Introduction to Geospatial Analysis in R Studio

DATA SERVICES
Bobst Library 5th floor
212-998-3434
data.services@nyu.edu
http://guides.nyu.edu/gis

Introductions

- Tell us your:
  - Name
  - Role within NYU (i.e. undergraduate, graduate student, staff, faculty etc.)
  - Why you are interested in using R for geospatial analysis.
  - Your prior experience with GIS software or R Studio (i.e. Are you completely new to GIS software? Have you used other types of GIS software, but are new to R?)

Roadmap

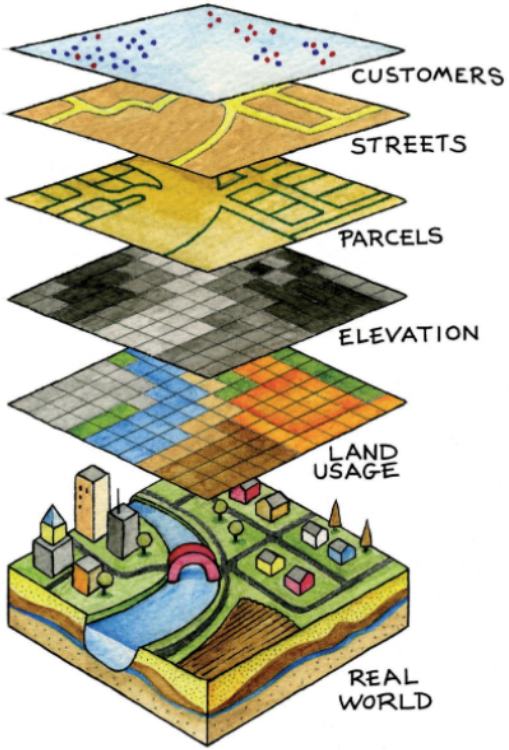
- Brief Introduction
  - What is GIS?
  - What are some examples of GIS software and how does R fit into this universe of GIS options?
- Part 2: Hands-on Tutorial
  - How can we produce a map using GIS packages in R Studio?
  - Basic geospatial analysis in R Studio (time permitting)

What is GIS?

GIS is...

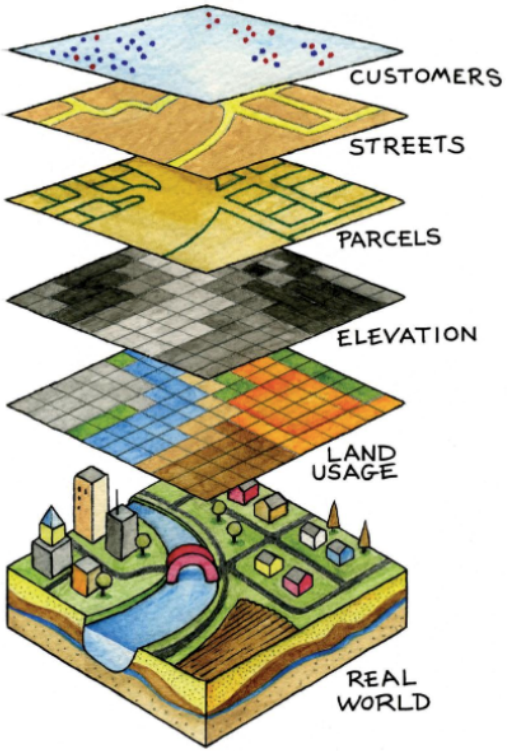
- A geographic information system (GIS) is a platform used to capture, store, analyze, manage, and display various forms of geographically referenced information.
- Today's GIS software is the product of collaboration collaborations between various disciplines like geography, cartography, statistics, computer science, and information science, web design etc.
- GIS is used in dozens of disciplines, ranging from ecology to literary studies to economics.
    - The flexibility of GIS, and its diverse uses, means that by learning about GIS, you acquire a language or framework of analysis that can allow you to make sense of research in a variety of disciplines.

With GIS we can...



- Combine disparate layers of geospatial information to explore and highlight interesting and relevant connections and relationships
- Analyze geospatial information (i.e. calculate distances, find intersections between different data, compute spatial statistics etc.)
- Visualize variables (such as crime prevalence, election results, Census data) in a spatial context

GIS Data Types



- There are two main GIS data types used to represent real-world spatial relationships: vector data and raster data.
- The building blocks of vector data are georeferenced points, which can also be combined to form lines and polygons.
    - In the picture on the left, customers (points), streets (lines), and parcels (polygons) are examples of vector data.
- Raster data consists of georeferenced grid-cells (pixels) that contain data of interest (such as temperature) .
    - In the picture on the left, elevation and land usage are represented using raster data.

GIS Software

- ESRI products such as ArcMap and ArcGIS Pro (proprietary software, which NYU licenses)
- Dedicated GIS packages written for the R programming language (such as sf and tmap).
- Dedicated GIS packages written for Python (such as GeoPandas)
- Google Products (Maps, Earth Engine, BigQuery etc.)
- QGIS (point-and-click GIS software that is the open-source analogue of ArcGIS)

- The advantage of R over Esri products and Google geospatial applications is that R is free and open-source; its advantage over QGIS is that it is relatively more flexible, and better at handling larger datasets and complicated spatial statistics.

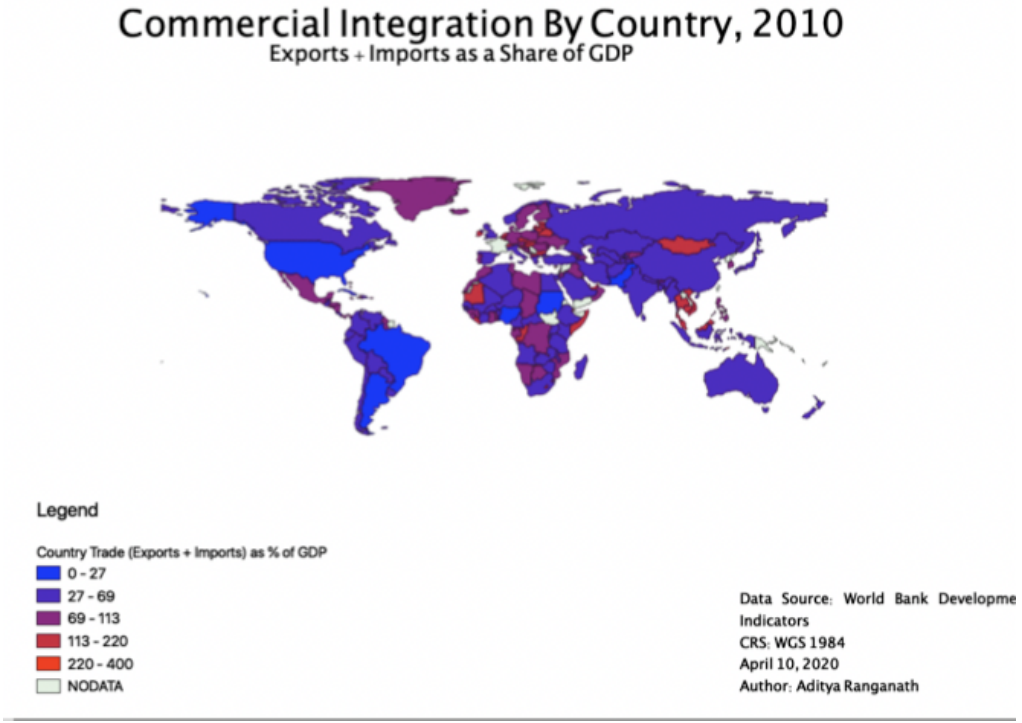
Tutorial: Part 1
Making a Choropleth Map

Objectives

In this part of the tutorial, we will cover the following:

- Downloading and cleaning data for use in R Studio
- Loading a shapefile (a data format used to store vector data) into R Sutdio.
- Joining tabular (non-spatial) data to a GIS layer so that it can be visualized on a map
- Displaying and visualizing data on a map
- Exporting the map for use outside R Studio

More concretely, you will learn how to generate a map that looks something like this...



Part 2

In Part 2, we will cover the following:

- Importing and visualizing a CSV containing latitude and longitude information into R studio
- Querying data in a shapefile's attribute table according to specified criteria
- Joining shapefiles that don't have a common field based on their spatial locations to derive interesting or relevant information