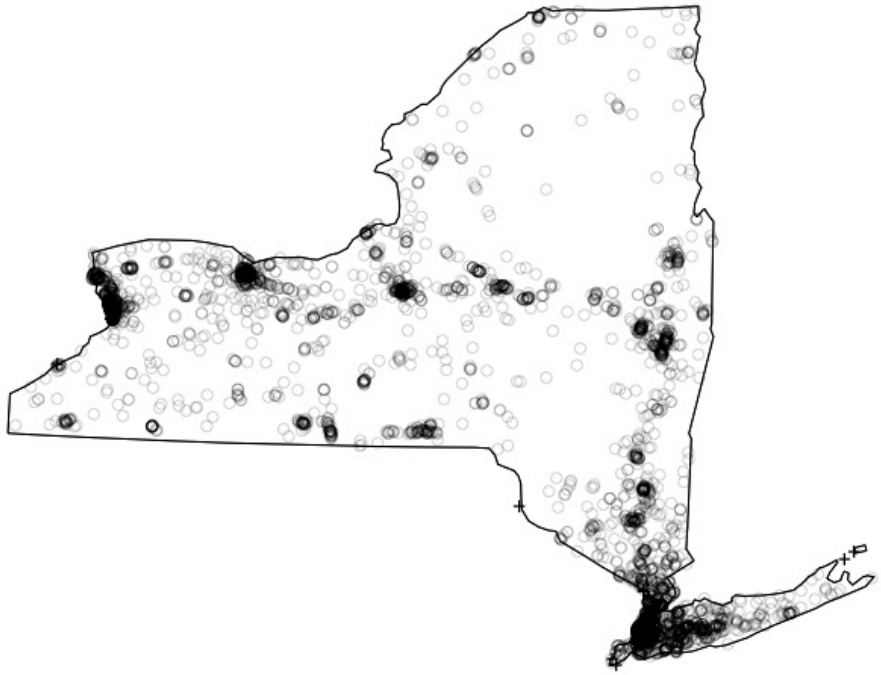


Introduction to Mapping and Spatial Analysis in R

FINTAN MOONEY

Geospatial Consultant
MPH Candidate

Yale UNIVERSITY LIBRARY

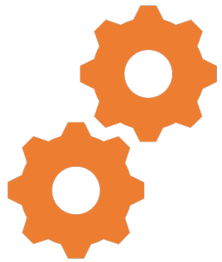
Workshop Structure

- Introduction
- Getting Started with Spatial Data
  - Loading spatial data into R
  - Basic data manipulation with sp/sf packages
- Visualization with Raster and ggplot2
  - Basic raster manipulation with raster package
  - Plotting raster data with ggplot2
- Advanced Visualization
  - Combining vector and raster data
  - Mapping with tmap and leaflet
- Spatial Analysis with spatstat
  - Basic spatstat objects and data manipulation
  - Kernel density analysis
  - Local and global Moran's I
- Conclusion

Why R?

Open-source

Advanced spatial statistics packages (spatstat, CARBayes, etc.)

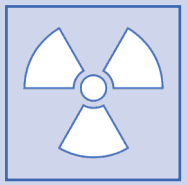


Automation and looping

Introduction

Make sure you have the R markdown file and data downloaded onto a computer folder

We will be working with multiple packages in this session, including sp, ggplot2, raster, tmap, leaflet, tigris and spatstat



We will map and analyze remediation sites (hazardous waste, Superfund sites, etc.) across New York and PM2.5 in NYC

DATA

Remediation Site Boundaries:
https://www.dec.ny.gov/chemical/102009.html

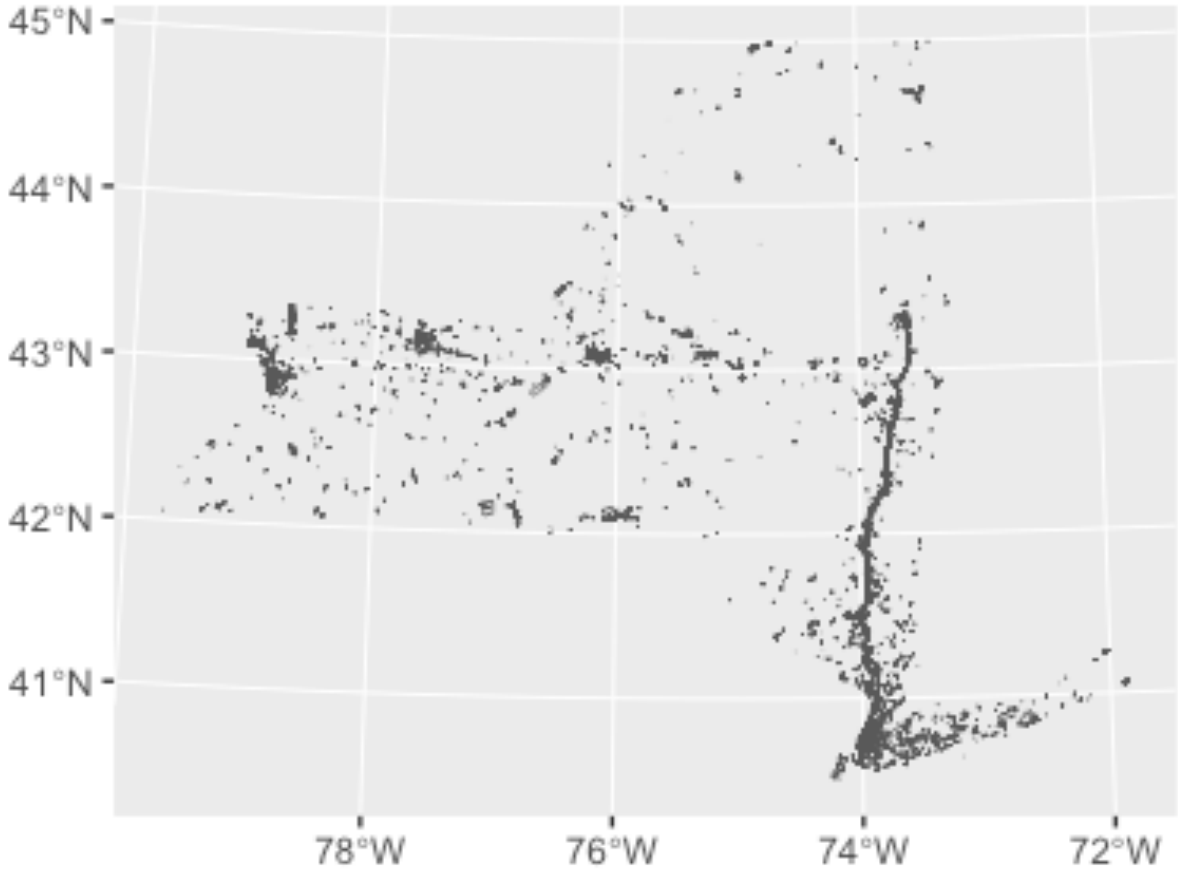
Air Pollution (NYC):
https://data.cityofnewyork.us/widgets/q68s-8qxv

Loading Data

with sf and ggplot2

If we have our shapefile in a folder in our working directory, we can read it in via st_read from the sf package

We can then plot it using ggplot2

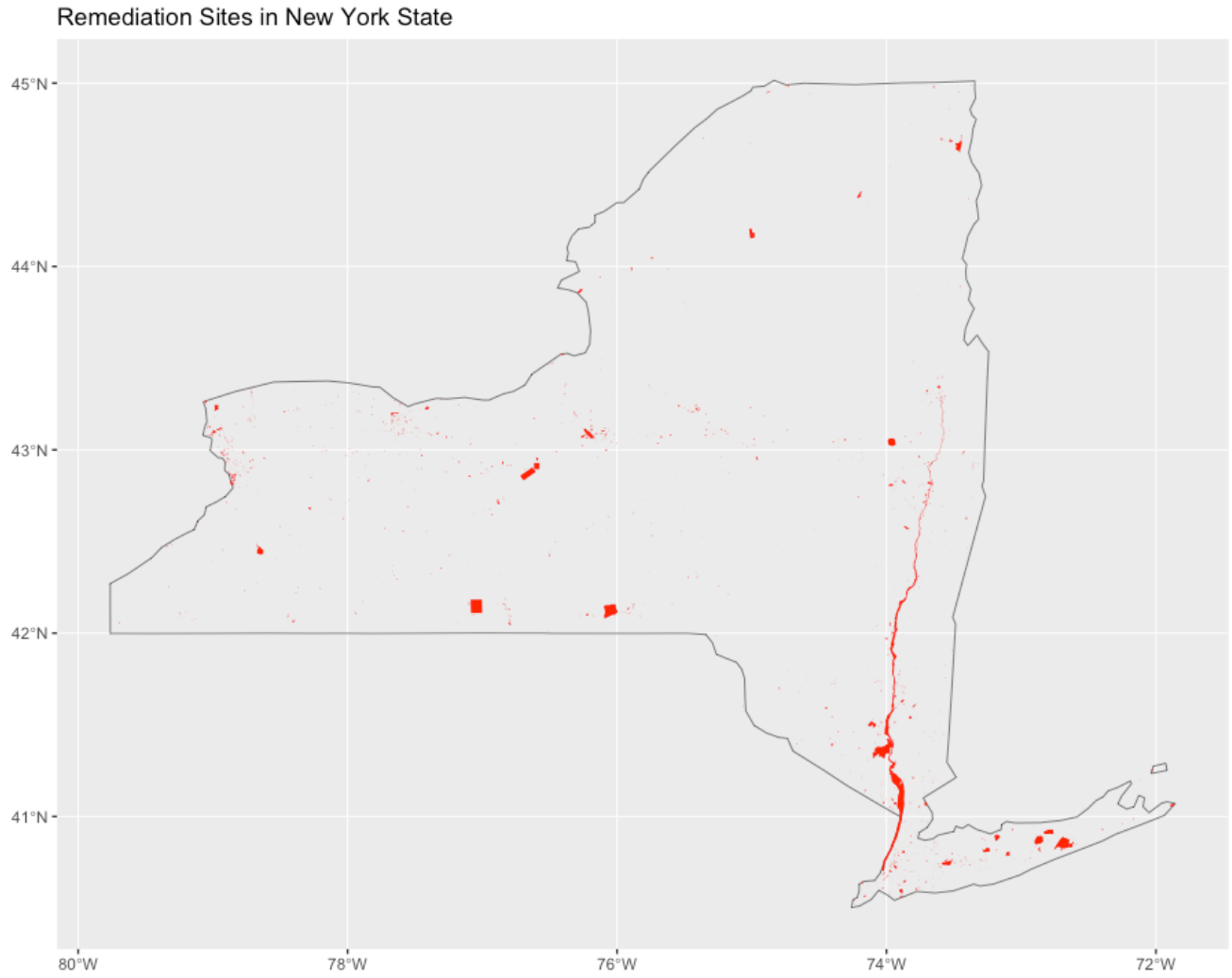


Adding in Detail

with tigris and ggplot2

The tigris package allows for the download of spatially referenced census data

It is useful for getting boundaries of states, counties census blocks/tracts, etc.



Tip If we have tabular data with census information we can easily map the data by merging it with a tigris shapefile

Difficulties in R

- We need to make sure our data layers are in the same coordinate system
  - st_transform from the sf package
  - projectRaster from the raster package

- Geographic coordinate system
  - Defines where the data is located on the earth's surface
  - Use when visualizing the globe, continents, and large countries

- Projected coordinate system
  - Transforms the spherical surface onto a two-dimensional plane
  - Countries have their own PCS
    - The U.S. commonly uses UTM

Merging Vector Data

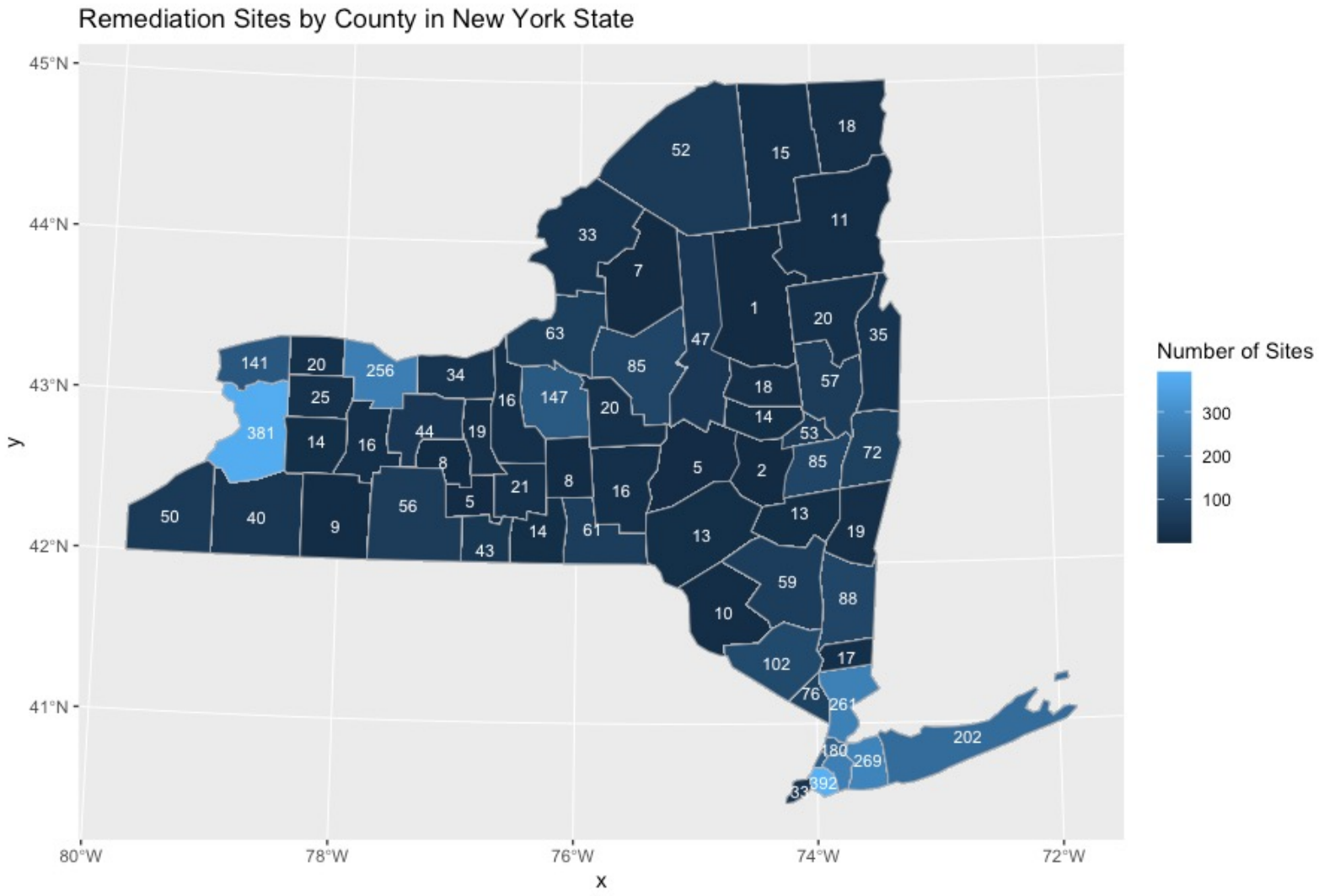
with sf and tigris

st_as_sf () function converts various data formats into an 'sf' object

st_crs () function allows us to check and match coordinate systems

st_join () function is used to join spatial data

st_make_valid () function corrects spatial datasets that have errors due to overlapping polygons, self-intersecting lines, or gaps between polygons



Tip: Make sure the coordinates systems match between layers. Use the st_transform function to project the census data into the remediation sites projected coordinate system

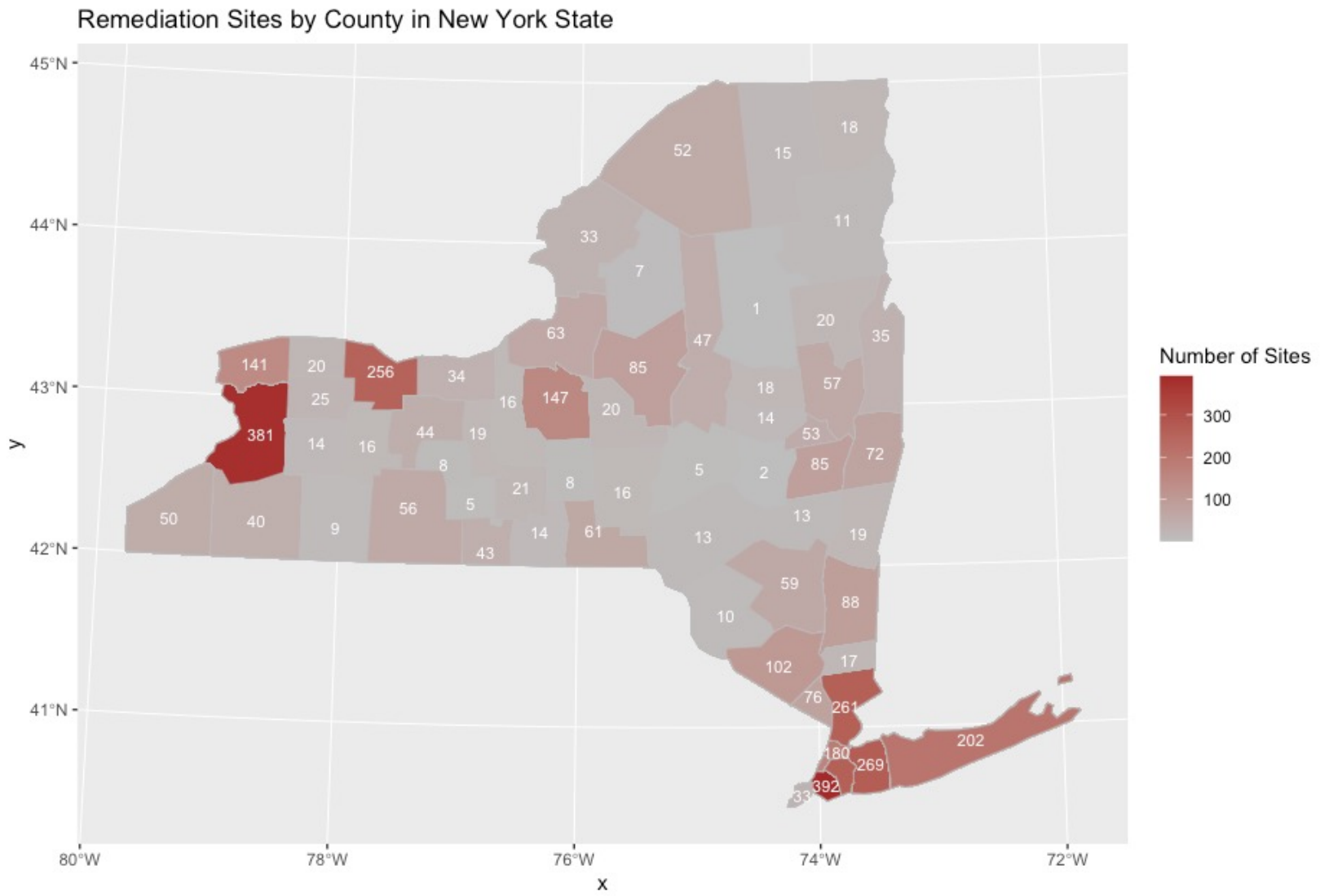
Alter Symbology

Using the scale_fill_gradient function in ggplot

scale_fill_gradient() produces a two-colour gradient

scale_fill_gradient2() produces a three-colour gradient with specified midpoint

scale_fill_gradientn() produces an n-colour gradient

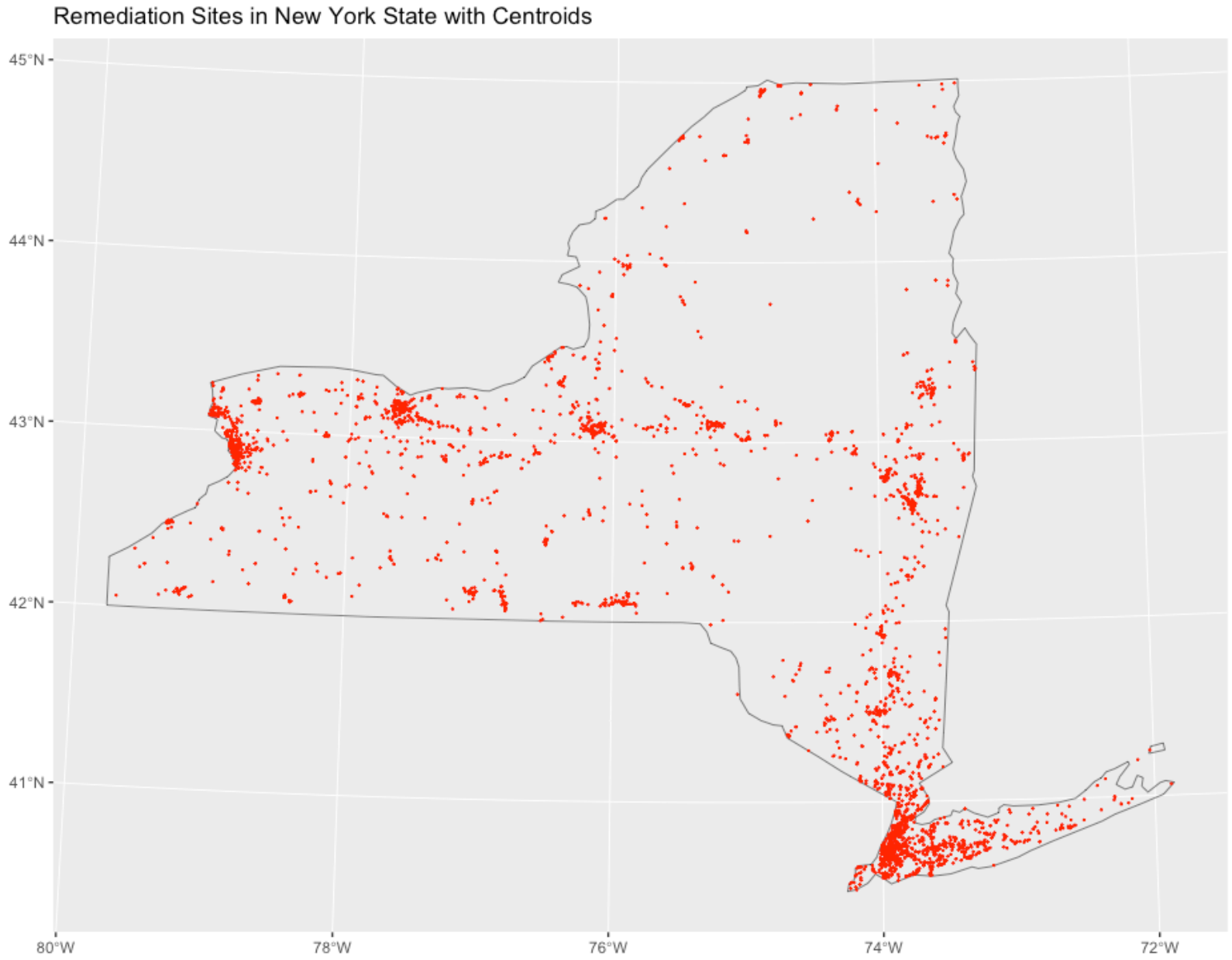


Mapping point data

Using the st_centroid function from the sp package

The sp package can also be used to map point data

We can create point data from our polygons by finding the centroids of each remediation site



Reading Raster Files

Many of the same functions in ArcGIS can be utilized in R with the raster package

Useful functions from the raster package:

- plot() - creates a plot of the raster data.
- extract() - extracts values from the raster at specified locations.
- crop() - crops the raster to a specified extent.
- resample() - changes the resolution of the raster.
- projectRaster() - changes the projection of the raster.
- reclassify() – reclassify raster values

Merging raster files:

- merge() - merges two or more rasters into a single raster object. This function works by aligning the rasters based on their extents and resolutions, and can handle rasters with different projections.

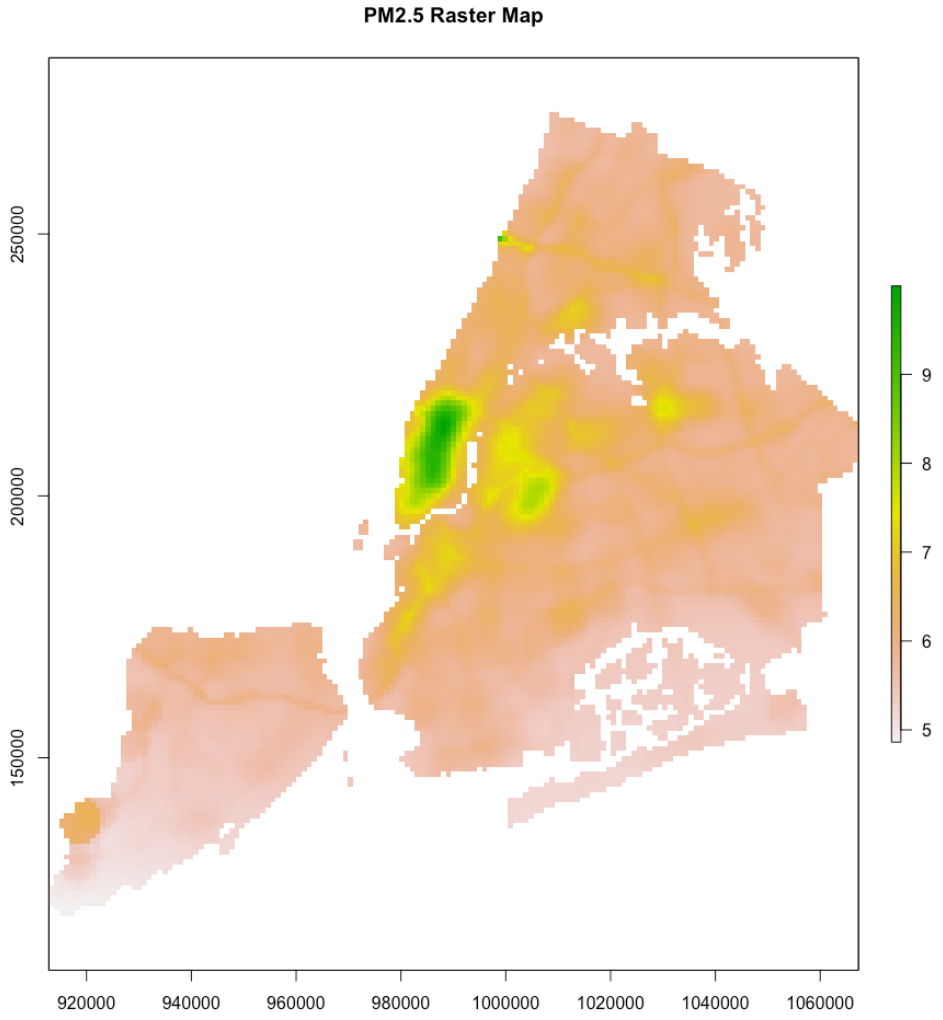
- mosaic() - merges two or more rasters into a single raster object, but requires that the rasters have the same extent and resolution. This function is faster than merge(), but can only be used with rasters that meet these criteria.

- stack() - creates a RasterStack or RasterBrick object from two or more rasters, stacking them on top of each other. This function does not merge the rasters into a single layer, but instead creates a multi-layer raster object.

Map a Raster File

Using the raster package and plot function

Terrain.colors was used for symbology



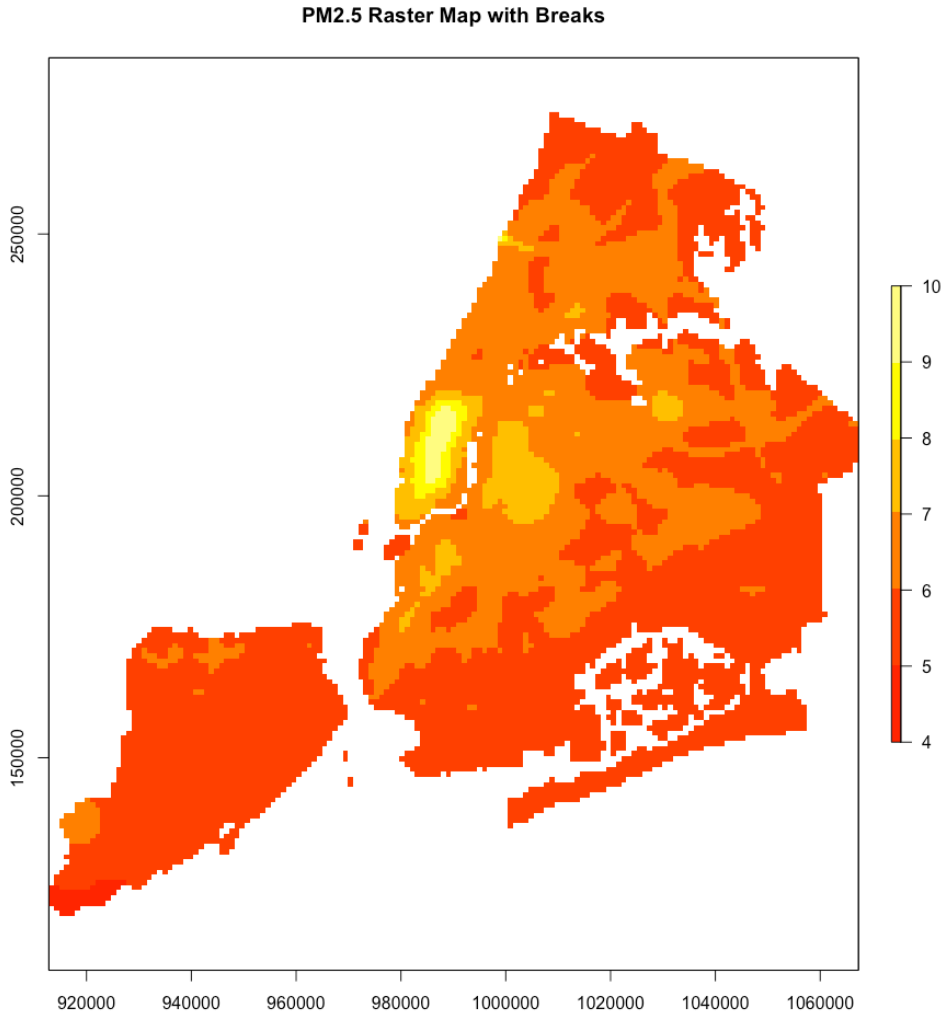
Tip: Change the parameters to fit the title and legend onto the map.
For example: par(mar = c(5, 4, 4, 2) + 0.1)

Modifying Classification

Using breaks

Adding in breaks helps us to redefine classification and symbolization

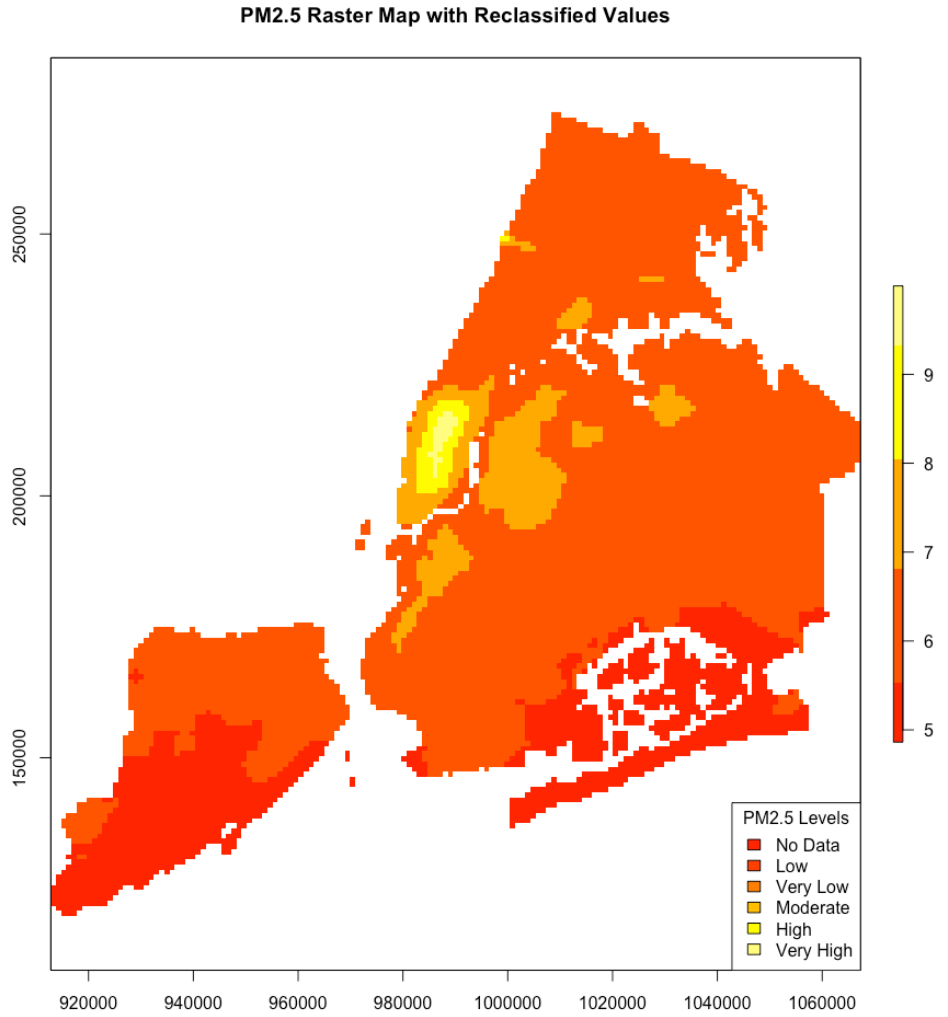
Use the hist () function to find reasonable breaks



Modifying Classification

Using reclassify()

To use the reclassify () function we need to define a reclass table using cbind ()



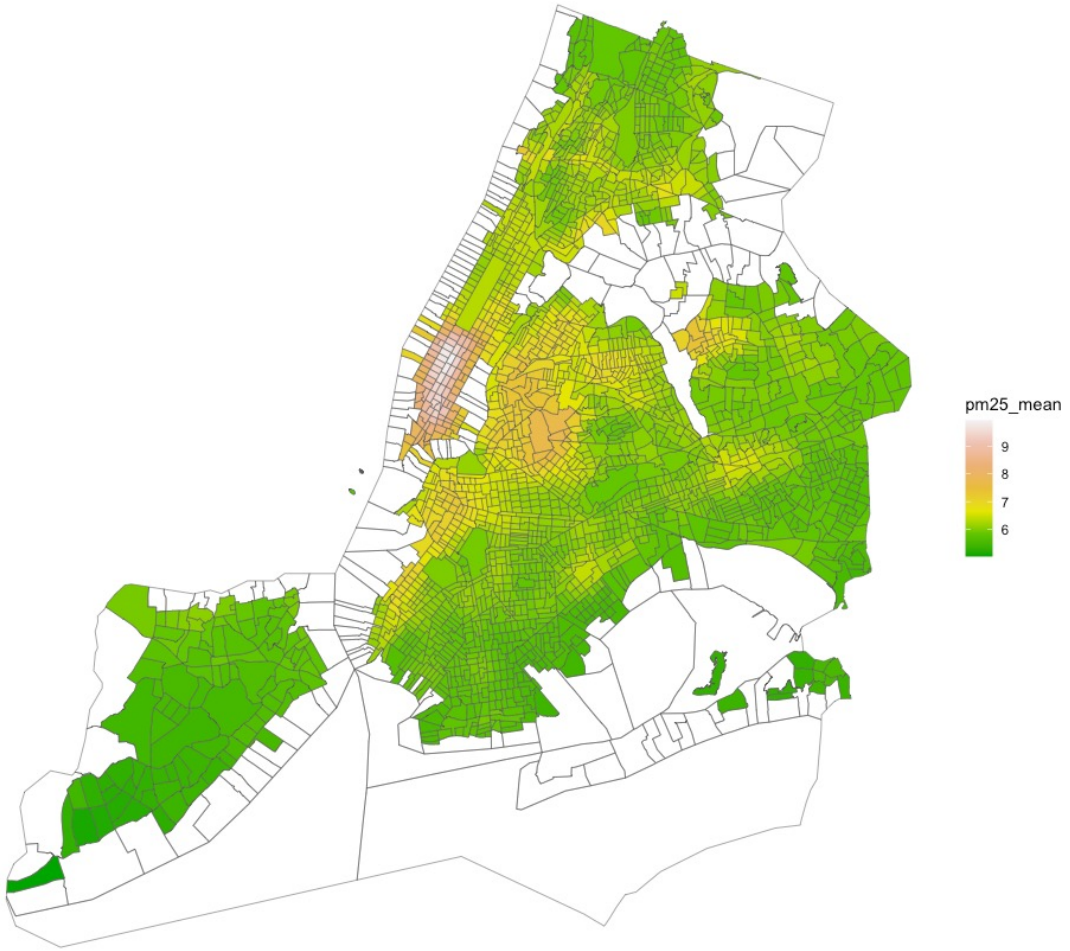
Tip: Need to add a custom legend to plot ()

Combining Vector and Raster Files for Spatial Analysis

Using the extract () function

The sapply function was used to get the mean PM 2.5 concentration for each census tract

The extract function can take a few minutes to run

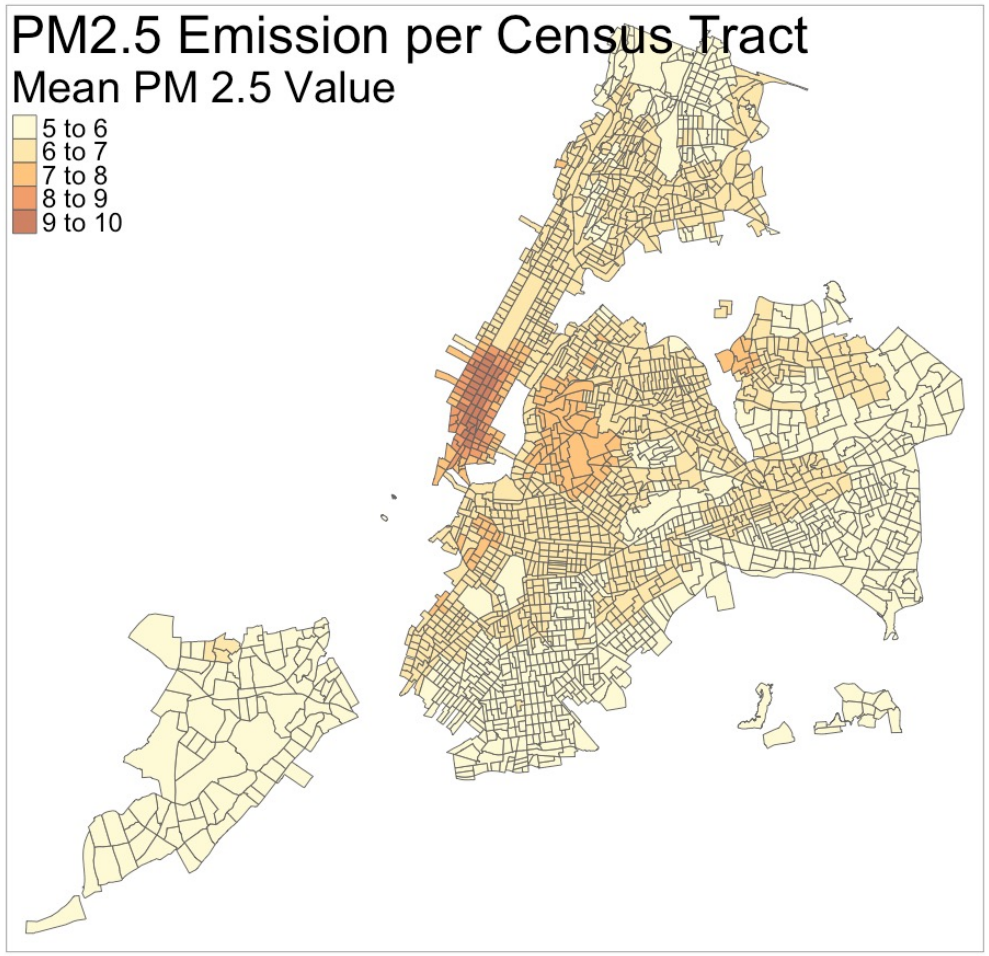


Tip: In the next slides we will learn how to make this prettier

Visualize with tmap

Functions needed:

- tm_shape ()
- tm_polygons ()
- tm_layout ()



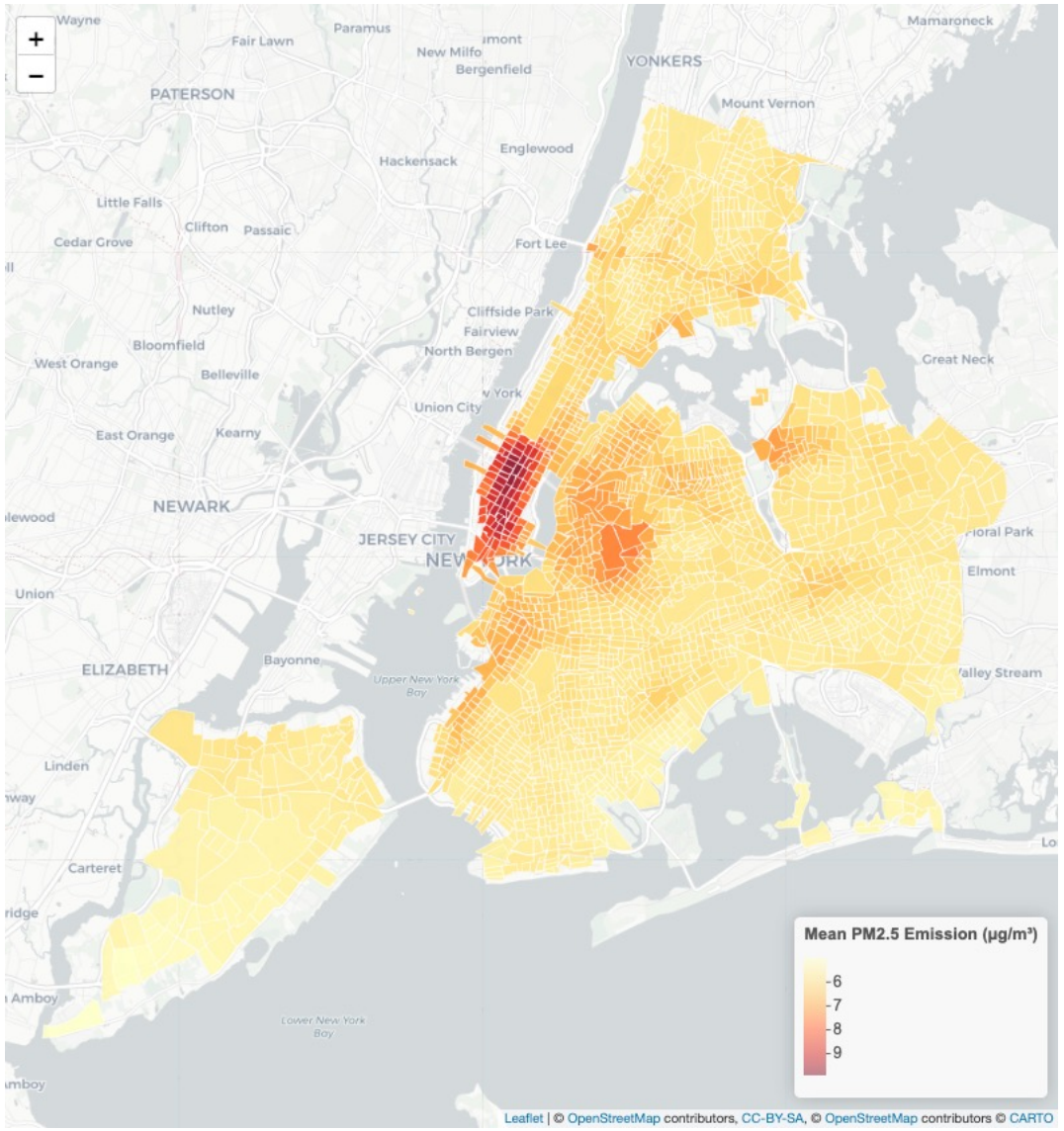
Caption: Lots of white space ☹

Visualize with leaflet

Allows for interactive maps!

Notice the ability to add in a basemap through addProviderTiles()

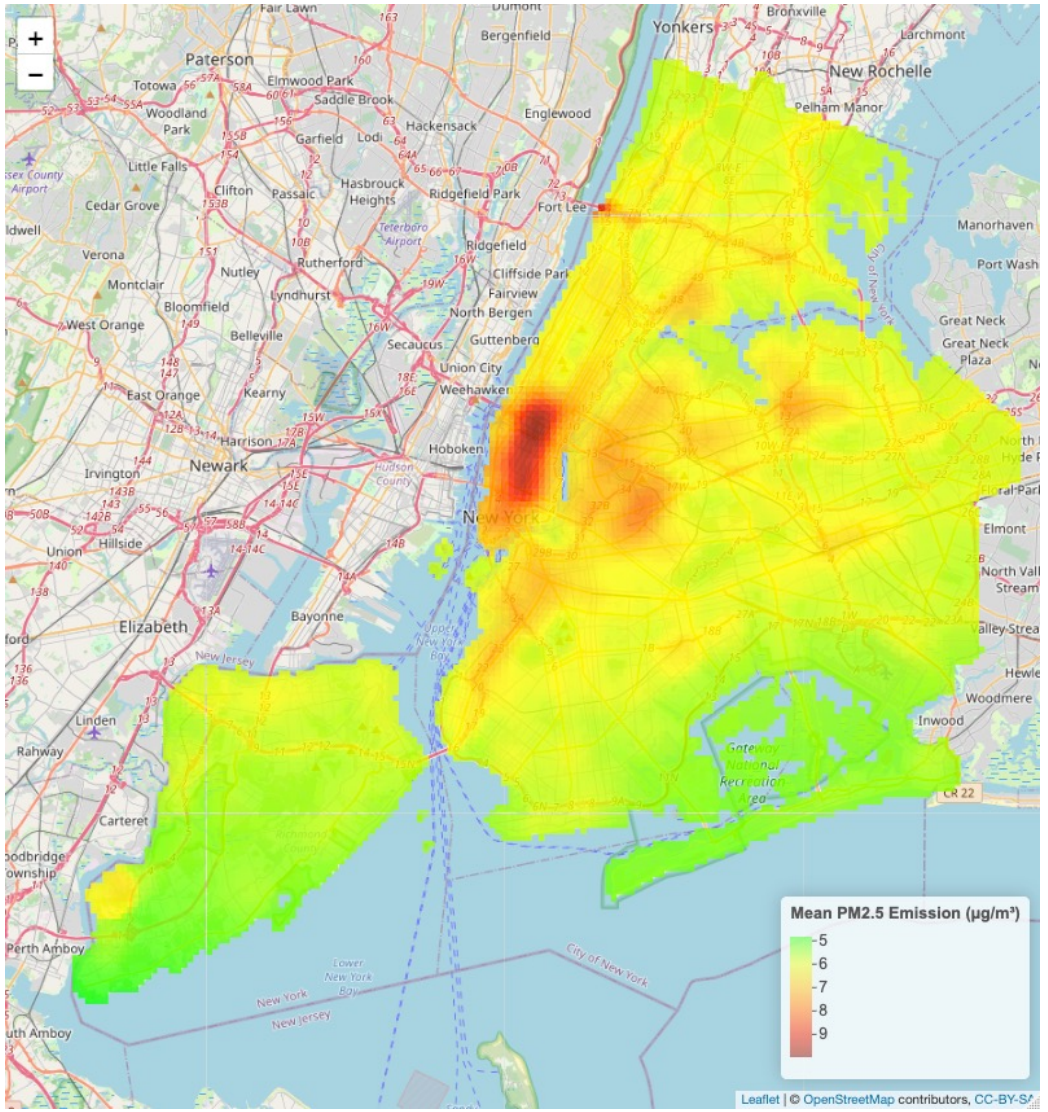
Also, we can use setView () to create a default extent



Problem: We will need to interpolate to give census tracts with NA values the nearest neighbor. For now, that is too advanced, and it is much easier to interpolate via ArcGIS.

Visualize Raster with Leaflet

Need to use addRasterImage()



Tip: It is fun to mess around with the palette

Point Pattern Analysis

Using the Spatstat Package

spatstat provides a wide range of functions for analyzing point patterns, including data input and manipulation, visualization, exploratory data analysis, statistical modeling, and hypothesis testing.

- ppp(): This function is used to create point pattern objects. It takes as input the x and y coordinates of the points and returns a ppp object.
- owin(): This function is used to create a spatial window object. The window defines the spatial domain of the point pattern.
- plot(): This function is used to create a plot of the point pattern or other spatial objects. By default, it creates a scatterplot of the points, but it can be customized with various arguments.
- density.ppp(): This function is used to estimate the intensity (density) of the point pattern. It returns a density estimate as a function of spatial location.
- Kest(): This function computes the K-function, which is a measure of the spatial dependence of a point pattern. The K-function is used to detect clustering or regularity in a point pattern.

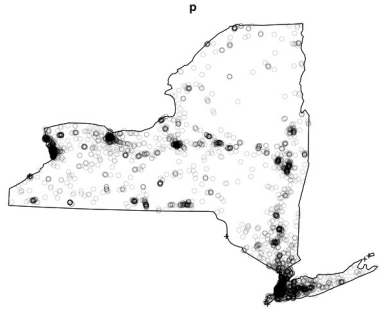
Kernel Density

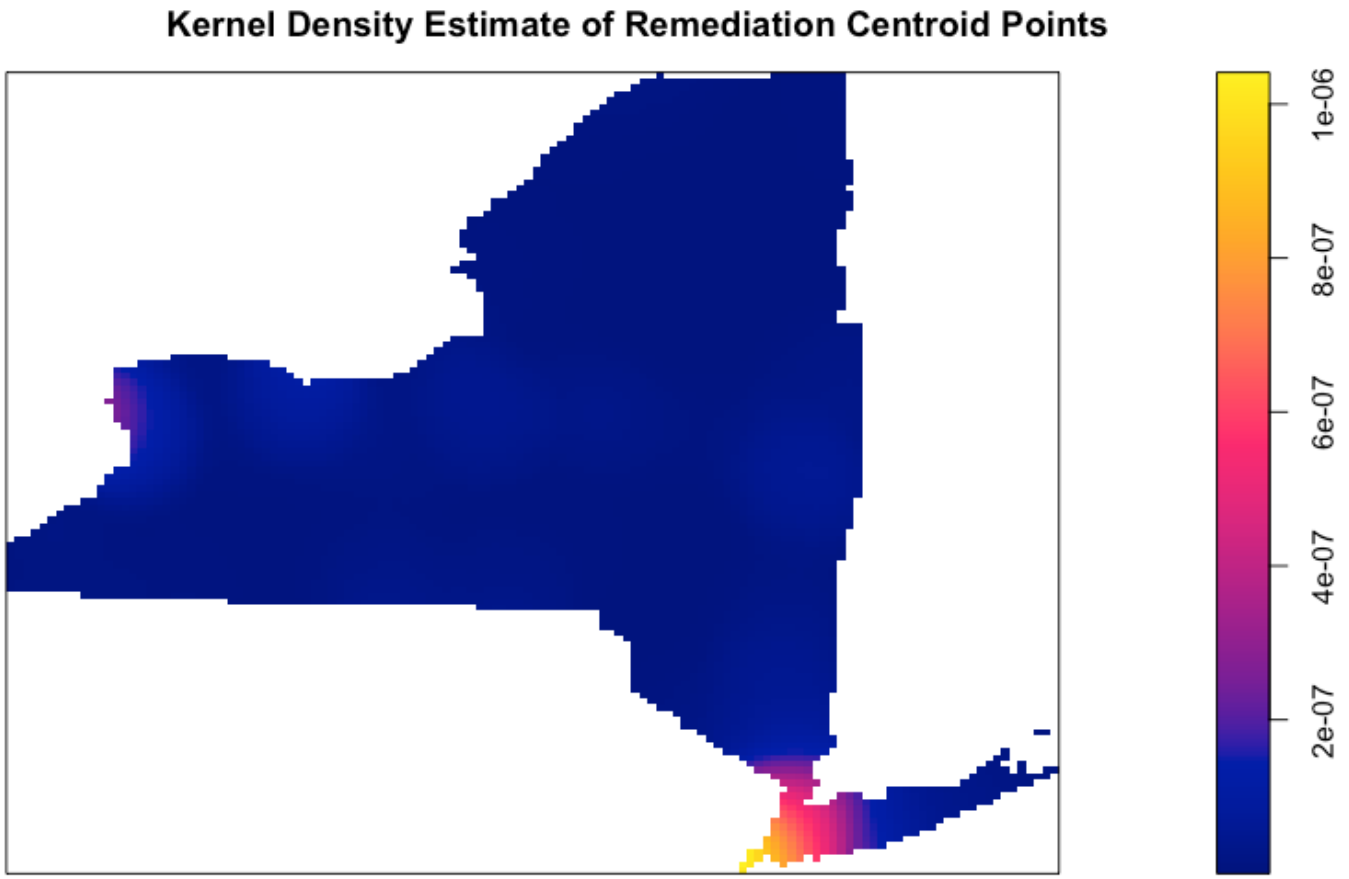
Using the density.ppp function

Adjust sigma to vary spatial influence of points

The choice of smoothing kernel is determined by the argument kernel:

"gaussian", "epanechnikov", "quartic" or "disc"

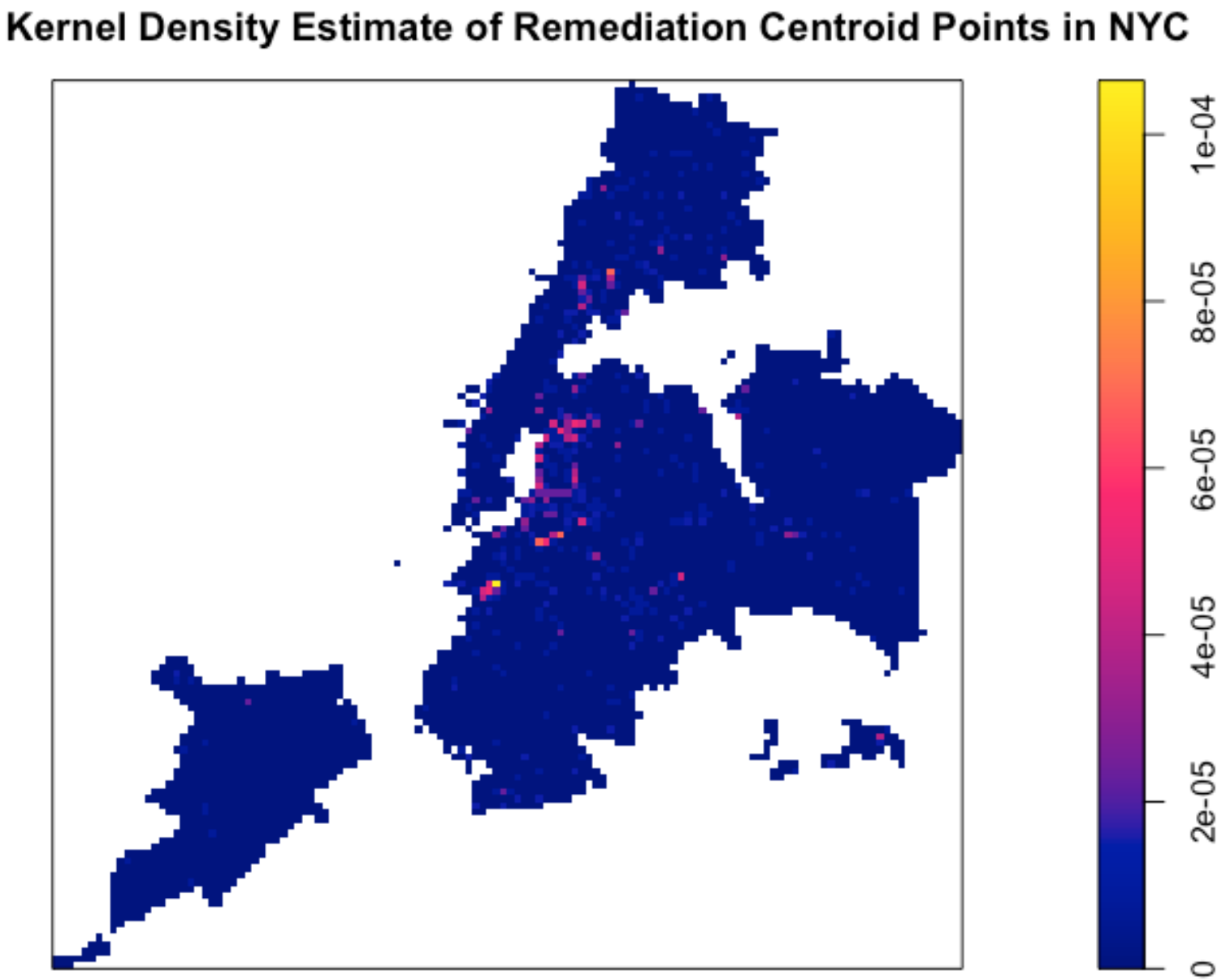




Observation: Remediation points are densely clustered in NYC

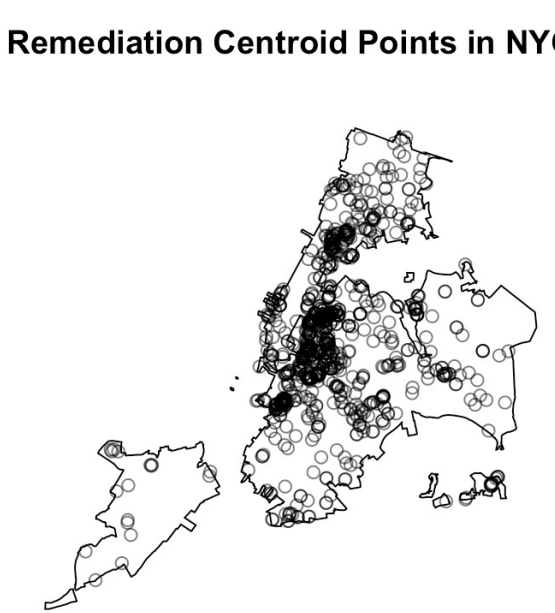
Zoom into NYC

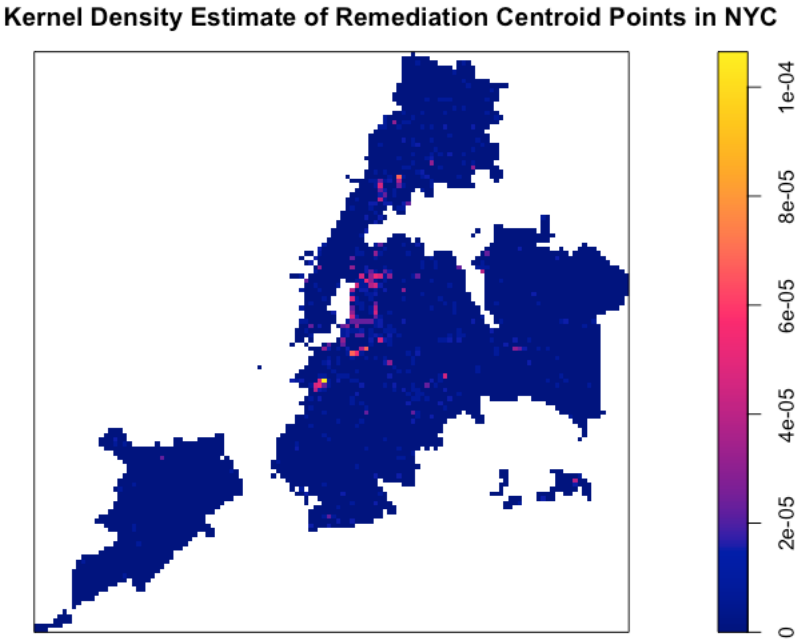
Need to create a spatial window for NYC



Tip: Make sure the shapefile for the as.owin () function has been projected

Quick Analysis of Clustering

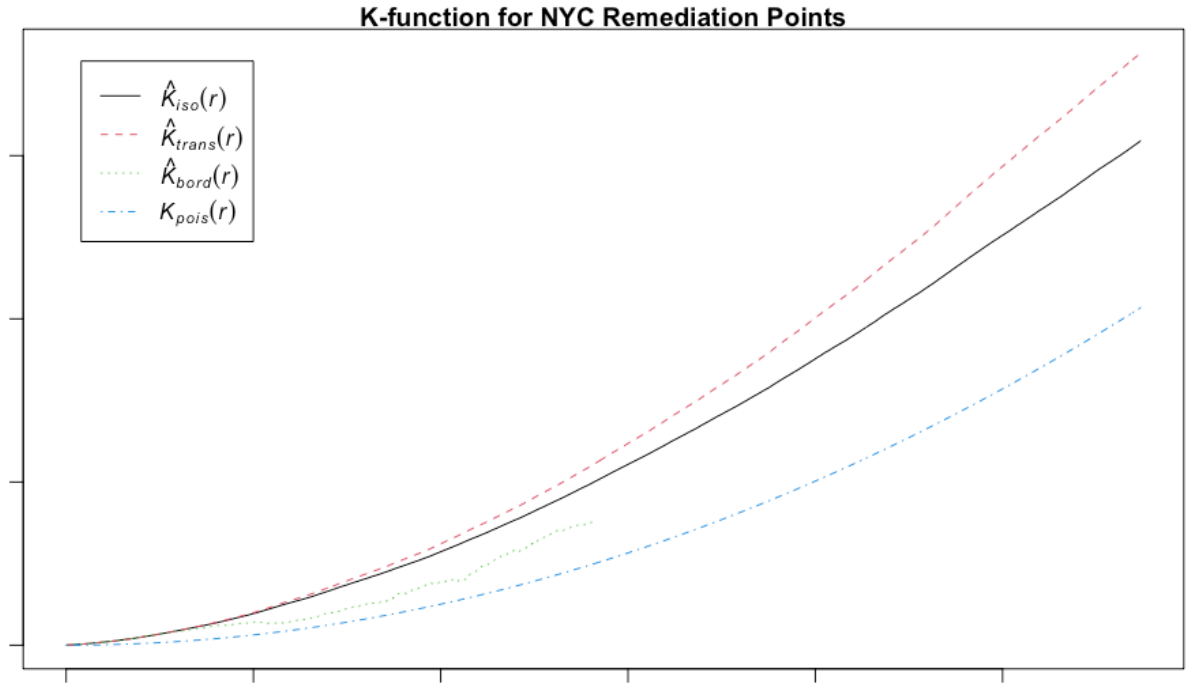




Using the Kest() function

Interpretation: the expected points (solid line) is below the observed points (red dash), indicating that the points are more dispersed than expected

It is important to consider spatial scale, if we clipped the points by borough, we may see more clustering



Extended Spatial Analysis with R

- Geographically Weighted Regression (GWR)
  - A local statistical technique that allows the relationships between dependent and independent variables to vary across space
  - Can help identify spatial heterogeneity and improve model performance
  - Available in the spgwr package

- Local Moran's I
  - A statistic that measures the spatial autocorrelation of a variable at a local level
  - Can help identify clusters of high or low values and spatial outliers
  - Available in the spdep package

- CARBayes Package
  - Allows for fitting Bayesian spatial regression models

Spatial autocorrelation refers to the degree to which the values of a variable observed at nearby locations tend to be more like each other than to those at more distant locations

Fin's Mapping in R Tips

- The sf (vector) and raster (raster) packages are the easiest to use for basic data manipulation and map presentation
- The leaflet package is the best for creating good-looking and interactive maps!
- Make sure you project your spatial data, especially before conducting any sort of analysis
- Do not attempt to read in files that are named with a number at the start– add a letter or word
- Sometimes for projection and basic data manipulation (clip, buffer, union, append, project, etc.) it is best to stick with ArcGIS

Thanks for coming!

- Please email fintan.mooney@yale for any questions or extended resources

Yale UNIVERSITY LIBRARY