

Intro to GIS

Instructor: Jeff Larson, jal2301@columbia.edu

Time: Fridays 9-12

Location: 511A Journalism

Website: thejefflarson.github.com/columbia-gis/

In this class we'll learn one of the most important tools in modern day data journalism: GIS. I'll assume that you have had a bit of exposure to mapping in previous classes, but I've also incorporated a bit of a buffer in the later sessions in case we move too fast and there are confusing parts.

A typical class will be 2 hours of lecture and an hour of hands on work. You should be able to mostly complete the homework in that time period, and I hope that we all can be collaborative to help each other figure out strategies to complete the assignments.

Grading for the class is pass / fail, with homework being worth 60% and the final project worth 40%. I've structured the homework such that if you complete most of it you'll have the tools necessary to complete the final. Your final project is pretty open ended and doesn't need to be flashy — the best maps are simple.

Week 1: Class outline and Review of QGIS

An intro into what we'll cover in the class and a survey of the state of mapping in journalism. A quick refresher tour on how to view shape files in QGIS and a quick tour of key functionality.

Homework: Install PostGIS, QGIS, GDAL, python, r

Week 2: Working with TIGER/Line Shapefiles

An overview of the most used spatial datasets in news: the Census's TIGER/Line shapefiles. We'll go over the hierarchy of different summary areas that the census

compiles, and the associated spatial data that is available and I'll introduce the concept of map projections — why they are necessary and the different types.

Homework: Make a map using various census sourced shape files.

Week 3: Working with the Census and ACS

This week we'll cover how to join Census decennial data and American Community Survey data to TIGER/Line vector data, and cover a few simple statistical techniques for cluster analysis. We'll also look at external data sources like NHGIS — historical data — and state and local GIS repositories. We'll continue our discussion of projections, and delve a bit into the math behind projections.

Homework: take the [Natural Earth Shapefile](#) and join it on a Census or [World Bank](#) dataset.

Week 4-5: Basic Web Mapping with SVG

Now onto the fun stuff. We'll walk through how SVG works, and after a crash course in javascript we'll use d3 to make a simple map on the web. We'll cover how to add interaction and styling. Finally, we'll look through impressive examples from various newsrooms.

Homework: Make an SVG map with d3 from your joined dataset.

Week 6: PostGIS, QGIS, and GDAL

GIS data can be unwieldy and keeping track of various shape files is often a recipe for disaster. This week we'll cover how to use PostGIS to import many files into one database, and we'll visualize layers in QGIS. We'll also cover how to reproject and clip shape files using GDAL.

Homework: Download the different census geographies of the county you were born in / or the one you live in, import them into PostGIS and put it on a QGIS map. Extra credit: download some ACS data and join it to the tracts in your county to make a choropleth.

Week 7: Spatial Operations and Geocoding

Census geographies don't always line up, for example ZCTAs can and do cross county lines. This week we'll explore how to use PostGIS and its powerful suite of spatial functions to account for that. We'll also cover the GIS skill you'll probably be doing the most as a data journalist: Geocoding. We'll write some python code to geocode against a number of different services.

Homework: Export each layer in your PostGIS database as GeoJSON, but only include a 1 mile radius around the centroid of the county.

Week 8-9: Advanced Web Mapping

Since google maps burst onto the scene in 2005, news organizations have been increasingly using so called "slippy maps" to tell stories. There are advantages to using tile based maps over svg, especially when you have lots of data in a posts database. In these two lectures, we'll explore the Leaflet mapping library to create maps using GeoJSON from our GDAL and PostGIS data sources.

Homework: Using your extracted dataset from last week, create a Leaflet map with many layers.

Week 10: Raster Data: Space Journalism!

The census is pretty great, but ever since the 70s government satellites have been taking pictures of every place on earth every two weeks or so. This data is a true human good, and often enhances our reporting especially when covering the environment. This week we'll cover the fundamentals of Raster datasets, especially pictures taken from Landsat 8, learn how to process them, and how to deal with their massive sizes.

Homework: Find a news worthy event on earth explorer, like a fire, war zone, natural disaster aftermath and process it with landsat-util.

Week 11: Pixels as data, other types of rasters

Not all raster data is image data. This week we'll cover the different types of raster data, and look at interesting examples like the National Elevation Dataset and the national lancer database. We'll also learn how to use GDAL to resize and clip large

rasters for the web. Finally, we'll cover how to use vector data and raster data together.

Homework: Crop and resize your raster from last week for the web. Draw administrative boundaries with d3 or put it on a slippy map with Leaflet.

Week 12-14: Advanced Topics / Final Project

Things I've missed: advanced raster processing; using python and bumpy for more intensive processing; working with vector data in code using shapely; averaging rasters to remove clouds; hosting your own image tiles; webgl?

Final Project: use vector and or raster data to tell a story. Doesn't need to be investigative, the best maps are explanatory.