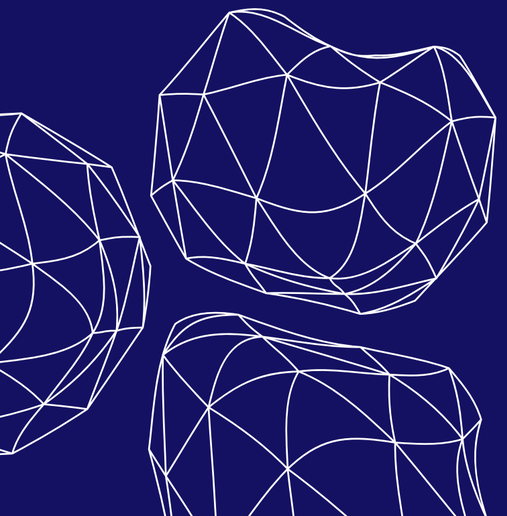
Analysis of Inequality

among Educational institutions in the US South



We determined the extent to which parental income affects 4-year graduation rate across ten states in the Southeast U.S. We operationalize parental income as median parental income and 4-year graduation rate as the percentage of students who complete college or university within four years.

Full User Guide

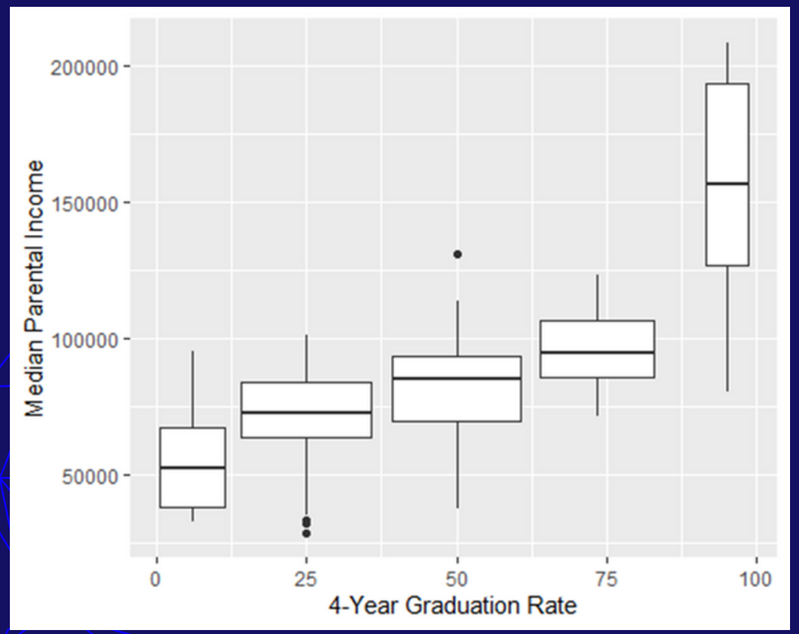
Tiding and manipulating the Data

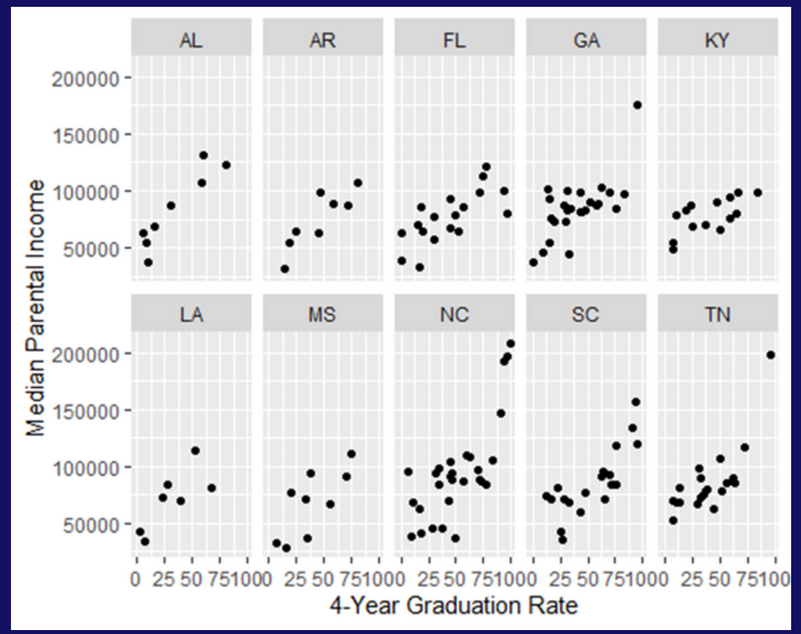
```r
race<- read.csv("HEsegDataviz_CollegeData_4-year_v5.csv")
mrc<- read.csv("mrc_table1.csv")
grad<- read.csv("cc_institution_details.csv") %>%
subset(select = c("chronname","grad_100_percentile"))
state <- read.csv("state_abbrev_crosswalk.csv")%>%
  rename(state = Code)
grad <- grad %>%
rename(name = chronname)
race <- race %>%
rename(name = inst_name)
race_mrc <- mrc %>%
  inner_join(race, by = "name")
college <- race_mrc %>%
  inner_join(grad, by = "name")%>%
  filter(year == 2013)%>%
  filter(state == "AL"| state == "FL"| state == "GA"| state == "AR"| state == "KY"| state == "LA"| state == "MS"| state == "NC"| state == "SC"| state == "TN")
college$grad_100_percentile <- as.numeric(college$grad_100_percentile)
college$grad_100_percentile[is.na(college$grad_100_percentile)]<-mean(college$grad_100_percentile,na.rm=TRUE)
college <- college%>%
  left_join(state, by = "state")
college <- college %>%
  rename(NAME = State)
```

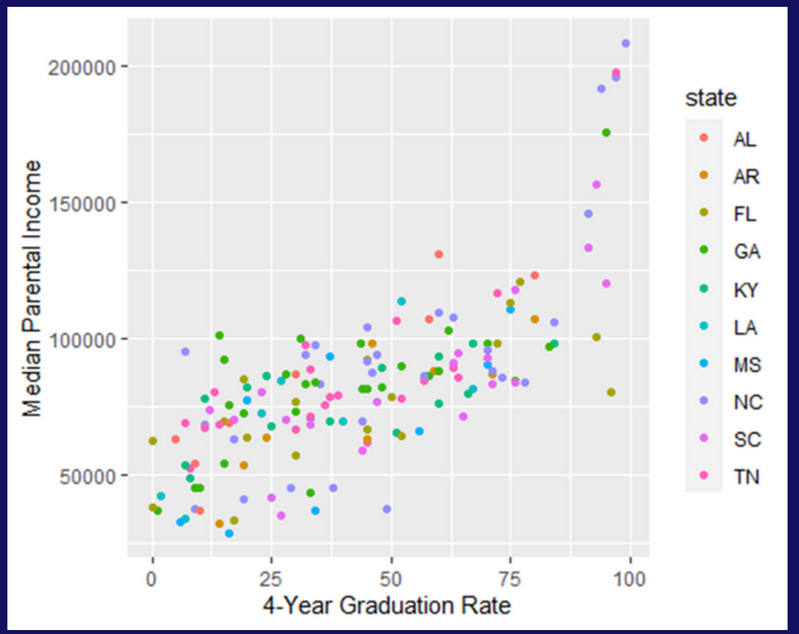
Reading four datasets. Then, renaming variables to facilitate joining the datasets. After joining them, we filter the data to the states that we are interested in. Finally, we clean the data and prepare it for analysis.

Data Visualization

```r
#Box Plot
college %>%
  ggplot(mapping = aes(x = grad_100_percentile, y = par_median)) +
geom_boxplot(mapping = aes(group = cut_width(grad_100_percentile, 25))) +
  labs(x = "4-Year Graduation Rate", y = "Median Parental Income")
#Facet Wrap
college %>%
  ggplot(mapping = aes(x = grad_100_percentile , y = par_median )) + geom_point() +
  facet_wrap(~state, nrow = 2) +
  labs(x = "4-Year Graduation Rate", y = "Median Parental Income")
#Scatter Plot
college %>%
  ggplot(mapping = aes(x = grad_100_percentile , y = par_median , color = state)) + geom_point() +
 labs(x = "4-Year Graduation Rate", y = "Median Parental Income")
```



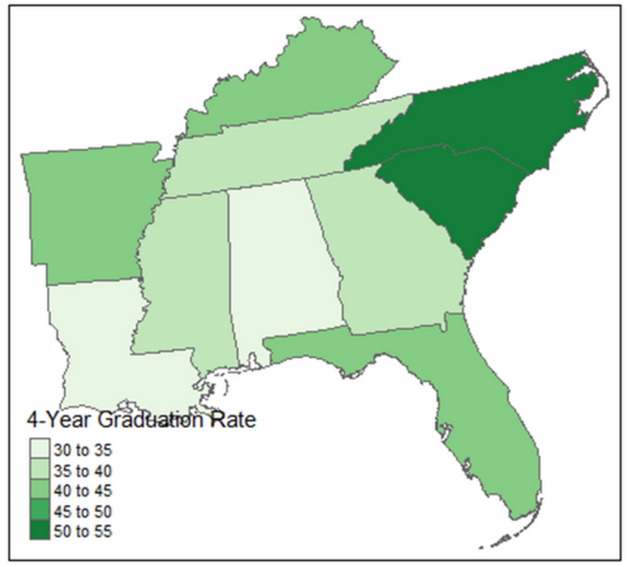


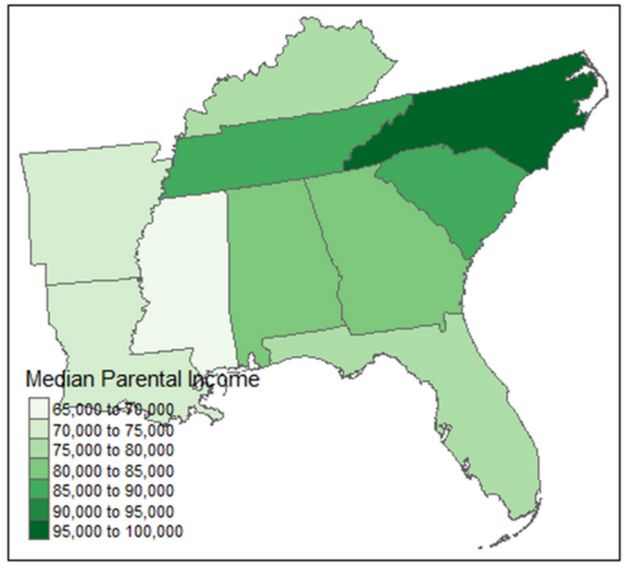


Tiding and manipulating the Data

```r
library(tidycensus)
library(tmap)
library(sf)
library(rgdal)
library(cartogram)
library(igraph)
library(ggraph)
```

```r
us <- get_acs(geography = "state",
                              variables = "B02001_001",
                            year=2013,
                            geometry = TRUE,
                            shift_geo = TRUE)
college_map <- college %>%
  inner_join(us, by = "NAME")
college_map_grad <- college_map %>%
  group_by(NAME)%>%
  summarise(NAME, geometry, `4-Year Graduation Rate` =  mean(grad_100_percentile))
st_sf(college_map_grad) %>%
  tm_shape() +
  tm_polygons("4-Year Graduation Rate", palette = "Greens")
college_map_income <- college_map%>%
  group_by(NAME)%>%
  summarise(NAME, geometry, `Median Parental Income` =  mean(par_median))
st_sf(college_map_income) %>%
  tm_shape() +
  tm_polygons("Median Parental Income", palette = "Greens")
```





All of that is INTERACTIVE!

Our R Shiny app features five tabs, three with interactive plots and the other two with interactive maps.