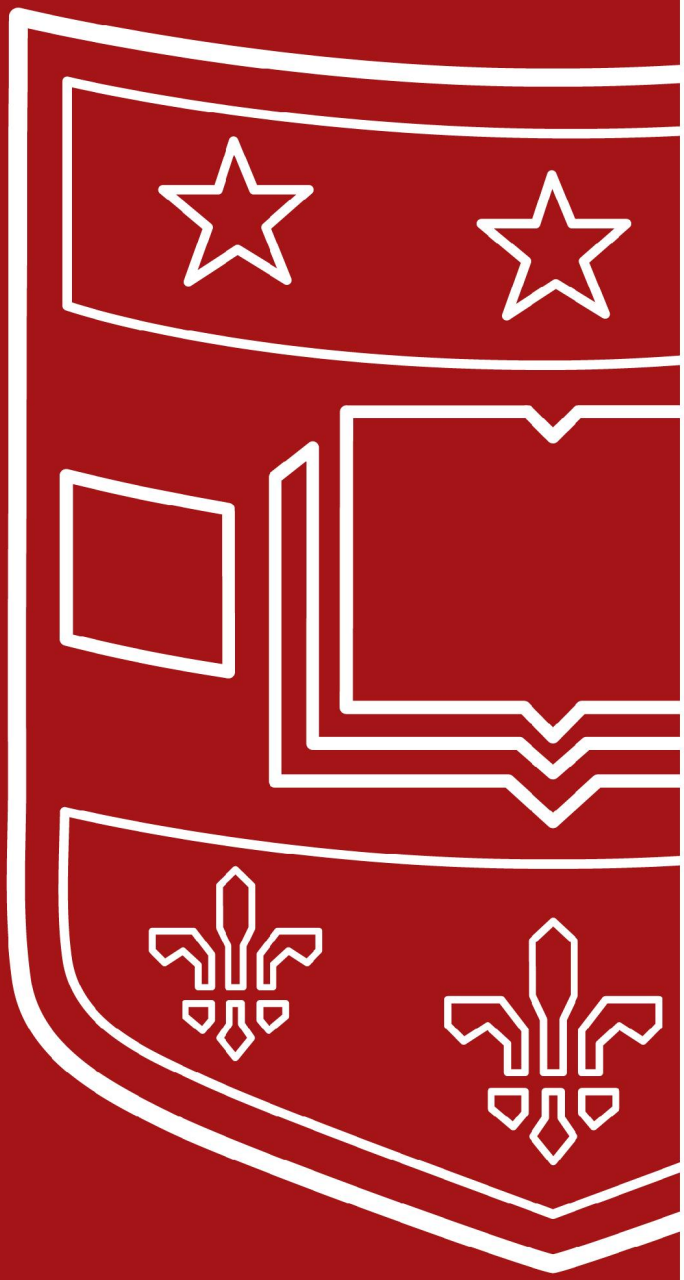
Tools for using spatial data within R

Saint Louis R User Group

Cindy Traub, PhD



Washington University in St. Louis

How do we store locational data?

- Lat/Lon pairs
- Addresses
- Place names
- Driving routes (turn by turn directions)
- Boundaries of areas
- By state/county/region/etc.

What kinds of objects?

- **[0-D] Single point locations**
  - (Latitude, Longitude)
  - local coordinate systems
- **[1-D] Line segments**
  - streets
  - boundaries
- **[2-D] Polygons**
  - counties, states, nations
  - lakes
  - building footprints

How do we present location data?

- Points/icons on a map
- Paths of roads on a map
- Images of regions on a map

- ON A MAP (2D image of our 3D planet)

Some challenges of map making

Some challenges of map making

- "Not only is it easy to lie with maps, it's **essential**."

  Mark Monmonier, *How to Lie with Maps*

- Impossible to avoid distortion on 2D image

- Geographic coordinates / projected coordinates

  [WGS84, NAD83] [Mercator, MO-east]

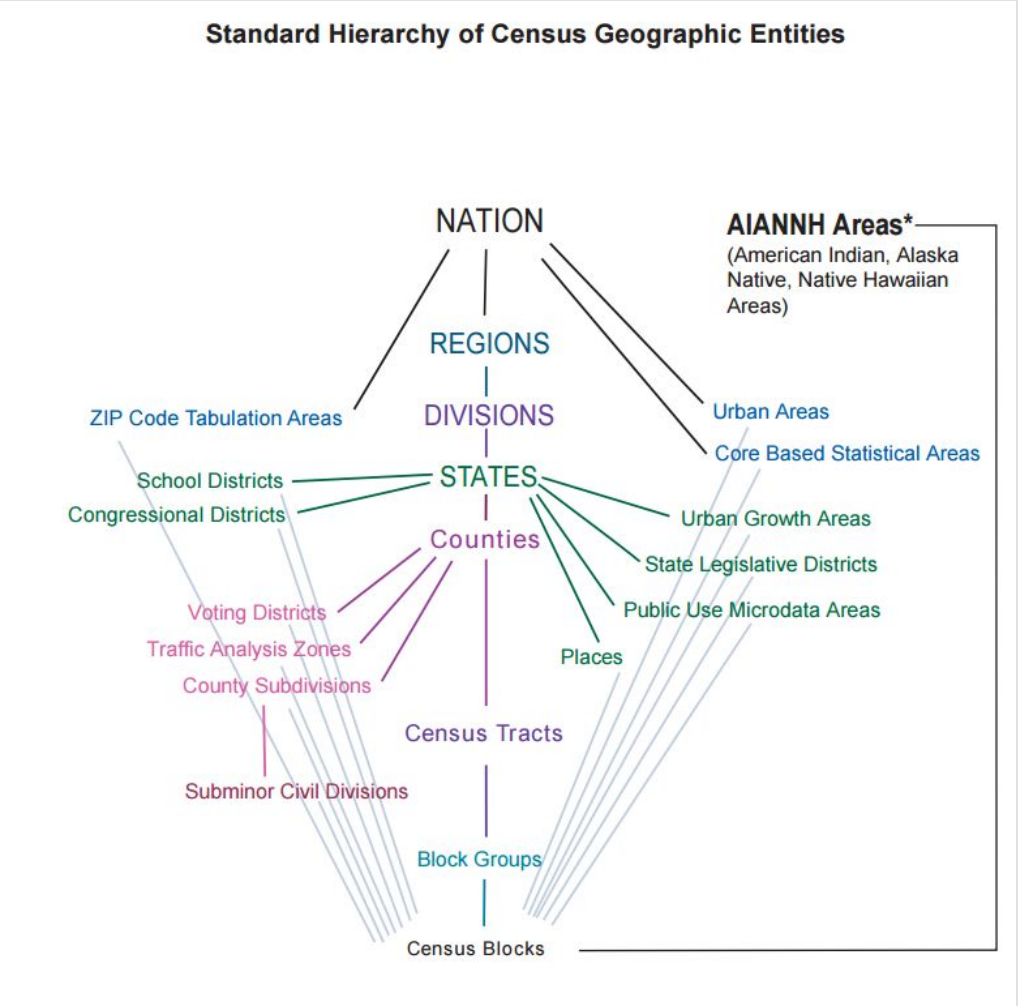
  [lat/long] [meters, feet, etc.]

Some Spatial File Types

- Shapefile (the geospatial equivalent of “csv”)
- KML (works with Google Earth)
- GeoJSON
- Geodatabase (ESRI proprietary format)

- SpatialPolygonsDataFrame (inside R)

How are countries organized?



Source: http://www2.census.gov/geo/pdfs/reference/geodiagram.pdf

Motivation for First Example

- Map of Santiago's transit system

  http://jkunst.com/r/plotting-gtfs-data-with-r/

PLOTTING GTFS DATA WITH R

OCTOBER 2014

R

3 COMMENTS

Structure of Transit Data

- GTFS = General Transit Feed Specification
  - Where does transit go?
  - Where does transit stop?
  - When does transit stop?
  - How are trips organized?

Structure of Transit Data

- GTFS = General Transit Feed Specification
  - Where does transit go? [on roads and rails]
  - Where does transit stop? [at stops and stations]
  - When does transit stop? [timetable]
  - How are trips organized?

    [Vehicles travel on sequence of roads/rails,
    stopping at designated locations,
    at specific times]

Structure of Transit Data

- GTFS = General Transit Feed Specification
  - Where does transit go? shapes.txt
  - Where does transit stop? stops.txt
  - When does transit stop? stop_times.txt
  - How are trips organized? routes.txt; trips.txt

Mapping Transit Data

To draw system map:

- Lines for roads/rails where transit goes
- Colored routes for Metrolink
- Option to symbolize points train, bus stops

Mapping Transit Data in R

- We will adapt code from
  http://jkunst.com/r/plotting-gtfs-data-with-r/

Follow along with .Rmd file (or the HTML) from
http://Libguides.wustl.edu/R (Meetup tab)

What else can I do in R?

- CRAN Task Views
  - Spatial
  - SpatioTemporal
  - Environmetrics
- Look for **vignettes**

Make a sound choice of technique

CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
WebTechnologies	Web Technologies and Services
gR	gRaphical Models in R

Resources

- Coordinate Systems and Projections
  https://education.usgs.gov/lessons/coordinatesystems.pdf
  http://www.msdis.missouri.edu/data/help/
- ArcGIS help files
  http://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/what-are-map-projections.htm
  http://desktop.arcgis.com/en/arcmap/10.3/guide-books/geocoding/what-is-geocoding.htm
  http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm
  http://desktop.arcgis.com/en/arcmap/10.3/manage-data/kml/what-is-kml-.htm
- Leaflet help
  https://rstudio.github.io/leaflet/
  http://leafletjs.com/
- CRAN task views
- See Markdown PDF for additional links to data

Thanks for coming!

- Any questions?

- Contact Cindy at ct@wustl.edu