

hyang390
903320189

Activity 2: Visualize Geospatial Data

Data

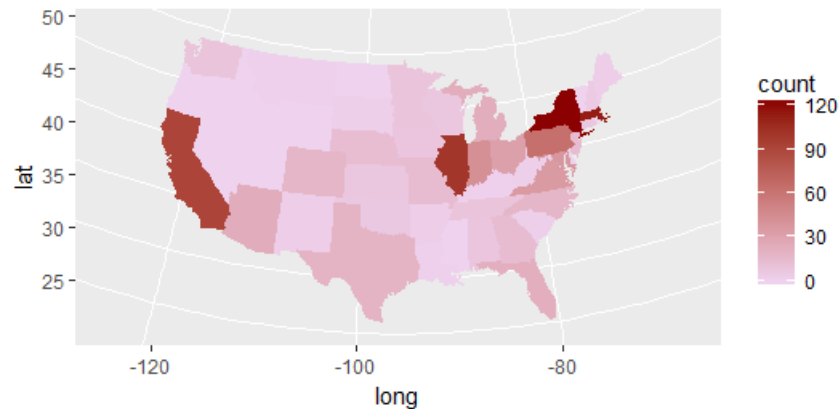
The data for this assignment was retrieved from Kaggle representing information regarding data science universities in the United States, it is uploaded as part of this assignment.

Analysis

For my analysis, I aggregated all the universities according to which state they reside in to get a count of the number of schools in each state; I filled out the data frame with the missing states that do not contain data science universities as well as merged them with the state map data from *ggplot2* to get the longitude and latitude coordinates for each state.

```
aggregated_schools = aggregate(cbind(count = SCHOOL) ~ STATE, data = raw_data, FUN = NROW)
states_data = map_data("state")
missing_states = data.frame(setdiff(state.name, aggregated_schools$STATE), 0)
names(missing_states) = c("STATE", "count")
aggregated_data = rbind(aggregated_schools, missing_states)
```

Visualization



The visualization was plotted in the above plot, with the gradient representing the number of data science universities in each U.S. state.