Data Science in Education: HW #4

In this assignment, you will recreate the plots in the school-level aggregate data analysis from the Chapter 9 Walkthrough using data for South Carolina schools. The dataset is the 180-day active student headcounts by school for gender, ethnicity, and pupils in poverty for 2021-2022 available at https://ed.sc.gov/data/other/student-counts/active-student-headcounts/. This is the dataset that was used to construct choropleth maps in Lab 6 and is posted to Canvas under both assignments.

Setup

- Create a project in RStudio for this lab.
- Download the school headcount data file from Canvas and move it to your project folder.
- Create a new R Markdown document with HTML specified as the output format.
- In the setup code chunk:
    - Delete the `include = FALSE` option and add the `message = FALSE` option.
    - Load the packages you will need for this assignment: `readxl`, `tidyverse`, and `dataedu`.
    - Read in the school headcount data file using the `read_excel` function and assign it a name. As in Lab 6, use the `skip` argument to skip over the formatting in the first few rows. In this analysis, we are not interested in regional differences, thus we can keep the charter and other special schools at the bottom of the data file. However, the last two rows of summary information need to be omitted, and thus this time you should use `n_max = 1219`.
    - Process the school headcount data by dropping the male, female, and two missing columns and renaming the columns to convenient names.

Exercises

In your R Markdown file, delete the sample code and text. Then add in R code chunks and commentary as necessary to complete the exercises below. Include appropriate headings, hide any messages and warnings, and adjust font sizes in visual displays as necessary to create a polished report. When you have finished working through the exercises, knit your document and upload the resulting HTML file to the HW #4 assignment in Canvas.

1. Tidy and transform the school headcount data such that it is in a format that will allow you to recreate the plots from the Chapter 9 Walkthrough. This includes determining:
    - the percentage of the student population that each subgroup comprises,
    - the percentage of white students at each school,
    - the percentage of students in poverty at each school, and
    - the percentage of each subgroup that attends a high poverty school (where high poverty is defined as a school where more than 75% of the students are in poverty).

    You can choose to either follow the given code in the textbook or you can use code that you come up with so long as the data ends up in a format suitable for plotting the graphs found in Chapter 9.

2. Create a bar graph of the percentage of the student population that each subgroup comprises (similar to Figure 9.1 of the textbook) for the SC school headcount data. Comment on the racial diversity in SC schools.

3. Create a histogram of the distribution of the percentage of white students at each school (similar to Figure 9.2 of the textbook) for the SC school headcount data. Comment on what this reveals about the distribution of demographics across schools.

4. Create a bar graph of the percentage of each subgroup that attends a high poverty school (similar to Figure 9.3 of the textbook) for the SC school headcount data. Determine the percentage of students across the state who are attending high poverty schools and use this information to comment on whether any subgroups are disproportionately attending high poverty schools.

5. Create a scatterplot of the percentage of students in poverty at a school versus the percentage of white students at that school (similar to Figure 9.4 of the textbook) for the SC school headcount data. Determine the correlation between these two variables and comment on the strength of the relationship.