Geocoding and Mapping in R

Hans Lueders

Stanford University
hlueders@stanford.edu

July 19, 2018
Why care?—Data visualization
Why care?—Data visualization

2015
Why care?—Data visualization
Why care?—Data visualization
Why care?—Data analysis

- Merge geocoded information to data on specific geographic units (e.g., municipalities, counties, districts, regions, states, countries, ...)
- What are the characteristics of municipalities targeted by political campaigns?
- What is the effect of campaign stops on turnout?
- Which city is closest to a particular point?
- Where do immigrants from a particular Mexican municipality typically move to in the US?
- What is the effect of climate change on domestic migration or political conflict?
- ...

...
Outline

1. Motivation
2. Geocoding in Theory
3. Geocoding in Practice
4. Mapping in Theory
5. Mapping in Practice
6. Using GIS for empirical analyses
7. Resources
Outline

1 Motivation
2 Geocoding in Theory
3 Geocoding in Practice
4 Mapping in Theory
5 Mapping in Practice
6 Using GIS for empirical analyses
7 Resources
Longitude and Latitude

X values measure Longitude - distance in degrees east or west of the Prime Meridian
Y values measure Latitude - distance in degrees north or south of the Equator
Getting geocodes on google maps
Getting geocodes on google maps
Getting geocodes on google maps

![Map with Google Maps interface showing geocoding information]

- Directions from here
- Directions to here
- What's here?
- Search nearby
- Print
- Add a missing place
- Report a data problem
- Measure distance

Encina Hall
616 Serra St, Stanford, CA, 94305
37.427372, -122.164626
Getting geocodes on google maps
Geocoding using the Google API

Pros:
- Works worldwide
- Accepts any kind of location (i.e., full addresses, cities, counties, countries, ...)

Cons:
- Daily limits
- Still need to check quality of geocodes
- Quality decreases for more remote places
Let’s do some geocoding!

R code
Outline

1. Motivation
2. Geocoding in Theory
3. Geocoding in Practice
4. Mapping in Theory
5. Mapping in Practice
6. Using GIS for empirical analyses
7. Resources
The Earth is *not* flat!
The Earth is *not* flat!
The Earth is *not* flat!
The Earth is *not* flat!
Lambert Projection
Coordinate Reference Systems (CRS)

- Also called Spatial Reference System (SRS)
- Coordinate-based local, regional, or global system used to locate geographical entities
- Defines a specific map projection
- Important to know when joining different shapefiles or geocoded points onto maps

GIS data

- GIS: Geographic information systems
- Storage of a variety of geographic information:
  - Points (e.g., Encina Hall)
  - Lines (e.g., Serra St)
  - Polygons (e.g., Santa Clara county)
GIS data

- **Data**
  - Shape file (.shp): vectorized geographic coordinate ("shape") data (points, lines, polygons)
  - Shape index (.shx): positional index of the feature geometry to allow seeking forward and backward quickly
  - Data frame (.dbf): information associated with each point, line, or polygon
  - Projection info (.prj): contains projection data and information on the CRS used
Outline

1 Motivation

2 Geocoding in Theory

3 Geocoding in Practice

4 Mapping in Theory

5 Mapping in Practice

6 Using GIS for empirical analyses

7 Resources
Let’s do some mapping!

R code
Outline

1 Motivation
2 Geocoding in Theory
3 Geocoding in Practice
4 Mapping in Theory
5 Mapping in Practice
6 Using GIS for empirical analyses
7 Resources
Let’s use some of the GIS tools for actual analyses!

R code
Outline

1 Motivation
2 Geocoding in Theory
3 Geocoding in Practice
4 Mapping in Theory
5 Mapping in Practice
6 Using GIS for empirical analyses
7 Resources
Some useful online resources

- Nick Eubank’s GIS in R scripts
- Making Maps with R
- Natural Earth Data shapefile collection
- Colors in R
- Color Brewer Advice for Cartography
Thank you!

For questions, please reach out to
hlueders@stanford.edu