

Geospatial and Census Data in R

Yun Tai | Postdoctoral Fellow
Research Data Services | UVa Library

Topics

Geospatial Data

- ▶ What's special about spatial data
- ▶ Handling spatial data in R

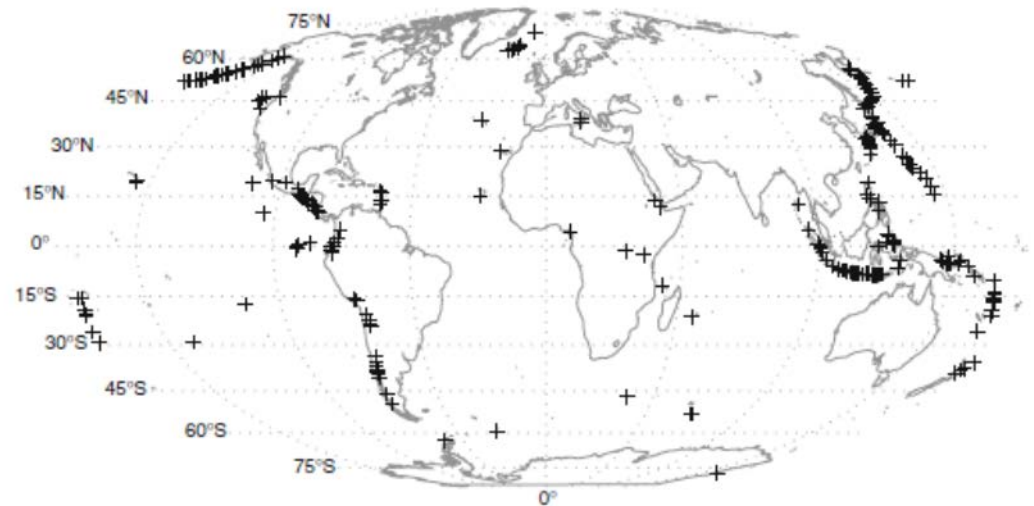
US Census Data

- ▶ US Census data structure
- ▶ Using Census 2000 and 2010 data in R



Spatial Data

- ▶ Data with spatial informational, and more
 - Coordinate values
 - Pairs of longitude/latitude degree values
 - Coordinate Reference Systems (CRS)
 - Defines how an ellipsoid, the shape of the earth, is represented on the plane, e.g., a piece of paper or computer screen
 - Attributes
 - Non-spatial information



Source: Bivand et al. 2008:7



Spatial Data

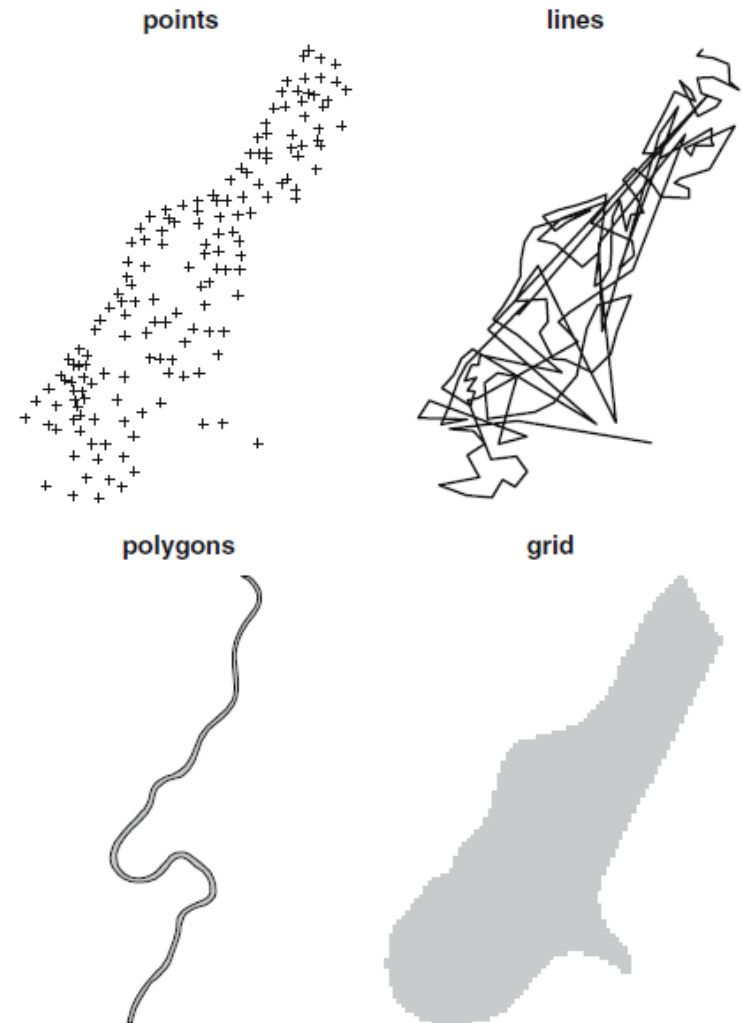
Representing positional information -

► Vector

- Point: a single point location
- Line: a set of points connected by straight line segments
- Polygon: an area marked by enclosing lines

► Raster

- Grid: a collection of points or rectangular cells organized in a regular lattice



Source: Bivand et al. 2008: 58

Spatial Data in R

► R and GIS

- Spatial data in Geographical Information System
- R may be sufficient itself or used in conjunction with GIS software

Why use R for spatial data?

- Easy to store and implement analysis components and approaches
- Many standard and innovative statistical analysis methods available
- An open source with contributed “packages”



The **sp** package and Spatial objects

- ▶ The **sp** package provides classes and methods for handling and analyzing spatial data, i.e., points, lines, polygons and grids.
- ▶ Spatial objects in R: the two slots of spatial class
 - Bounding box: an area defined by two longitudes and two latitudes contains the spatial object, e.g., Greater London is enclosed by `bbox = -0.49, 51.28, -0.24, 51.68`
 - CRS class object: defines the coordinate reference system, e.g., WGS84



Subclasses of Spatial objects

- ▶ **SpatialPoints**: point data, represented by x and y coordinates
- ▶ **SpatialPointsDataFrame**: contains spatial point coordinates with “their” associated data
- ▶ **SpatialPolygons**: polygon data, representations of sets of polygons
- ▶ **SpatialPolygonsDataFrame**: brings together the spatial representations of the polygons with data
- ▶ **Other spatial objects**: SpatialLines, SpatialGrid and SpatialPixel

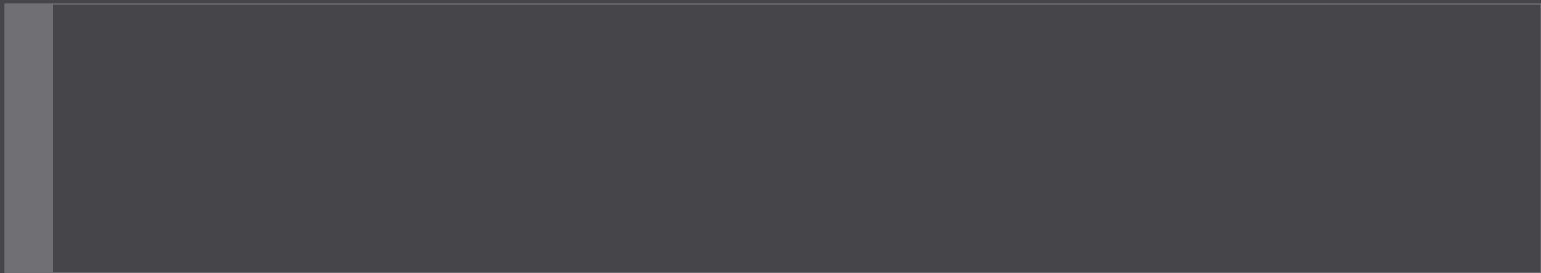


The rgdal package

► **rgdal**

- R's interface/plugin of the Geospatial Spatial Data Abstraction (GDAL) library (<http://www.gdal.org/>)
- Provides access to projection/transformation operations from the PROJ.4 library (<http://proj4.org/>)
- readOGR() function: read vector maps into R spatial objects





Let's go to R.

The US Decennial Census

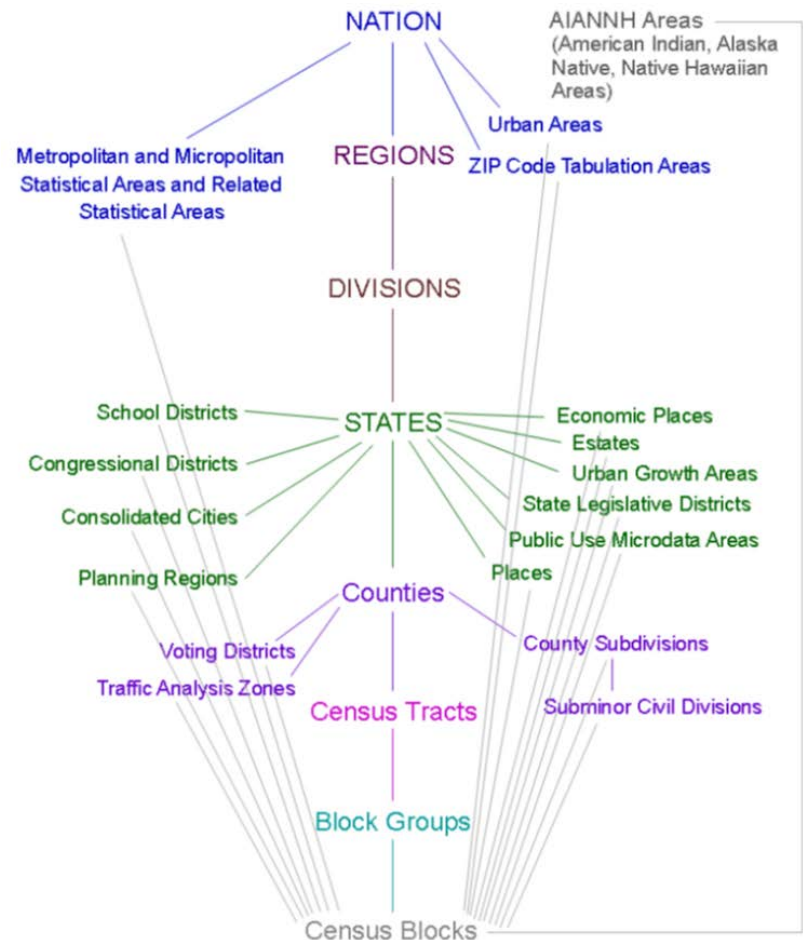
- ▶ The census has been taken decennially since 1790
- ▶ The long form had been sent to a sample of households from 1940 to 2000, to gather detailed information
- ▶ The American Community Survey (ACS) began nationwide full implementation in 2005, replacing the long form
- ▶ Census 2010 is the 23rd and most recent decennial census



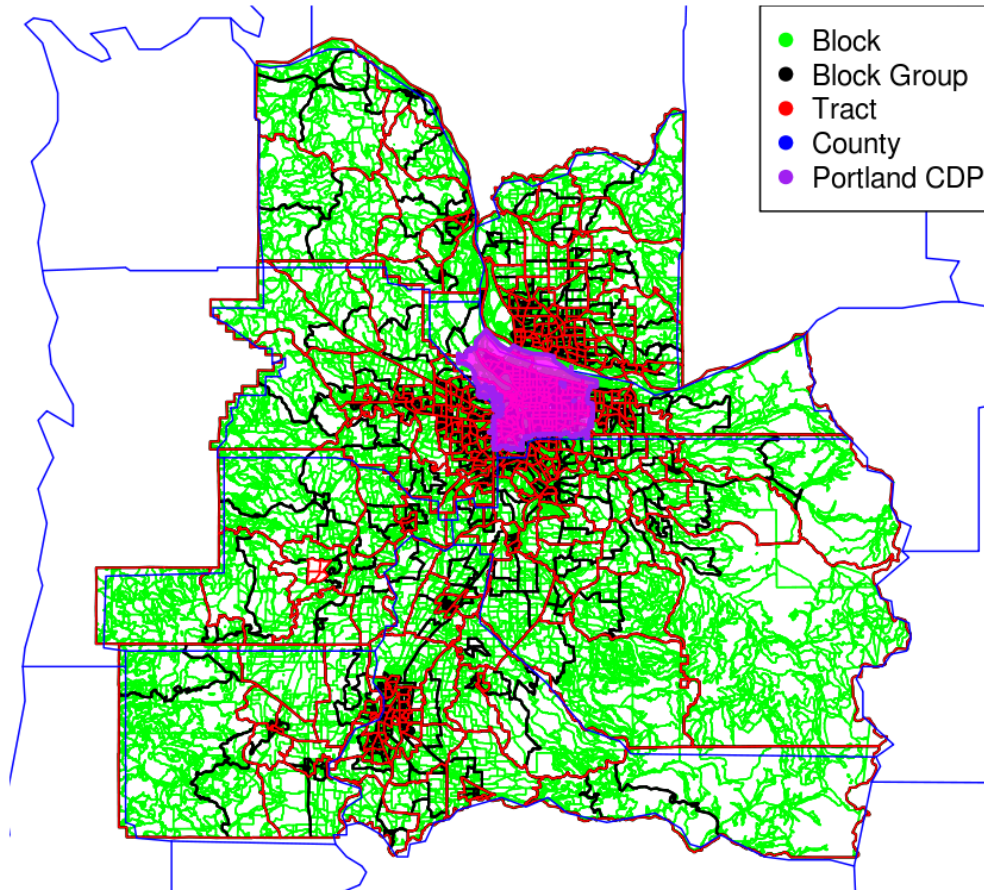
Census Geographic Areas

- ▶ The four basic geographic levels: county, tract, block group and block
- ▶ Two other geographic conglomerations: metropolitan statistical areas (MSA) and census designated places (CDP)

Standard Hierarchy of Census Geographic Entities



Portland MSA Eight Counties



Census Geographic Areas: Portland

- Representation of block, block group, tract, and county hierarchy in US Census geography
- MSAs are composed of counties
- CDPs do not observe boundaries of the other polygons

(Almquist 2010)

R Census packages:

UScensus2000 and **UScensus2010**

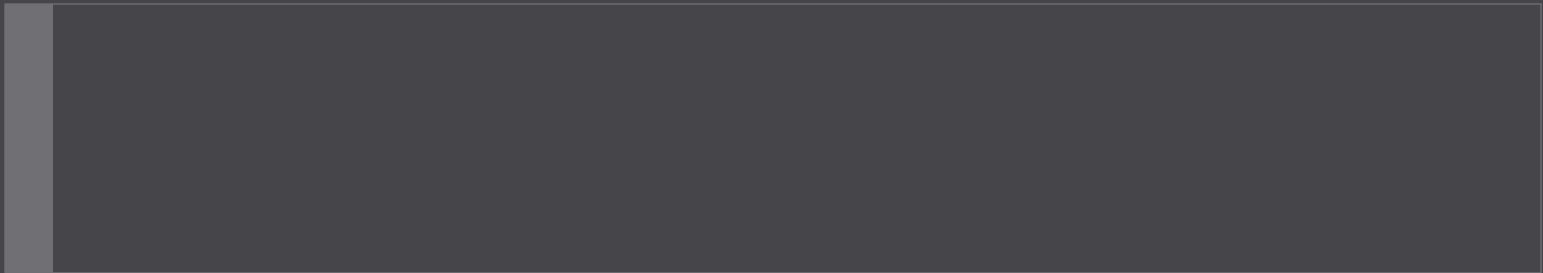
- ▶ Organize separate entities of the US Census data files including the actual demographic information, shapefiles and metadata into a single package
- ▶ Provide examples and functions such as identifying and extracting data



Available Data in **UScensus2010**

- ▶ Block: UScensus2010blk
- ▶ Block group: UScensus2010blkgrp
- ▶ Tract: UScensus2010tract
- ▶ Census designated places: UScensus2010cdp





Let's go to R.

Other Census Data Tools

- ▶ Social Explorer

<http://data.library.virginia.edu/datasources/licensed/>

- ▶ IPUMS-USA

<https://usa.ipums.org/usa/>

- ▶ National Historical Geographic System (NHGIS)

<https://nhgis.org/>

- ▶ American FactFinder

<http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>



References

Almquist, Z.W. 2010. US Census Spatial and Demographic Data in R: The UScensus2000 Suite of Packages. *Journal of Statistical Software* 37(6), 1–31.

Bivand, R.S., E.J. Pebesma and V. Gomez-Rubio. 2008. *Applied Spatial Data Analysis with R*. New York: Springer.

- Find ebook in Virgo: <http://search.lib.virginia.edu/catalog/u4869290>
- Codes and dataset: <http://www.asdar-book.org/courses.php>

Lovelace, R., J. Cheshire and R. Oldroyd. 2015. *Introduction to visualising spatial data in R*.

- For latest version: <https://github.com/Robinlovelace/Creating-maps-in-R>

Massey D. S. and N.A. Denton. 1988. The Dimensions of Residential Segregation. *Social Forces* 67 (2), 281-315.

Sparks, C. 2014. Measuring Residential Segregation using R. *Spatial Demography* 2(1): 72-78.

U.S. Census Bureau. 2008. *A Compass for Understanding and Using American Community Survey Data: What Congress Needs to Know*. U.S. Government Printing Office, Washington, DC.

