GIS and Spatial Data

Robert Dinterman
NC State University
October 28, 2013

What is GIS?

Geographic Information System is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data.

Typically, one thinks of ArcGIS as the mack-daddy of GIS software. Commercially produced by esri, this is a powerful program for just about any GIS application one can think of. One typically needs a license for this ($$$).

We can use R to act as GIS for no (monetary) cost. Our goals today:

1. Load up Spatial Data
2. Simple Plots
3. Map Projections

R for Spatial Data

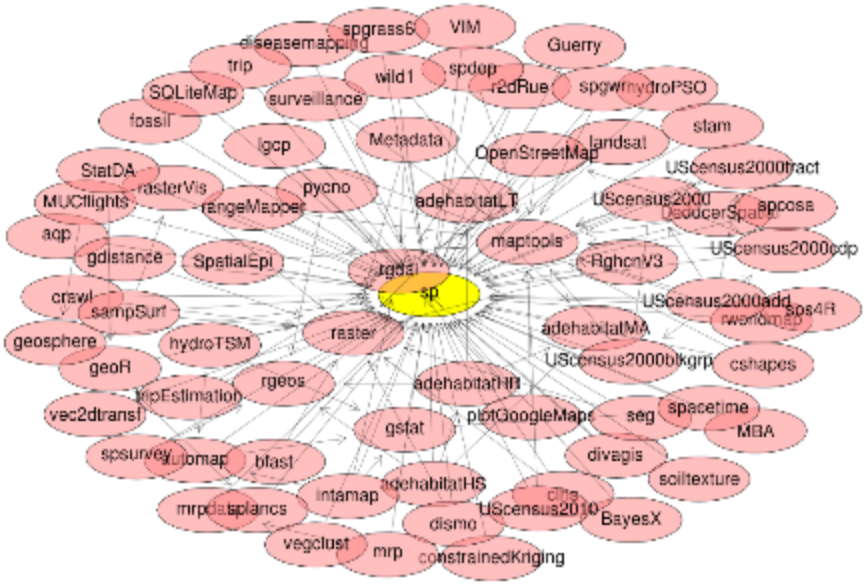
Packages used for today:

- gpclib
- mapproj
- maps
- maptools
- rgdal
- rgeos
- sp

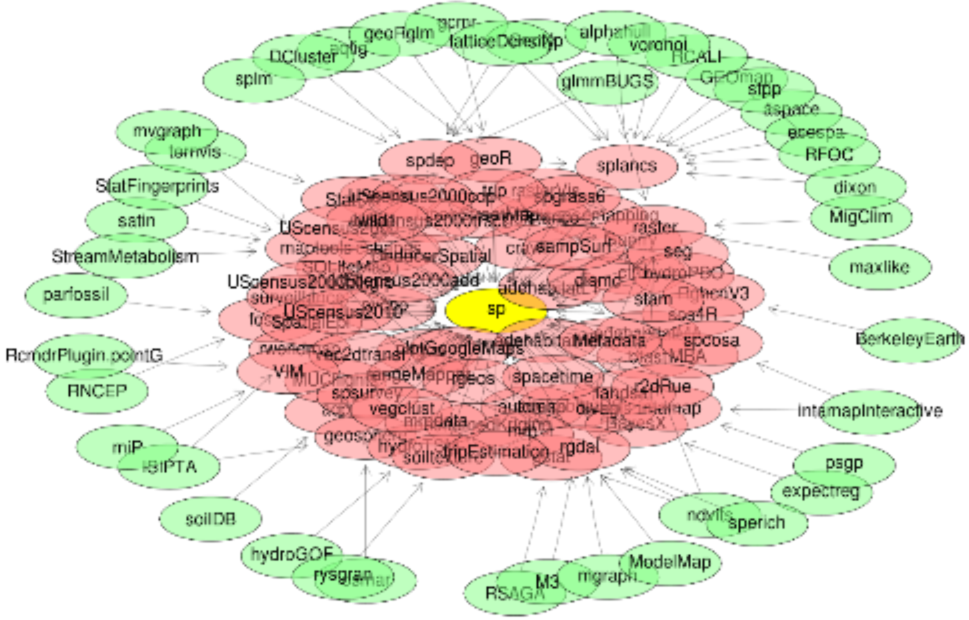
Gee, that seems like a lot.

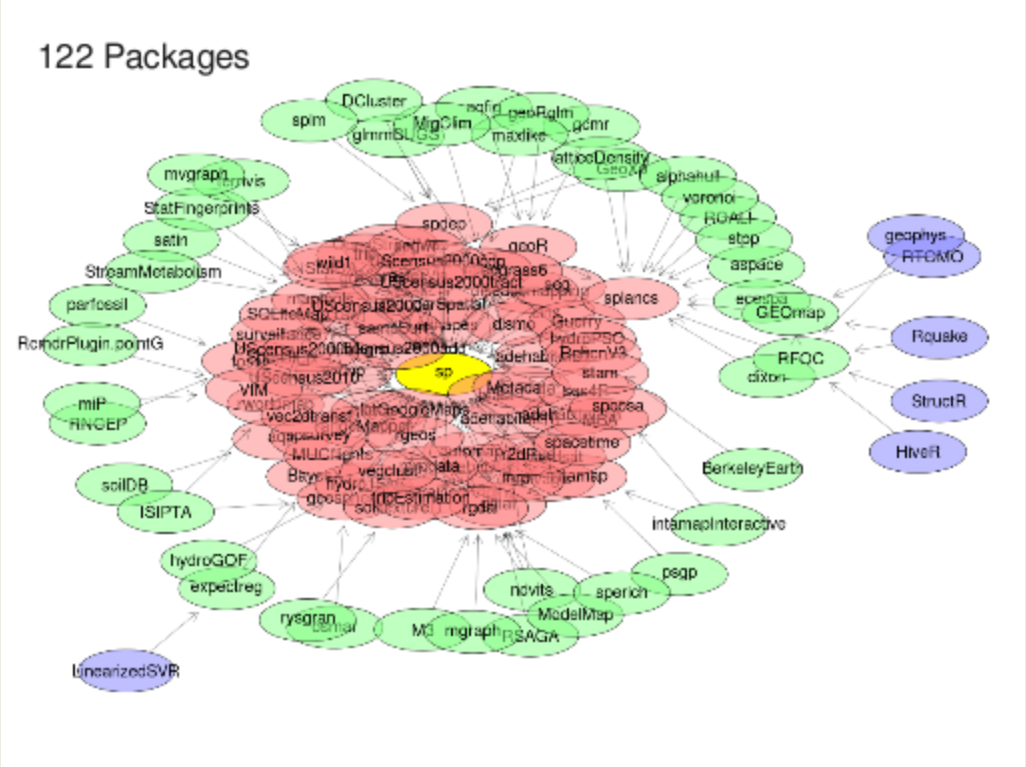
1 Package

73 Packages


spgrass6
VIM
Guerry
trip
wild1
spdep
r2dRue
SQLiteMap
surveillance
fossil
lgcp
Metadata
OpenStreetMap
landsat
stam
StatDA
rasterVis
pycno
MUCflights
rangeMapper
adehabitatLT
UScensus2000tract
UScensus2000
apcose
aqp
maptools
gdistance
SpatialEpi
RghcnV3
UScensus2000cdp
rgdal
sp
crawl
sampSurf
UScensus2000add
sos4R
raster
adehabitatMA
geosphere
hydroTSM
UScensus2000blkgrp
cshapes
geoR
rgeos
adehabitatHR
vec2dtransf
plotGoogleMaps
seg
spacetime
MBA
gstat
spsurvey
bfast
divagis
intamap
adehabitatHS
sciltexture
mrpdata
splancs
UScensus2010
BayesX
dismo
vegclust
mrp
constrainedKriging

116 Packages




122 Packages
DCluster
spim
MigClim
maxlike
gcmr
mvgraph
StatFingerprints
sdpdc
gcoR
sedt
StreamMetabolism
splancs
perfosil
GEOmap
RcmdrPlugin.pointG
sp
Rquake
RFOC
VIM
StructR
miP
RINGEP
spacetime
HiveR
BerkeleyEarth
soilDB
ISIPTA
intamapInteractive
hydroGOF
psgp
expectreg
novits
sperich
ModelMap
rysgran
linearizedSVR

R for Spatial Data

Types of files:

- *.csv with Longitude and Latitude (order is important)
- *.E00 Arc/Info ASCII grid
- *.shp Esri shapefile (most common)
  - *.shx and *.dbf are **required** files for shapefile to function
  - Attributes/Optional files: *.prj, *.sbx, etc.

Shapefile needs to have all files with the same name but with different extensions to denote the component that R is reading.

R operates with different classes of Spatial Data.

DATA TYPE	CLASS	CONTAINS	CAN BE data.frame?
points	SpatialPoints	Spatial/SpatialPoints	yes
pixels	SpatialPixels	SpatialPoints/SpatialPixels	yes
full grid	SpatialGrid	SpatialPixels/SpatialGrid	yes
lines	Lines	Line	no
lines	SpatialLines	Spatial/SpatialLines	yes
rings	Polygons	Line/Polygon	no
rings	SpatialPolygons	Spatial/Polygons/SpatialPolygons	yes

R for Spatial Data

You will need to utilize different commands for reading in Spatial Data dependent upon the type of data you have.

readShapePoints, readShapeLines, readShapePoly, readShapeSpatial

How to load:

Let's jump into some R Code for today.

Also, let's figure out how to merge with other data.

Types of Spatial Data

Three main types of Spatial Data:

1. Areal
2. Geospatial
3. Spatial Point Pattern

Determines how to read in data for you.

You don’t want a polygon as a line...
or a point as a Polygon.

Areal Data

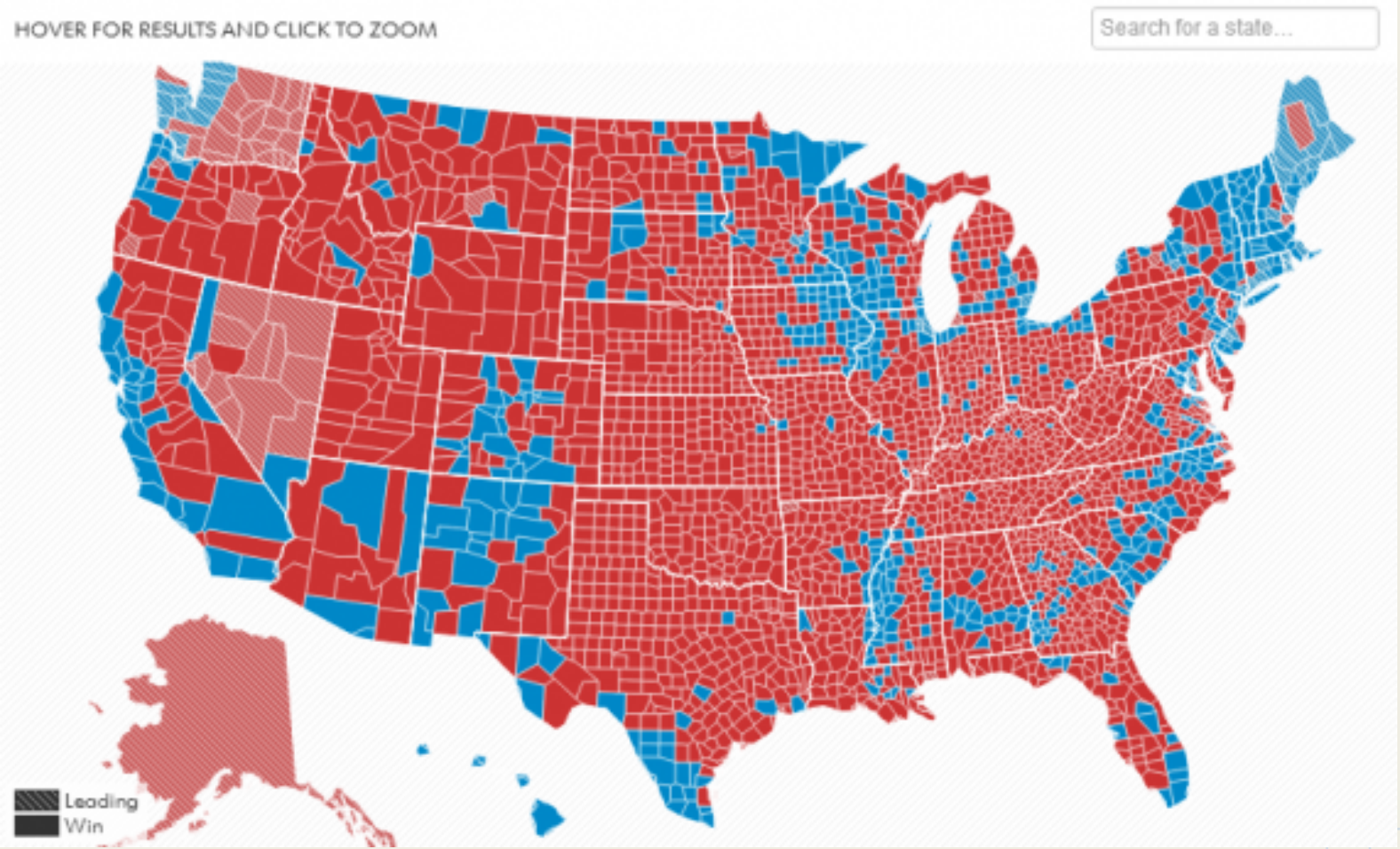
Definition: There are a fixed number observations/locations/regions in a spatial domain.

Also known as Regional or Lattice Data. Lattice refers to a grid-like structure. Counties fall into this category, although most are irregular lattice.

Sources:

National Atlas - collection of over 20 Federal organizations and their data
TIGER - **T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing

Census Geographic Hierarchy, WARNING Alaska and Virginia suck. Every other state is OK by me.



Areal Data Example

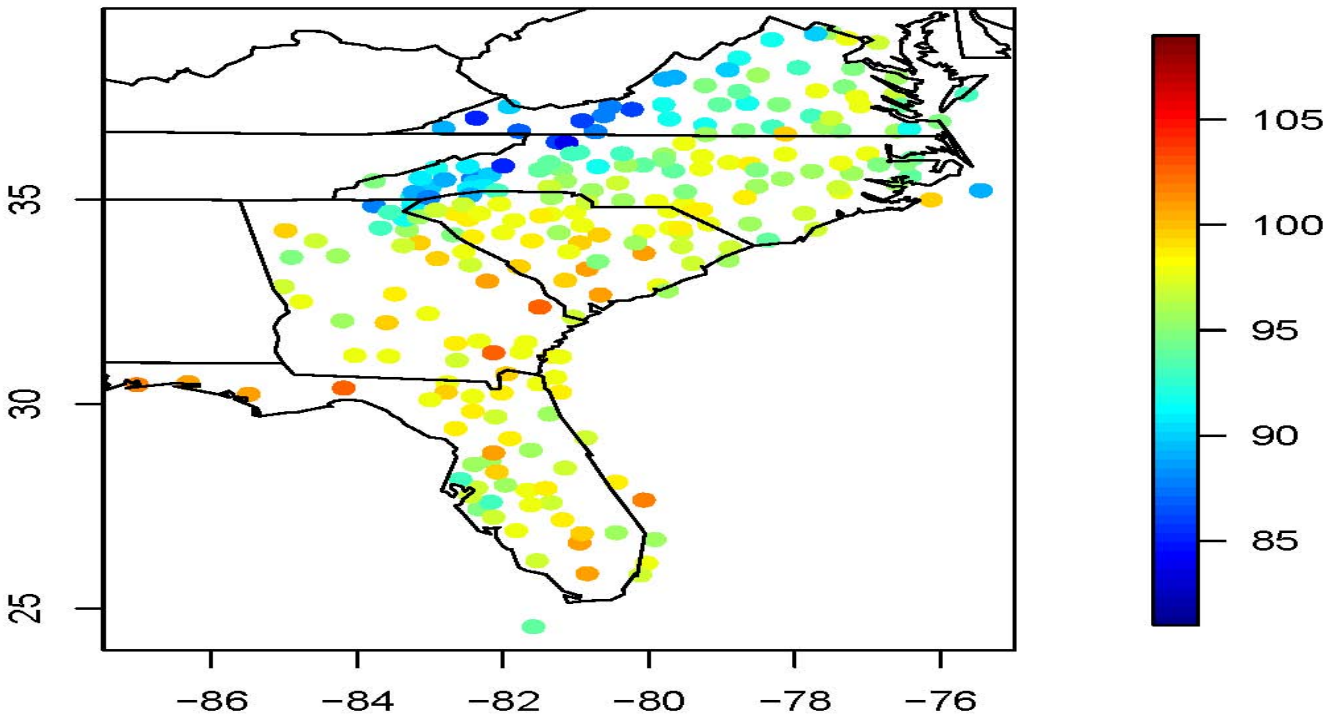
Geospatial Data

Definition: Y(s) is the response at location s. s is an element of D where D is some spatial domain with infinitely many points.

Also called Point-Reference Data.

One interpretation is that is a response for all points/locations in D. The response is therefore a continuous surface. The response is thought of as the random variable.

Mainly used in Spatial Statistics.



Geospatial Data Example

Spatial Point Pattern Data

Same set-up as Geospatial except the locations themselves are random and not the response.

Typically seen in “hot spot” analysis.

Simple Plots

Let's go back to our R code to see how we can plot our data.

Notice, R plots with different layers. We can get fancy and plot multiple layers to indicate some intricate relationship. Or even not so intricate relationships, like just making the ocean blue.

Code today is just for simple plotting. Nothing fancy and, frankly, these plots don't even look nice. But that is OK, next time we will delve into publication quality figures.

In case you get lost on some commands, here is a good overall reference guide to R: Cookbook for R

Map Projections

A map projection is a system in which locations on the curved surface of the earth are displayed on a flat sheet or surface according to some set of rules.

Angles, areas, directions, shapes and distances become distorted when transformed from a curved surface to a plane.

Economists want to make sure we don't look like buffoons when using maps.

We can use the package mapproj for this.

Also, here is a listing of EPSG projections that can be used with proj4string
And HERE is another one!

References

Parts of this presentation utilizes information from other sites:

http://www.ncgia.ucsb.edu/giscc/units/u002/

http://www.ncgia.ucsb.edu/giscc/gateway.html

http://www.maths.lancs.ac.uk/~rowlings/Teaching/UseR2012/

http://geography.uoregon.edu/bartlein/courses/geog495/index.html

http://www4.stat.ncsu.edu/~reich/st733/