# Project 2 - Areal Data

ST 433/533: Applied Spatial Statistics

Your project report should include a title, introduction, methods, results, discussion. The report should be approximately 3 pages long including relevant figures. If you need more than that, see me! Computer output must be summarized appropriately and presented as it would in a journal article, (e.g., tables – I don't want to see R output!) and figures must include proper labels, captions, legends, etc.

Explore county level data across the state of North Carolina. The dataset `NCcounty2024data.csv` consists of demographics, education, health, and economics data for each county. See here for more details: https://indd.adobe.com/view/3f0db37e-7e56-4b42-9a16-b6a218100a6a.

1. Conduct statistical tests for spatial dependence for three variables of your choice. Report the results of your tests both numerically and in context. Make sure to include details on the proximity matrix you chose.
2. Conduct a spatial areal regression analysis using any response variable of interest. Provide motivation for your choice in response variable and the predictor variable(s) you include. Feel free to use the internet to obtain any other county level data that you find interesting. Write out your model and discuss how it was chosen. Interpret the model in context of the problem. Make sure to assess your residuals for spatial dependence. Be creative and make it interesting!
3. Produce at least one spatial map of the county level estimates from your model in part 2.

The following is code provides a nice starting point for your analyses:

```
library(geodata,sp,spdep,spmodel)

## Load in the demographic data and county data for North Carolina
ncdata=read.csv("NCcounty2024data.csv",header=T)
load("nc_counties.Rdata")

## Plot the counties
plot(nc_counties)

## Names of the counties
nc_counties[[7]]

## Obtain neighborhoods given the polygons
nc_neigh=poly2nb(nc_counties)
```

In making a map, you can add a variable or (any county-level estimates) as a variable to the object `nc_counties` and use the `spplot` function. For example, the following produces a map of the median age for each county across North Carolina. Obviously `ggplot` can also be used to make nicer figures, this is just an easy and quick way to visualize your data.

```
nc_counties$age=ncdata$Median.Age
spplot(nc_counties,"age")
```