

Funding and Spending of U.S. Schools

Shankar Prabhu

July 2020

In this section, we will take a look at how the enrollment demographics of U.S schools have changed in the past 30 years.

```
data = read.csv(
  "C:/Users/shank/Documents/Data_Science_Projects/education/states_all_extended.csv",
  header = TRUE, sep = ",")
data[is.na(data)] = 0
data$AM <- rowSums(data[grepl('.._AM_|..._AM_|$', names(data))])
data$AS <- rowSums(data[grepl('.._AS_|..._AS_|$', names(data))])
data$HI <- rowSums(data[grepl('.._HI_|..._HI_|$', names(data))])
data$BL <- rowSums(data[grepl('.._BL_|..._BL_|$', names(data))])
data$WH <- rowSums(data[grepl('.._WH_|..._WH_|$', names(data))])
data$HP <- rowSums(data[grepl('.._HP_|..._HP_|$', names(data))])
data$TR <- rowSums(data[grepl('.._TR_|..._TR_|$', names(data))])

race_by_year = setNames(aggregate(cbind(data$AM, data$AS,
                                         data$HI, data$BL,
                                         data$WH, data$HP,
                                         data$TR)
                                , by=list(YEAR=data$YEAR),
                                FUN = sum, na.rm = TRUE), c("YEAR", "American_Indian",
                                                            "Asian", "Hispanic",
                                                            "African_American",
                                                            "White", "Hawaiian_Native",
                                                            "Two_or_More"))
```

We have a lot of missing data for years before 2009 on enrollment demographics. It would not be wise to assume a mean/median value for years prior to 2009 since we are really trying to analyze trends of enrollment over time. The most practical method would be to replace na values with 0 and ignore those years in the graphs that we produce. This will lead to less meaningful results, as I think it would have been far more interesting to see how demographics changed in the late 1900's, but giving this dataset it does not look like that is possible.

Still, we can see how the race makeup of U.S. schools has shifted in the past decade.

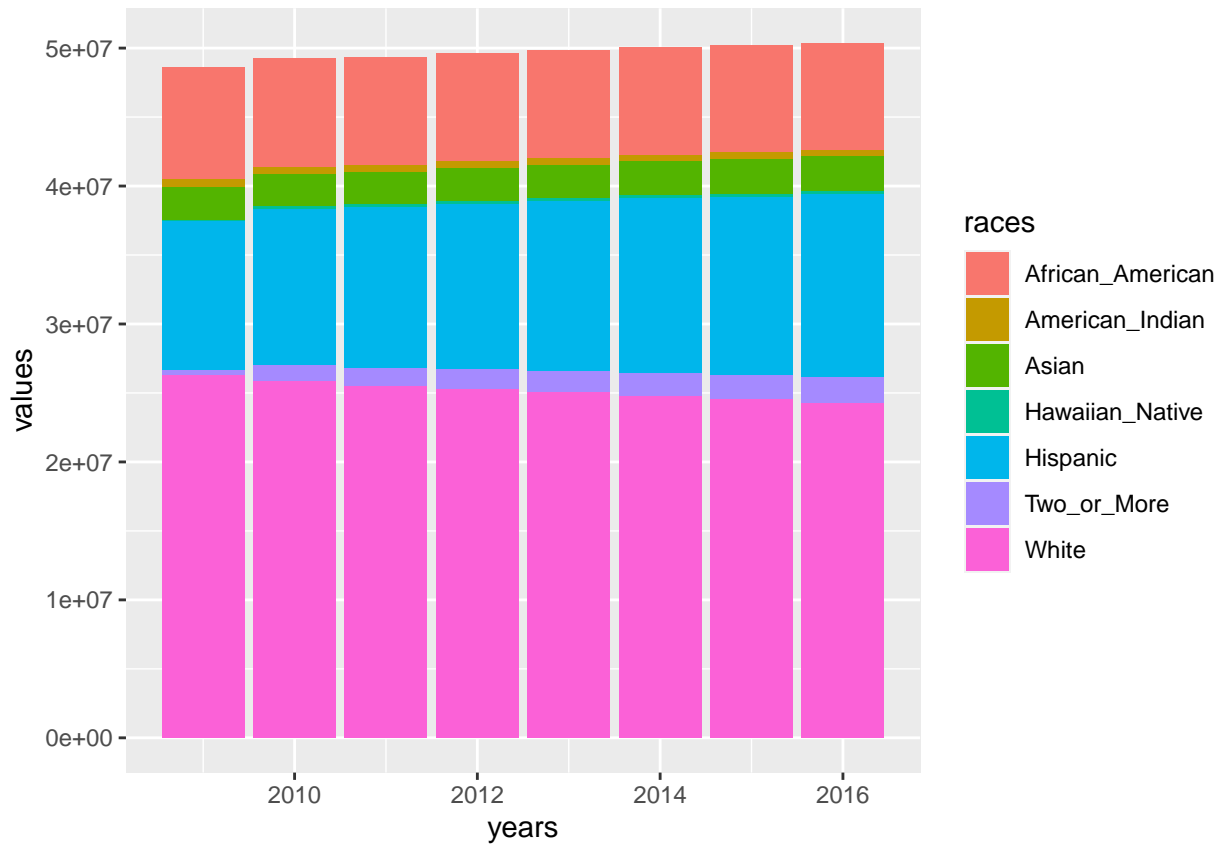
```
filtered_race = race_by_year[c(24:31),]
years = rep(filtered_race$YEAR, 7)
races = c(rep("American_Indian", 8),
          rep("Asian", 8), rep("Hispanic", 8),
          rep("African_American", 8), rep("White", 8),
          rep("Hawaiian_Native", 8), rep("Two_or_More", 8))
values = c(filtered_race$American_Indian, filtered_race$Asian,
```

```

    filtered_race$Hispanic, filtered_race$African_American,
    filtered_race$White, filtered_race$Hawaiian_Native,
    filtered_race$Two_or_More)
race_data = data.frame(years, races, values)

library(ggplot2)
ggplot(race_data, aes(fill=races, y=values, x=years)) +
  geom_bar(position="stack", stat="identity")

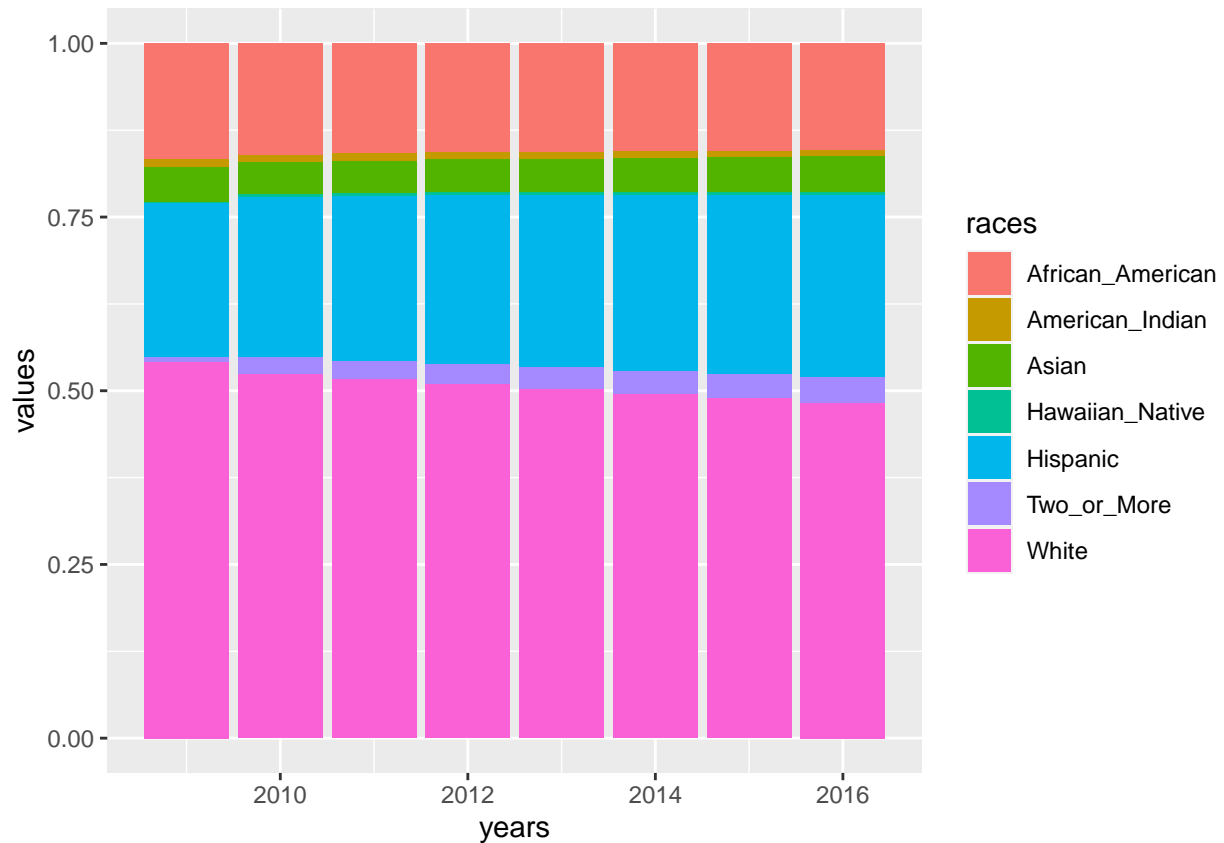
```



```

ggplot(race_data, aes(fill=races, y=values, x=years)) +
  geom_bar(position="fill", stat="identity")

```



Over time, it looks as though White enrollment is decreasing, with its percentage share shifting primarily toward those who identify as Hispanic or multiracial.

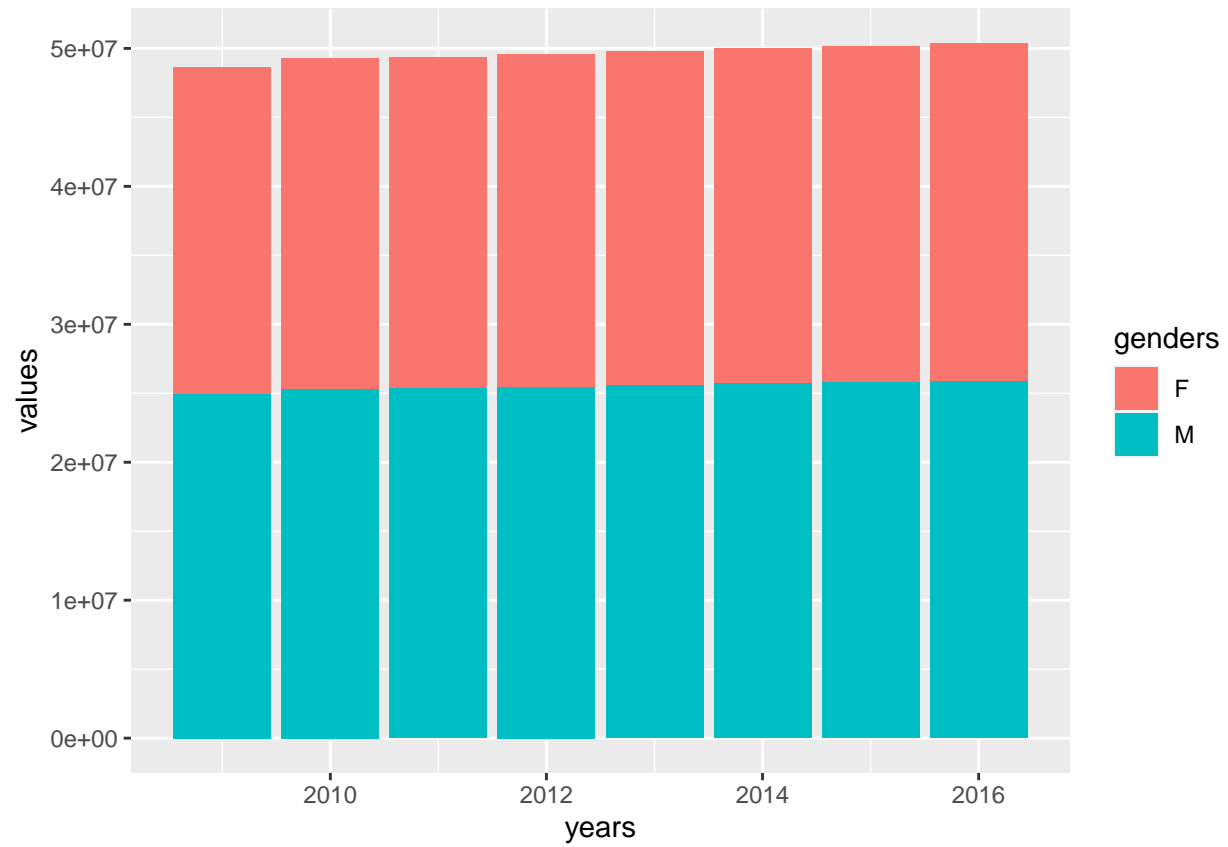
Next, we can take a look at the gender makeup of U.S. schools.

```
data = read.csv(
  "C:/Users/shank/Documents/Data_Science_Projects/education/states_all_extended.csv",
  header = TRUE, sep = ",")
data[is.na(data)] = 0
data$F <- rowSums(data[grepl('..._F$|..._F$', names(data))])
data$M <- rowSums(data[grepl('..._M$|..._M$', names(data))])

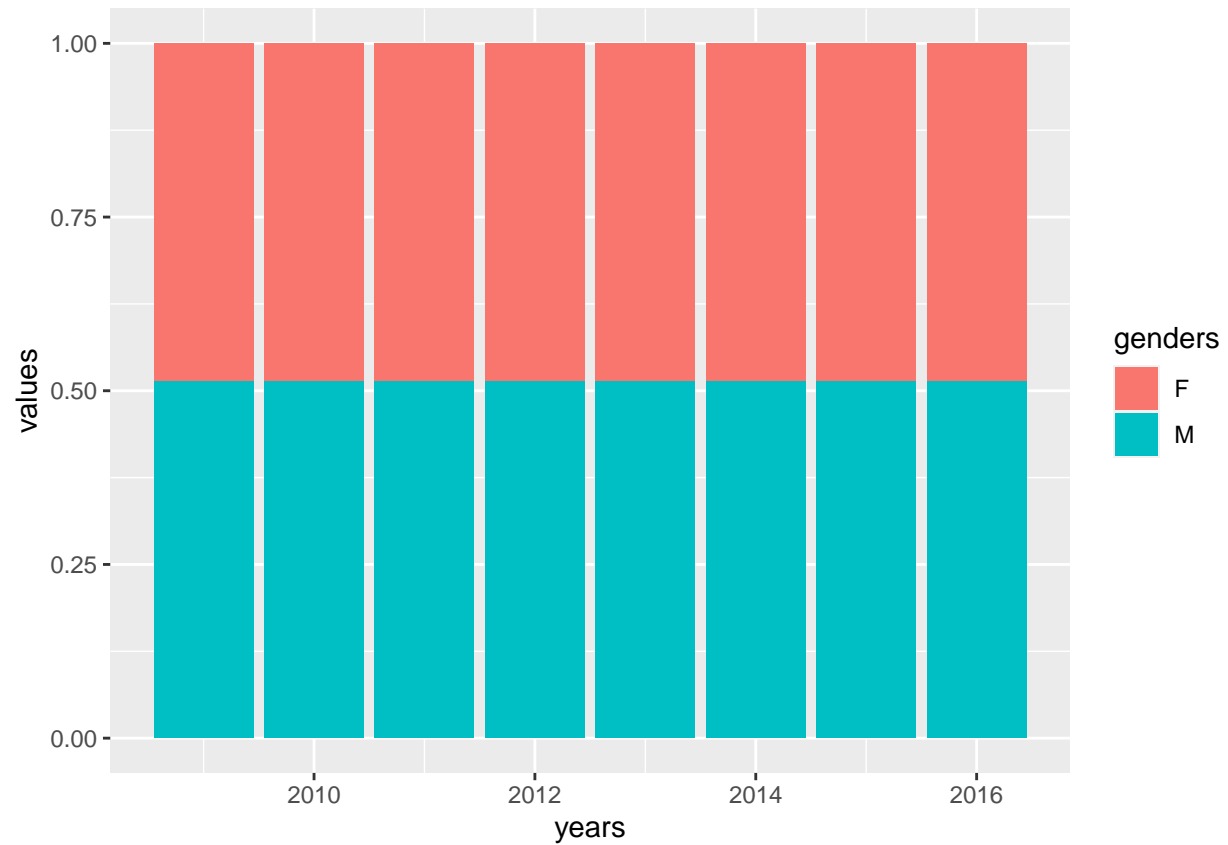
gender_by_year = setNames(aggregate(cbind(data$F, data$M)
  , by=list(YEAR=data$YEAR),
  FUN = sum, na.rm = TRUE), c("YEAR", "F", "M"))

filtered_gender = gender_by_year[c(24:31),]
years = rep(filtered_gender$YEAR, 2)
genders = c(rep("F", 8), rep("M", 8))
values = c(filtered_gender$F, filtered_gender$M)
gender_data = data.frame(years, genders, values)

library(ggplot2)
ggplot(gender_data, aes(fill=genders, y=values, x=years)) +
  geom_bar(position="stack", stat="identity")
```



```
ggplot(gender_data, aes(fill=genders, y=values, x=years)) +  
  geom_bar(position="fill", stat="identity")
```



We can see the school system still has slightly more enrolled male students, with the percentage distribution of female and male students remaining nearly constant throughout the years we have data for.

Overall, there is not a rich trend observable in the data, primarily because so much data is missing from earlier years. Given this condition, I do not think there is much to be gained in further exploration of specific states or regions.