

## Lab 5: Working with Geospatial Data

For this week's lab, we are going to look at college going outcomes for California school districts. We will be using several files:

- "DistrictAreas2122.shp": Shapefile of the school district boundaries
- "District College Outcomes.csv": College outcomes by school district
- "college locations.csv" College locations

### Part I. Cleanup

1. Read in the District Areas Shapefile. Transform the CRS to CA Albers Equal Area.
2. Read in the District College Outcomes csv file. Convert the CDCODE to character. Create a variable that represents the CSU going rate. Create a variable that represents the UC going rate.
3. Left Join the District Areas sf object to the District College outcomes.
4. Read in the college locations csv file. Convert the data frame to a sf object. Transform the CRS to CA Albers Equal Area. Create a sf object for campuses that are part of the CSU system. Create a sf object for campuses that are part of the UC system

### Part II. Analysis

1. Create a map of CSU going rates by district with the CSU locations overlaid.
2. Create a map of the UC going rates by district with the UC locations overlaid.
3. Discuss the relationship between college proximity and college going rates in the two maps. Which do you expect to display a stronger spatial relationship and why?
4. Let's focus on school districts in Riverside County. Find the following:
  - a. The largest school district by area (sq miles)
  - b. The largest school district by #enrolled students
  - c. The school district with the highest CSU going rate
5. Calculate the distance between the Riverside School District Centroids and CSU San Bernadino (the closest CSU). Is the school district with the highest CSU going rate the closest to CSUSB?
6. Create one additional visualization of your choice that is related to college going rates (map or otherwise). Discuss the findings.

### Writeup Guidelines:

- Your writeup should use complete sentences. Recommended word length: 300 words.
- Include the R script you used to answer the questions. The code should be well commented so that the reader understands what your code does. Your code should use concepts that we have covered in lecture. There are many ways to the correct answer, but we are trying to reinforce the approaches we cover during lecture.
- If ChatGPT was used, include the link to your chatGPT conversation regarding the lab. The conversation should be titled "[Name] Lab 5". You can see an example conversation here: <https://chat.openai.com/share/2e45a859-52f9-4477-a7a3-f6e272b8619e>

**Submit your writeup through Canvas by Sunday, 11:59pm.**