Reproducible Research

Rich Majerus

You can write your entire paper (text, code, analysis, graphics, etc.) all in R Markdown. As an example, the following maps show the the geographic distribution of Reed College's enrolling students. You can reproduce this example in RStudio with this code and this data. Additionally, You can access a dynamic version of this analysis online. Figure 1 shows the raw matriculant data from 2013 mapped by state. The darker a state's shading, the more matriculants from that state.
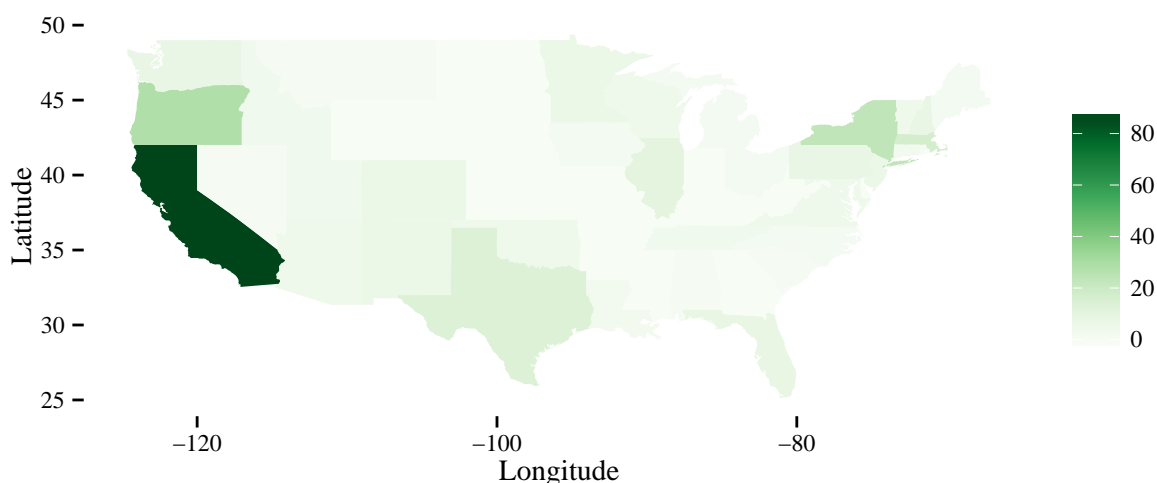

Figure 1: Domestic Geographic Distribution of 2013 Entering Class

However, we may be interested in learning more about the variation in matriculants across all states rather than identifying the states that account for the greatest number of matriculants. One way to approach this task is to map the log of matriculants. Log transforming a variable that contains exceptionally large values (i.e., a right skewed variable) pulls those large values closer to the mean and yields a more symmetrically distributed variable. As for the map, log transforming the number of matriculants increases the variation in the color gradient across states and enables us to better visualize the distribution of Reed's matriculants across the entire country.
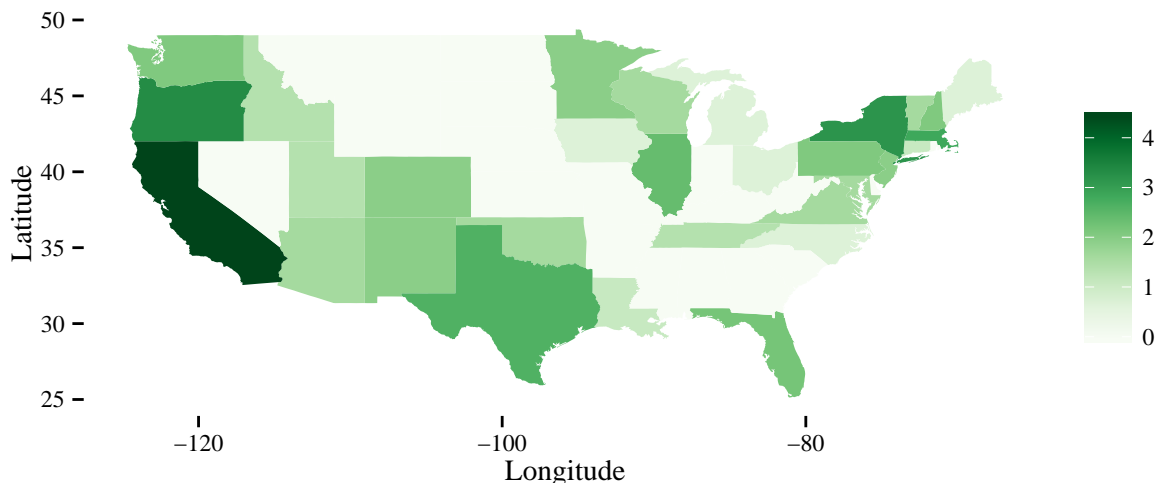

Figure 2: Domestic Geographic Distribution of 2013 Entering Class (Log Transformed)