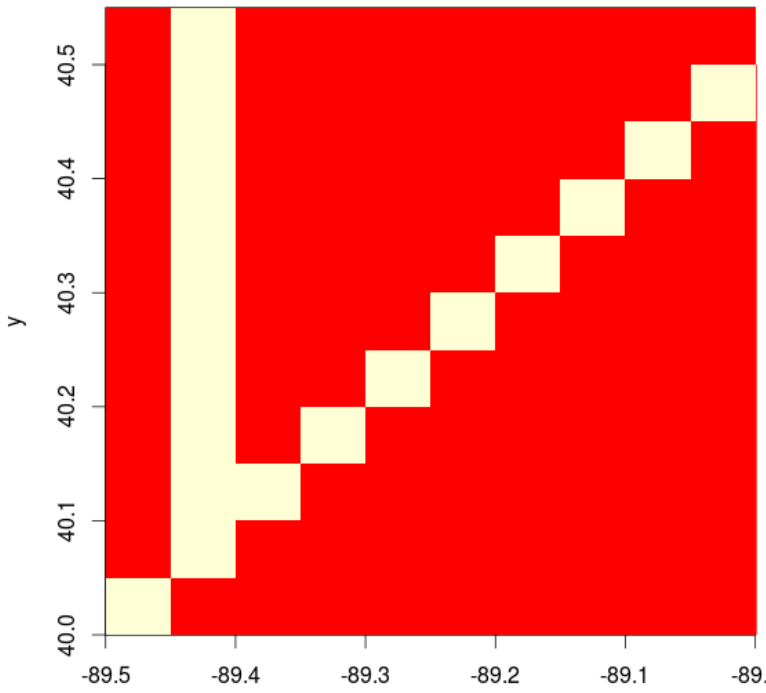
```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Document>
    <name>KML Samples</name>
    <open>1</open>
    <description>Unleash your creativity
          with the help of these examples!</
          description>
    <Style id="downArrowIcon">
      <IconStyle>
        <Icon>
          <href>http://maps.google.com/
                mapfiles/kml/pal4/icon28.png
                </href>
        </Icon>
      </IconStyle>
    </Style>
    <Style id="globeIcon">
      <IconStyle>
        <Icon>
          <href>http://maps.google.com/
                mapfiles/kml/pal3/icon19.png
                </href>
        </Icon>
      </IconStyle>
      <LineStyle>
        <width>2</width>
      </LineStyle>
    </Style>
    <Style id="transPurpleLineGreenPoly">
      <LineStyle>
        <color>7fff00ff</color>
        <width>4</width>
      </LineStyle>
      <PolyStyle>
        <color>7f00ff00</color>
```

```
        <latitude>36.14979247129029</
              latitude>
        <altitude>0</altitude>
        <heading>-55.42811560891606</
              heading>
        <tilt>56.10280503739589</tilt>
        <range>401.0997279712519</range>
      </LookAt>
      <styleUrl>#transYellowPoly</
            styleUrl>
      <Polygon>
        <extrude>1</extrude>
        <tessellate>1</tessellate>
        <altitudeMode>relativeToGround</
              altitudeMode>
        <outerBoundaryIs>
          <LinearRing>
            <coordinates>
                  -112.3348783983763,36.151400846
                -112.3372535345629,36.1488851755388
                -112.3356068927954,36.1478161267928
                -112.3350034807972,36.1484646902417
                -112.3358353861232,36.1489624162954
                -112.3345888301373,36.1502622937250
                -112.3337937856278,36.1497809602646
                -112.3331798208424,36.1504472788618
                -112.3348783983763,36.1514008468736
                  </coordinates>
          </LinearRing>
        </outerBoundaryIs>
      </Polygon>
    </Placemark>
```

Spatial Data Types and Formats - Raster

Image, matrix
- Data type choice for continuous phenomena/conditions
- Simple data structure, easy to read and analyze
- Mostly binary but could be just a text..

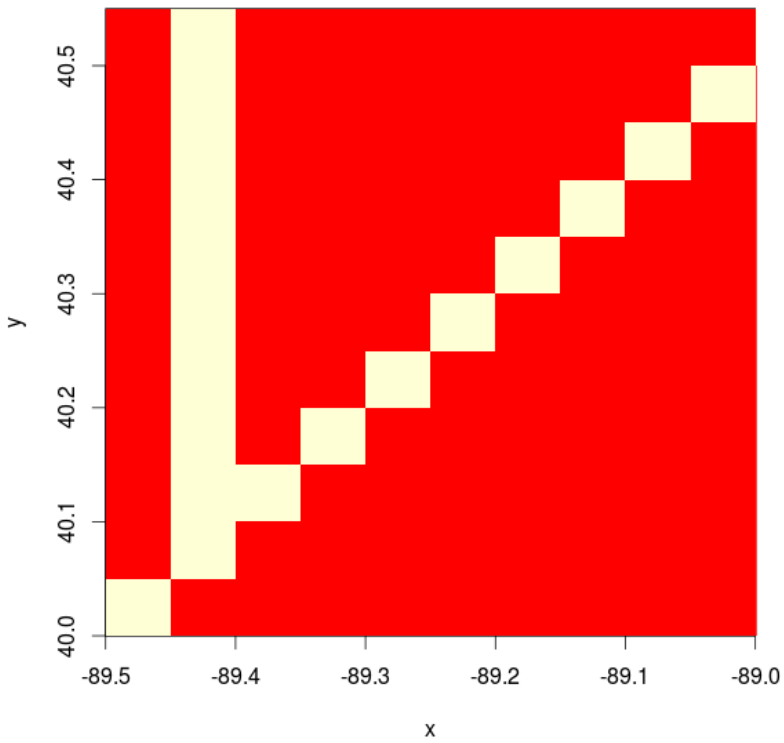
- ESRI GRID
- IMG (Erdas Imagine)
- SID (MrSID)
- BMP (Bitmap)
- TIF
- JPG
- .. also in databases



Spatial Data Types and Formats - Raster

Example (header with matrix)
- Header is NOT required (separate metadata or world file will do) but these XY-origins, cell size, projection info make the data "spatial"

```
ncols 11
nrows 11
xllcorner -89.5
yllcorner 40
cellsize 0.05
nodata_value -99
1 8 1 1 1 1 1 1 1 1 8
1 8 1 1 1 1 1 1 1 8 1
1 8 1 1 1 1 1 1 8 1 1
1 8 1 1 1 1 1 8 1 1 1
1 8 1 1 1 1 8 1 1 1 1
1 8 1 1 1 8 1 1 1 1 1
1 8 1 1 8 1 1 1 1 1 1
1 8 1 8 1 1 1 1 1 1 1
1 8 8 1 1 1 1 1 1 1 1
1 8 1 1 1 1 1 1 1 1 1
8 1 1 1 1 1 1 1 1 1 1
```



Coordinate Reference Systems/Map Projections

- XY coordinates - unique attributes pertaining to spatial data
- Problem: there are many ways to set XY
  - We need to know how XY was set to locate the 3-dimensional Earth objects on 2-dimensional space (map/graphics)
- Key concepts:
  - Datum - Earth's (shape) models
  - Coordinate reference systems - location reference system
    - Geographic (lon/lat)
    - Or, planar/grid systems? (UTM, US State Plane, etc.)
  - Map Projections - cartographical devices (Mercator, Lambert, etc.)
- Link: excellent guide for R users
- Link: list of EPSG numbers & proj4string

Spatial analysis

Comprehensive and authoritative spatial analysis textbook:
http://www.spatialanalysisonline.com/

Key concepts from the book:

Vector-base

"map overlay (combining two or more maps or map layers)"

"simple buffering (identifying regions of a map)", etc.

Raster-base

"a range of actions applied to the grid cells of one or more maps (or images) often involving filtering and/or algebraic operations (map algebra)"

Spatial analysis

Mapping/visualization

"I need to make a map."

Overlay/zone statistics

"How many businesses will be affected by the zoning change?"

"How much of wildfire area is contained?"

"How often and what type of crimes occurred in each police district?"

Adjacency/contiguity

"Tell me the names of adjacent cities and villages."

Proximity

"I need to find resources near XXXXX."

Distance

"How far is it between this and that?"

Spatial analysis

Geocoding (add XY)

"I have a list of addresses to geocode/locate.

Travel/flow time

"How long does it take to get there from here by foot, car or public transportation?"

Imagery

"Tell me what's on the ground based on the satellite image."

Terrain/elevation

"Calculate slope or viewshed area."

Interpolation

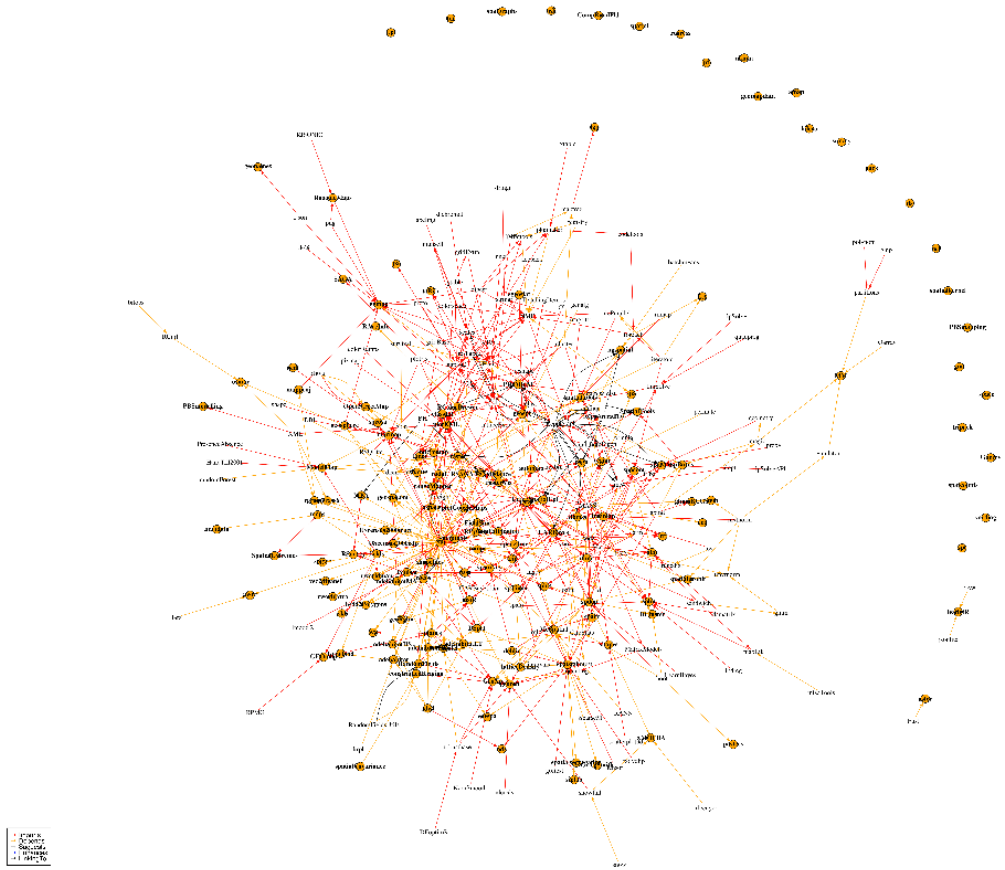
"Based on the nearby survey result, estimate the total snowfall for this location."

R

- R is free
- R is a dialect of the S language
- R is based on 40-year old technology
- R offers statistical analysis environment
- Learn R - Springer's UseR! series (PDF books from the library)
- IDE (integrated development environment) for R
  - RGui (Windows/Mac default) - use this today
  - R commander, RStudio, Revolution, Rattle, ..
- R functionality is divided into a number of packages
  - Our focus is on packages from CRAN's "Spatial" task view!
  - https://cran.r-project.org/web/views/

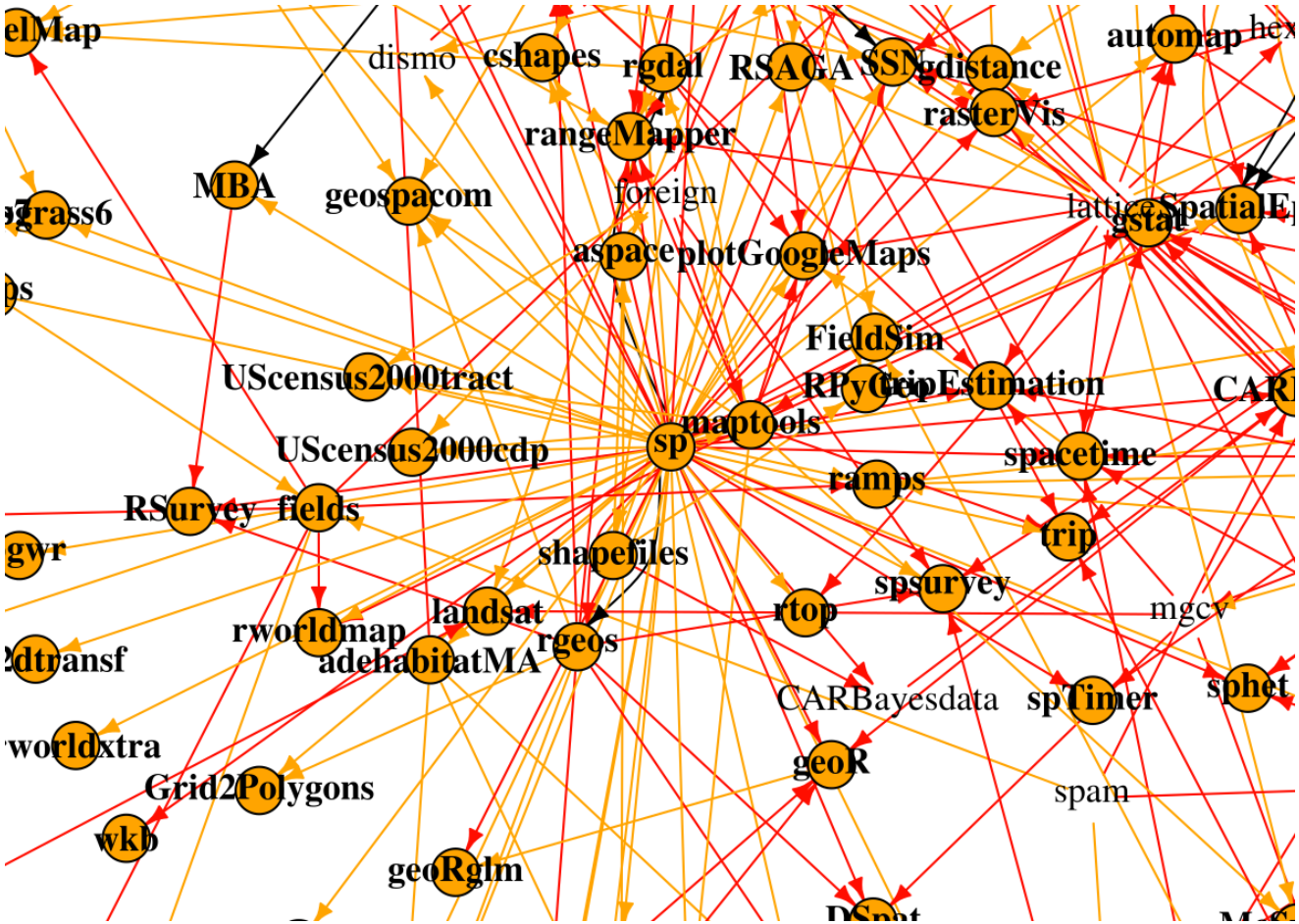
R Spatial

- http://rspatial.r-forge.r-project.org/
- Currently, 154 packages are listed under "Spatial" task view
  - About 2 % of all packages (my estimate as of 9/10/15)
- R-SIG-Geo: email list



R Spatial

- most "Spatial" packages depend on "sp" package



R Spatial - sp

- "sp" defines "spatial" data object classes (esp. for vector)
  - Great advantage for doing spatial analysis in R - standardized access.
- R objects
  - data.frame,matrix,array,vector,list,variables(char/numeric/logical)
- R-spatial objects are more complex
  - Vector: Spatial{Points:Lines:Polygons}, Spatial{Points:Lines:Polygons}DataFrame
  - - composed of multiple slots
    - geometry {points:lines:polygons} with coordinates
    - bbox (extent/bounding box information)
    - proj4string (projection/coordinate system information)
    - data
  - Raster - often extended with raster package
    - Grid (similar structure as vector sp objects)
    - Rasterlayer, RasterStack, RasterBrick (with raster package)

R Spatial - rgdal, rgeos

rgdal package - R interface to GDAL

- Requires GDAL (must-have geospatial library)
- Full GDAL installation includes handy utilities.
- Three components
  - GDAL: raster data handling
  - OGR : vector data handling
  - PROJ4 : spatial/coordinate reference system

rgeos package - R interface to GEOS

- Requires GEOS, a spatial (vector) analysis library

Please consult this excellent guide for installation advice

- http://geoscripting-wur.github.io/system_setup/

R Spatial - reading data

Vector

- rgdal - check with ogrDrivers()
  - readOGR("roads.shp","roads")
- maptools - shapefiles only, don't read PRJ
  - readShapeSpatial("roads")

Raster

- rgdal - check with gdalDrivers()
  - readGDAL("elevation.asc") — SpatialGridDataFrame
  - To convert to "rasterLayer", simply pass it to raster()
- raster - GeoTIFF, IMG, GRD, BIL, BSQ, ArcASCII, SAGA, IDRISI
  - raster("elevation.asc") — rasterLayer (matrix/2-dimensions)
- Not "spatial" : tiff (readTIFF), jpeg (readJPEG), png (readPNG)
  - Like Matlab/other stats software, these return matrix/array (2-/3+-dimensions)

R Spatial - writing data

Vector

- rgdal - check with ogrDrivers()
  - writeOGR(roads, "roads2.shp","roads2", driver="ESRI Shapefile")
- maptools - shapefiles only, don't read PRJ
  - writeSpatialShape(roads, "roads2.shp")

Raster

- rgdal - check with gdalDrivers()
  - writeGDAL(elevation, "outputfile", drivername="GTiff")
- raster - RD, BIL, BSQ, ArcASCII, SAGA, IDRISI
  - writeRaster(elevation, "output", format="GTiff")