Practicum 2

Data Wrangling and Visualization

Prof. Jack Reilly

Fall 2026

Overview

You are to draw a series of maps of the State of New York, highlighting particular identified features of the state. You have the following data made available to you:

- `cb_2024_us_county_500k.shp` - A shape file that includes shapes for all counties in the United States. Variables include:
  - STATEFP - State portion of FIPS code
  - COUNTYFP - County portion of FIPS code
  - GEOID and GEOIDFQ - County identifier (combined FIPS code)
  - NAME and NAMELSAD - Name of county
  - STUSPS and STATE_NAME - State abbreviation and name
  - ALAND - Land area of county (sq meters)
  - AWATER - Water area of county (sq meters)
  - geometry - County shape (multipolygon)
- `nyshighpoints.csv` - a data file containing information about county high points in New York State. Variables include:
  - County - name of county
  - High_Point_Name - name of high point
  - Elevation_Feet and Elevation_Meters - elevation in feet and meters
  - Latitude and Longitude - latitude and longitude
  - Data_Quality - whether the latitude and longitude are exact or estimated[1]

In addition, you will have data to use from online sources, including the US Census and the Harvard Dataverse Replication Archive.

R Work

[1] Data in this file is estimated by Claude, and is thus for class purposes only. Verify before using elsewhere.

1. The file `cb_2024_us_county_500k.shp` is a shape file that includes shapes for all counties in the United States. Use it to draw a map of the counties of New York State (and only New York State - do not include counties from neighboring states).
2. Using the same file, which is the largest county, by area, in New York? Which is the smallest?
3. Use `nyshighpoints.csv` to plot the location of high points of every county in the state of New York over the base county map.
4. There is a somewhat common scale that cartographers use to indicate elevation, where low lying areas are green, middle height areas are yellow to brown, and high areas are white and sometimes red. (See, for instance, the National Geographic map highlighted here.) Use a color scale to mark the high points according to their height. Make sure to have at least four different colors included in your scale.

   Tip

   Note that I am **not** asking you to color in elevation for the entire state - I'm just asking you to color the points that indicate the high points of the county.
5. What are the highest and lowest high points in the state?
6. Label the highest high point on a map.
7. Use the `tidycensus` package to load county population in New York. Shade in each county according to its population. Use a light yellow to dark orange scale.

   Tip

   Remember that you can load data from `tidycensus` with or without geography.
8. Calculate population density for each county using population and land area. Shade in each county according to its population density. Use a light lavender to dark purple scale.
9. What is the largest county, by population, in New York? What is the smallest? What is the most dense county in New York? What is the least dense?
10. **Bonus:** Use the Algara and Amlani's county level data to create maps of two-party vote for each county in New York in the 2000, 2004, 2008, 2012, 2016, and 2020 elections. Show vote by partisan lean, where more Democratic voting counties are bluer, more Republican voting counties are redder, and evenly split counties are white. You may discount any third party votes.

    Tip

    Remember there are multiple ways to plot polygon-based political and social maps, in both base R and the tidyverse. To first approximation, there is the *pre*-`sf` data management style, where each vertex in the map would be a row/observation of your data, and there is the newer `sf` style of mapping, where

each polygon is a single row in your data with a multipolygon column/variable containing geographic information.

If you do not have an `sf` object, you need to use fundamental mapping aesthetics (like the `mapping = aes(...)` argument in `ggplot`) to draw your map. If you already have an `sf` (simple features) R object, however, you can use the simpler `geom_sf` in ggplot to begin drawing your map (or use appropriate `plot()` commands in base R).

Writeup Work (Quarto)

Answer the following questions in your Quarto write-up.

1. Include the map of New York counties and color-coded high points, with the highest point labeled. Which are the highest and lowest high points, and what are their elevations?
2. Include the map of county population by county. What are the most and least populous counties, by name?
3. Include the map of county population density. What are the most and least densely populated counties, by name?
4. **Bonus:** Include county maps of presidential elections in New York, by county. (I recommend doing this in a grid on a single page.) What do you notice about vote trends in New York over that time? What change and continuity has there been in partisan voting trends?

Other Details

- This Practicum is due **Thursday, November 19, by class time**. *As usual, we will cover the practicum in that class, so late work is not accepted.*
- The practicum **IS** open-book, open-note, and open-internet, but it **IS NOT** open-human. In short: you can use any resource you want, so long as that resource does not involve asking another human a question. *(The only exception is that you can ask the professor clarification questions.)*
- If you use AI, you must also turn in a record of your prompts as a plain-text `.txt` file.[2] You **MAY** use AI, but *only* to ask questions as you design your own work. You **MAY NOT** feed the entire practicum itself into a chatbot or other AI tool.

[2]This record does not have to be word for word, but it should encapsulate what you did. For instance, just saying "I used Claude on this practicum" isn't ok; you should tell me what code snippets you used Claude to help you with.

- Make sure your write-up document and code script files are **cleanly formatted**. You will be evaluated both for the accuracy of your output and the clarity of your code.
- Follow all the data and style guidelines we have discussed in class. Your `.R` files — once I change the working directory, if necessary — should properly execute all commands needed to reproduce the results of your practicum, and should do so without any errors. If your `.R` file does not do this, the relevant answers will be considered wrong. You may also use a `.Rproj` file to obviate the need to set a working directory.

Submission

Turn in 4 things:

1. Your R script(s).
2. A quarto file that addresses or answers the steps above and includes your graphics. Your quarto file can include your R code or your R code can remain separate in your R scripts, with the quarto file reading in figures from project directory.
3. A compiled quarto file, as a PDF, that includes your maps and answers to relevant questions above.
4. Your plain text .txt file, identifying if and how you used AI, and which AI you used. If you did not use AI, you still must turn this file in; all you need do is write in it, "I made no use of AI for this assignment."

Make sure that your quarto writeup follows the order of the writeup file specified above and clearly indicates what map is what.

> **Important**
>
> To submit your assignment to Blackboard, submit a clickable link to your cleanly organized GitHub repo with all of the above materials.

Good luck!