College Enrollment

Carie Spannagel, Kristen Tompkins, Harrison Bardo

10/22/2020

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2

## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.2

## Warning: package 'readr' was built under R version 4.0.2

## Warning: package 'forcats' was built under R version 4.0.2

## -- Conflicts ---------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(forcats)
library(modelr)
library(ModelMetrics)
```

```
## Warning: package 'ModelMetrics' was built under R version 4.0.2

##
## Attaching package: 'ModelMetrics'

## The following objects are masked from 'package:modelr':
##
##     mae, mse, rmse

## The following object is masked from 'package:base':
##
##     kappa
```

```r
library(Hmisc)
```

```
## Warning: package 'Hmisc' was built under R version 4.0.2

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
library(readr)
library(dplyr)
library(tictoc)
```

```
## Warning: package 'tictoc' was built under R version 4.0.2
```

```r
library(RColorBrewer)
```

```
## Warning: package 'RColorBrewer' was built under R version 4.0.2
```

```r
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.2
```

```r
library(colorspace)
```

```
## Warning: package 'colorspace' was built under R version 4.0.2
```

```r
library(scales)
```

```
## Warning: package 'scales' was built under R version 4.0.2

##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
library(ggplot2)
#tinytex::install_tinytex()
```

Background

In the initial phases of our project we expressed the idea of working with educational longitudinal data sets with Dr. Will Doyle. Dr. Doyle shared that since BPS is not publicly available, it would not be an option for our project. However, he directed us to the ELS 2002-2012 which did have data on students who are enrolled in college, and thought it would be a helpful resource. He referred us to:https://nces.ed.gov/OnlineCodebook and directed us to the video for how to access the data:https://www.youtube.com/watch?v=BsDP00PUEZI

While this might not seem a significant step, the important learning was reaching out to colleagues who have experience with large data sets and determining what was realistic for our interests, scope and timeline. This conversation helped us narrow our focus to college preparatory support systems which we will discuss in our introduction section.

Introduction:

Since local and state governments define college readiness standards, students may not be able to determine if their school adequately prepares them for college. By exploring college readiness indicators in the High School Longitudinal Study from the National Center for Education Statistics, our group will explore the correlation between access to college preparation supports and enrollment in post-secondary programs based on student demographic information and school location.

For background on why such a study is important, a report released by the e National Post secondary Education Cooperative Working Group on Access to Post secondary Education on K-12 Underrepresented youth reveals that "since the mid-70s, college-going rates for white, non-Hispanic students have increased significantly. In contrast, despite progress in the 1990s, students from historically underrepresented minority groups have not experienced substantial increases in college-going rates. Improving their college-going rates is an issue of growing urgency for colleges, universities, and states" (Paving the Way to Post Secondary Research, vii). Their report evaluated early intervention programs in K-12 and found that programs that double the rate of college-going participants included:

Providing a key person who monitors and guides the student over a long period of time—a "mentor," program director, faculty member, or guidance counselor. Studies are not clear on which of these is most effective.

Providing high-quality instruction through access to the most challenging courses offered by the school ("untracking"), through special coursework that supports and augments the regular curricular offerings (tutoring and specially designed classes), or by revamping the curriculum to better address the learning needs of the students.

Making long-term investments in students rather than short-term interventions. The longer students were in the program, the more likely they were reported to benefit from it.

Paying attention to the cultural background of students. Many programs reported having greater success with one group of students than another; it is likely that background and expertise of the staff and directors helped them to make cultural connections with students.

Providing a peer group that supports students' academic aspirations as well as giving them social and emotional support.

Providing financial assistance and incentives. Financial assistance is important for access to academic leveling experiences—college visits and SAT preparation courses— as well as to monetary support to make college a realistic possibility for some students. Scholarships make the difference between going to college or not for many low-income students (Thomas, 1998; St. John, 1990).

Despite the limitations of the report including attrition, lack of evaluation data on if the programs had an impact, and participation of males, the report was conclusive. The notion that the chances are equal for college readiness to students across all SES status, race, language, is a myth. The report concluded that "since 1990 college enrollment rates have been improving for underrepresented groups, however, given the rapid shifts in demographics in the United States, increasing the enrollment rates for traditionally underrepresented students is a matter of growing urgency" (Paving the Way to Post secondary Education, pg. 2)

Using the Paving the Way to Post secondary Education as a framework, our group will evaluate current college readiness programs and their impact. This data science report will dive into the High School Longitudinal (HSLS) study that followed 23,000 plus 9th-grade students in 2009 from 944 schools. The HSLS study followed up with these students in 2012 and 2016 following throughout their secondary and post secondary years. More on our data set can be found below in the data set summary. We are excited to share the outcomes that have the potential to guide college prep programs in high schools.

In order to load our file without crashing R Studio, we cleaned our data down from about 9600 variables and 23,503 observations to 242 variables. These variables were selected because they had a high probability of aligning with our research framework and provided data. Some of our original data was not available due to privacy.

Problem Statement

How does college enrollment vary based on student demographics (SES, race, language) and school location?

```r
precollege <- read_csv("~/Desktop/HS Project/Student.csv")
```

```
## Warning: Duplicated column names deduplicated: 'STU_ID' => 'STU_ID_1' [161]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```r
percentile <- read_csv("~/Desktop/HS Project/percentileCSV.csv")
```

```
## Parsed with column specification:
## cols(
##   Percentile = col_double(),
##   Ofe = col_double()
## )
```

In preparation for creating a model, we looked at our available data and narrowed down our independent variables based on our framework. This provided the opportunity to explore proportion of each potential independent variable.

```
college<-precollege%>%
        select (STU_ID,S3CLGFT, S3CLGFT, S3CLGFT,C1POSTSECREQ, C1CLGFAIR, C1VISITCLG, C1CLGPREP, C1INFO

#College Prep
prop.table(table(college$C1CLGPREP))
```

```
##
##          -9          -8           0           1
## 0.005190827 0.096540867 0.342807301 0.555461005
```

```
prop.table(table(college$C1VISITCLG))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.302557120 0.597072714
```

```
prop.table(table(college$C1CLGFAIR))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.068459346 0.831170489
```

```
prop.table(table(college$C1POSTSECREQ))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.034421138 0.865208697
```

```
prop.table(table(college$S3CNSLCLG))
```

```
##
##         -9         -8         -7         -4          1          2          3
## 0.00246777 0.21039867 0.06390674 0.04169680 0.49074586 0.13921627 0.05156788
```

```
#Family & Student Support
prop.table(table(college$C1FINANCEAID))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.046972727 0.852657108
```

```
prop.table(table(college$C1ASSISTOTH))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.559971068 0.339658767
```

```r
prop.table(table(college$C1INFOSESSN))
```

```
##
##          -9          -8           0           1
## 0.003829298 0.096540867 0.042462664 0.857167170
```

```r
prop.table(table(college$S3CNSLAID))
```

```
##
##          -9          -8          -7          -4           1           2
## 0.002382675 0.210398673 0.063906735 0.041696805 0.329234566 0.310300813
##           3
## 0.042079735
```

```r
#College Prep Courses
prop.table(table(college$S3ANYCLGCRED))
```

```
##
##          -9          -8          -4           0           1
## 0.007148024 0.210398673 0.045823937 0.352720929 0.383908437
```

```r
prop.table(table(college$S3AP))
```

```
##
##          -9          -8          -7          -4           1           2
## 0.008339361 0.210398673 0.352720929 0.045823937 0.294260307 0.063055780
##           3
## 0.025401013
```

```r
prop.table(table(college$S3IB))
```

```
##
##          -9          -8          -7          -4           1           2
## 0.008764839 0.210398673 0.352720929 0.045823937 0.017529677 0.336169851
##           3
## 0.028592095
```

```r
prop.table(table(college$S3DUAL))
```

```
##
##          -9          -8          -7          -4           1           2
## 0.008637195 0.210398673 0.352720929 0.045823937 0.139811939 0.213377016
##           3
## 0.029230311
```

```r
#School Location
prop.table(table(college$X1LOCALE))
```

```
##
##         1         2         3         4
## 0.2846020 0.3602519 0.1186232 0.2365230
```

```
prop.table(table(college$X1REGION))
```

```
##
##         1         2         3         4
## 0.1558099 0.2648173 0.4079054 0.1714675
```

```
#Attending college full-time or part-time as of Nov 1, 2013
prop.table(table(college$S3CLGFT))
```

```
##
##          -9          -8          -7           1           2           3
## 0.004084585 0.210398673 0.213674850 0.495341020 0.045483555 0.031017317
```

Data Cleanup

For all of the variables that we intended to include in our index measures, we re-coded those variables so that they were binary, converting any missing data into "0". We contemplated entirely eliminating observations with missing data, however, this would have greatly reduced the number of included observations, and we ultimately determined that they would not hurt the index measures, since these measures are only affected by the presence of the condition, and not the absence of it.

We did eliminate the students who did not have Socioeconomic Status information as we felt this was an important variable to control for, and we would not have been able to do so without dropping the missing observations.

We also re-coded Locale and Region variables so that they were more easily interpreted. Finally, we condensed our data frame to only include the variables of highest interest.

```
#Prep Courses
  college<-college%>%
            mutate(S3ANYCLGCRED_new=recode(S3ANYCLGCRED,`-4` =0, `-8` =0, `-9`=0))
  college<-college%>%
            mutate (S3AP_new=recode(S3AP, `2`=0, `3` =0, `-4` =0, `-7`=0, `-8` =0, `-9`=0))
  college<-college%>%
            mutate(S3IB_new=recode(S3IB,`2`=0, `3` =0, `-4` =0, `-7`=0, `-8` =0, `-9`=0))
  college<-college%>%
            mutate(S3DUAL_new=recode(S3DUAL,`2`=0, `3` =0, `-4` =0, `-7`=0, `-8` =0, `-9`=0))


#Family & Student Support
   college<-college%>%
            mutate(C1FINANCEAID_new=recode(C1FINANCEAID,`-8` =0, `-9`=0))
   college<-college%>%
            mutate(C1ASSISTOTH_new=recode(C1ASSISTOTH,`-8` =0, `-9`=0))
   college<-college%>%
            mutate(C1INFOSESSN_new=recode(C1INFOSESSN,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(S3CNSLAID_new=recode(S3CNSLAID,`2`=0, `3` =0, `-4` =0, `-7`=0, `-8` =0, `-9`=0))
```

```r
#College Prep
    college<-college%>%
            mutate(C1CLGPREP_new=recode(C1CLGPREP,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(C1VISITCLG_new=recode(C1VISITCLG,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(C1CLGFAIR_new=recode(C1CLGFAIR,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(C1POSTSECREQ_new=recode(C1POSTSECREQ,`-8` =0, `-9`=0))
     college<-college%>%
            mutate(S3CNSLCLGL_new=recode(S3CNSLCLG,`2`=0, `3` =0, `-4` =0, `-7`=0, `-8` =0,`-9`=0))

#Attending College
    college<-college%>%
            mutate(S3CLGFT_new_full=recode(S3CLGFT, `2` = 0, `3` = 0, `-7` = 0,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(S3CLGFT_new_part=recode(S3CLGFT, `1` = 0, `3` = 0, `-7` = 0,`-8` =0, `-9`=0))
    college<-college%>%
            mutate(S3CLGFT_new_full_part=recode(S3CLGFT, `2` = 1, `3` = 0, `-7` = 0,`-8` =0, `-9`=0))
#Demographics
    college<-college%>%
      mutate(X2RACE_new=recode(X2RACE, `1`=9, `8`=1))
    college<-college%>%
      mutate(X2SES_new=recode(X2SES, `-8`= NA_real_))
    college<-college%>%
      mutate(X1FAMINCOME_new=recode(X1FAMINCOME, `1`= "Less than $15,000", `2`="$15,000 to $35,000",
                                    `3`="$35,000 to $55,000", `4`="$55,000 to $75,000",
                                    `5`="$75,000 to $95,000",
                                    `6`="$95,000 to $115,000", `7`="$115,000 to $135,000",
                                    `8`="$135,000 to $155,000", `9`="$155,000 to $175,000",
                                    `10`="$175,000 to $195,000", `11`= "$195,000 to $215,000",
                                    `12`= "$215,000 to $235,000",
                                    `13`= "More than $235,000"))
```

```
## Warning: Unreplaced values treated as NA as .x is not compatible. Please specify
## replacements exhaustively or supply .default
```

```r
#School Location
    college<-college%>%
      mutate(X1LOCALE_new=recode(X1LOCALE, `1`="City", `2`="Suburb", `3`="Town",
                                 `4`="Rural"))
    college<-college%>%
    mutate(X1REGION_new=recode(X1REGION, `1`= "Northeast", `2`= "Midwest", `3`= "South",
                               `4`="West"))
college %>%
  select(STU_ID,S3CLGFT_new_full, S3CLGFT_new_part, S3CLGFT_new_full_part,C1POSTSECREQ_new, C1CLGFAIR_ne
```

```
## # A tibble: 23,503 x 19
##    STU_ID S3CLGFT_new_full S3CLGFT_new_part S3CLGFT_new_ful~ C1POSTSECREQ_new
##     <dbl>            <dbl>            <dbl>            <dbl>            <dbl>
##  1  10001                1                0                1                1
##  2  10002                1                0                1                1
```

```
##   3  10003               1                0                1                1
##   4  10004               0                0                0                0
##   5  10005               0                0                0                1
##   6  10006               1                0                1                1
##   7  10007               1                0                1                1
##   8  10008               1                0                1                1
##   9  10009               0                0                0                1
## 10  10010               0                0                0                1
## # ... with 23,493 more rows, and 14 more variables: C1CLGFAIR_new <dbl>,
## #   C1VISITCLG_new <dbl>, C1CLGPREP_new <dbl>, C1INFOSESSN_new <dbl>,
## #   C1ASSISTOTH_new <dbl>, C1FINANCEAID_new <dbl>, S3CNSLAID_new <dbl>,
## #   S3DUAL_new <dbl>, S3IB_new <dbl>, S3AP_new <dbl>, S3ANYCLGCRED_new <dbl>,
## #   X2RACE_new <dbl>, X2SEX <dbl>, X2SES_new <dbl>
```

Data Exploration

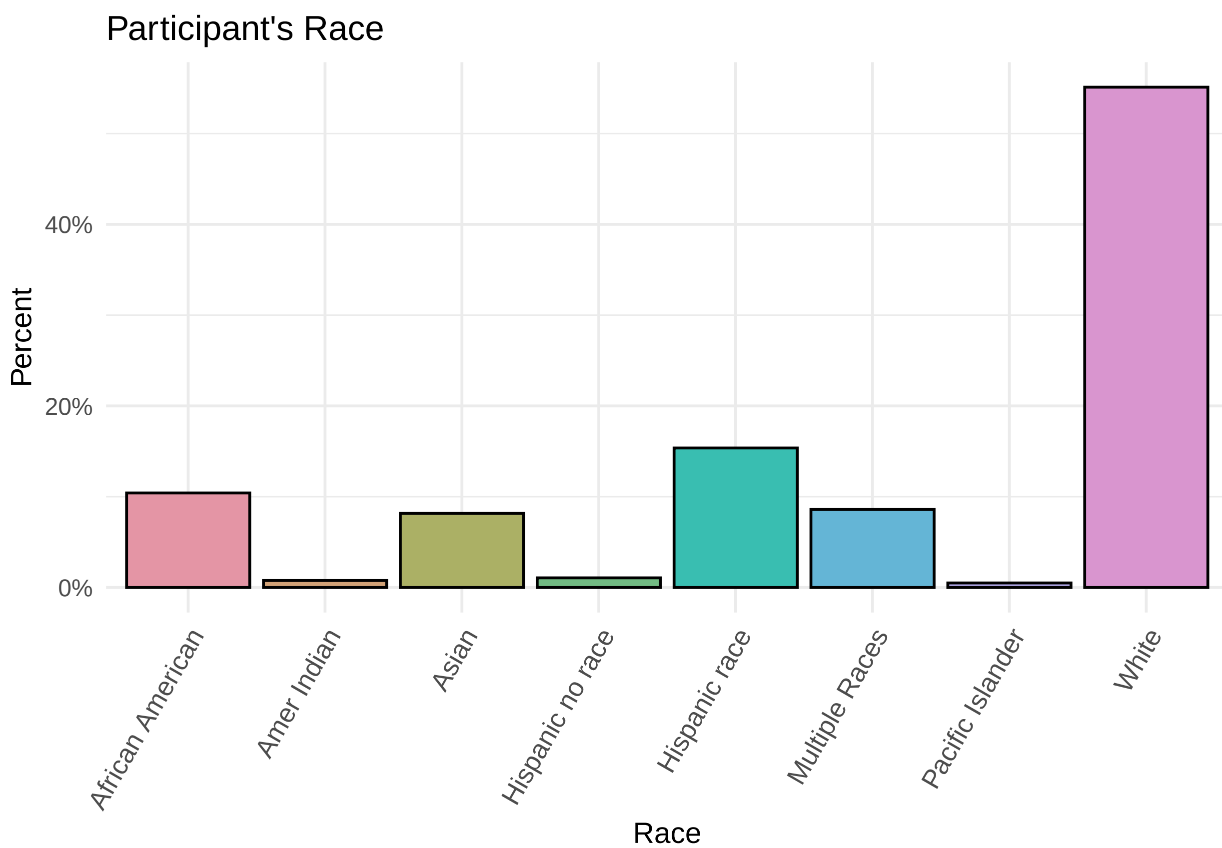
We began our exploration of our data by looking at the demographic information of our sample population. Our exploration included gender, race, and family income. In addition, we began exploring the percentage of college enrollment by these same sample demographics. This helped us narrow down our dependent variable to include students in our sample group that were attending college part or full time in November 1, 2013. This would have been the fall right after their high school graduation.

```
college<-college%>%
      mutate(X2RACE_new2=recode(X2RACE, `1`="Amer Indian", `8`="White", `2`= "Asian",
                                `3`="African American", `4`="Hispanic no race",
                                `5`= "Hispanic race",
                                `6`="Multiple Races", `7`="Pacific Islander"))
table(college$X2RACE_new2)
```

```
##
## African American      Amer Indian            Asian Hispanic no race
##             2448              181             1922              250
##    Hispanic race   Multiple Races Pacific Islander            White
##             3612             2021              118            12951
```

```
ggrace<-ggplot(college,
   aes(x = X2RACE_new2,
       y = ..count.. / sum(..count..))) +
  geom_bar(fill =rainbow_hcl(8), color= "black") +
  theme_minimal() +
  labs(x = "Race",
       y = "Percent",
  title  = "Participant's Race",
  color="Legend") +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggrace
```




```r
college_sumrace<-college%>%
  group_by(X2RACE_new2)%>%
  summarise(enrollment_avg=mean(S3CLGFT_new_full_part))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
college_sumrace
```

```
## # A tibble: 8 x 2
##   X2RACE_new2      enrollment_avg
##   <chr>                     <dbl>
## 1 African American          0.491
## 2 Amer Indian               0.354
## 3 Asian                     0.668
## 4 Hispanic no race          0.328
## 5 Hispanic race             0.460
## 6 Multiple Races            0.503
## 7 Pacific Islander          0.449
## 8 White                     0.567
```

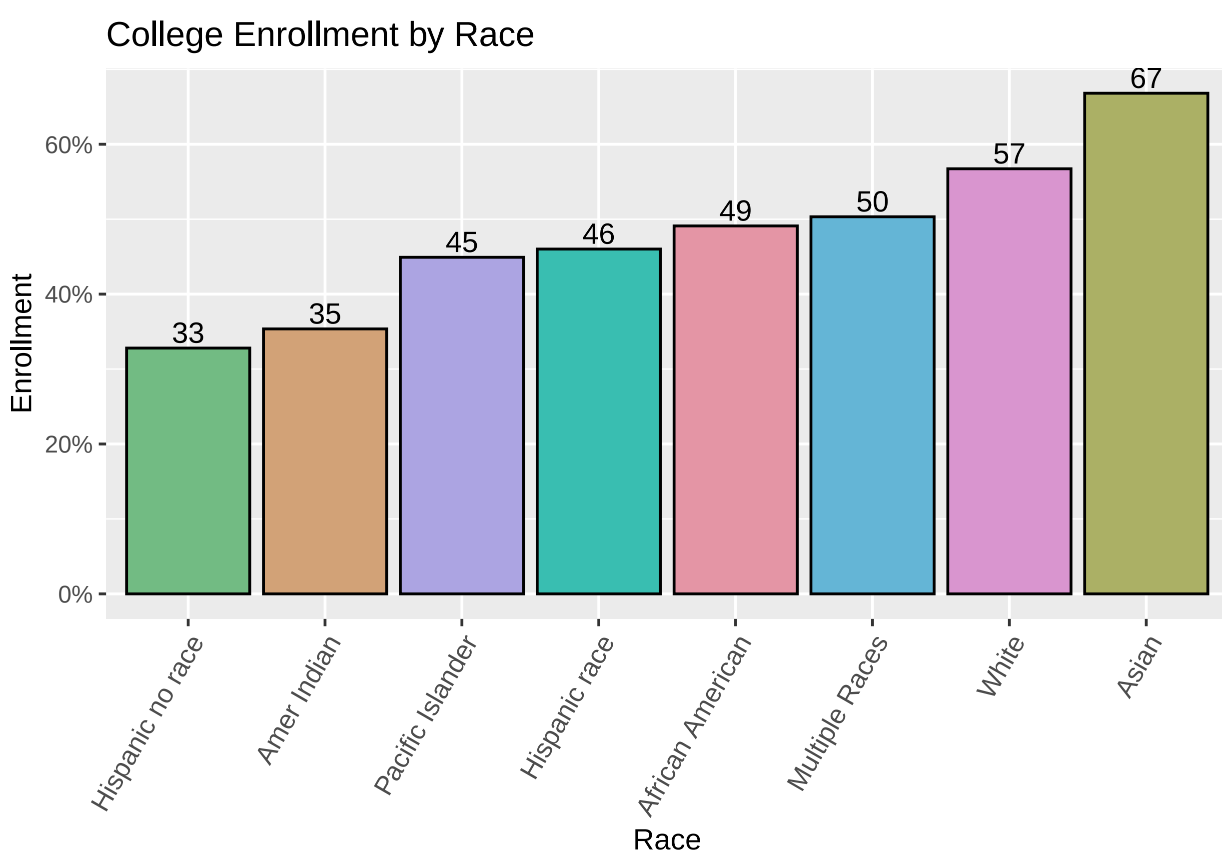
```r
college %>%
  count(X2RACE_new2,S3CLGFT_new_full_part) %>%
  group_by(X2RACE_new2)%>%
  mutate(prop = prop.table(n)) %>%
```

```r
select(-n) %>%
spread(S3CLGFT_new_full_part, prop)%>%kable()
```

X2RACE_new2	0	1
African American	0.5089869	0.4910131
Amer Indian	0.6464088	0.3535912
Asian	0.3319459	0.6680541
Hispanic no race	0.6720000	0.3280000
Hispanic race	0.5398671	0.4601329
Multiple Races	0.4967838	0.5032162
Pacific Islander	0.5508475	0.4491525
White	0.4327079	0.5672921

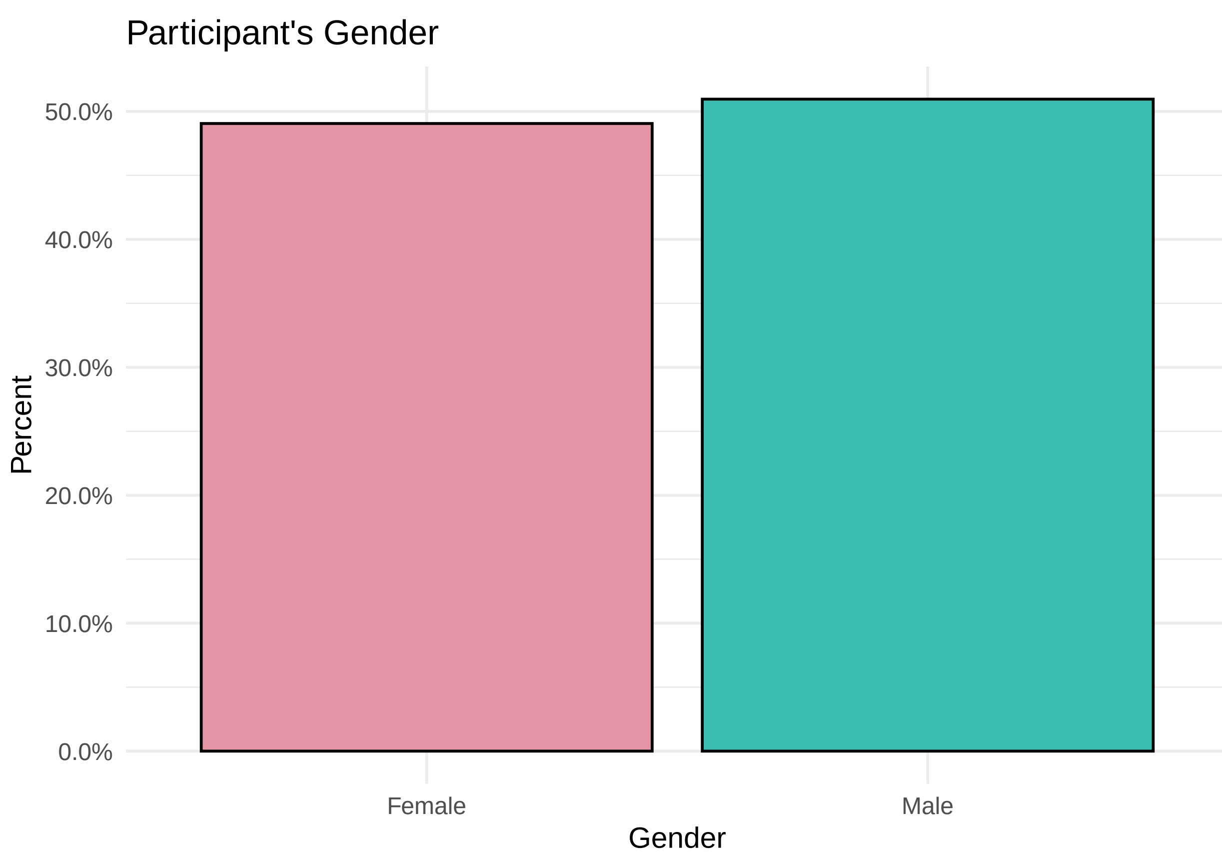
```r
ggsumrace<-ggplot(college_sumrace, aes(x=fct_reorder (X2RACE_new2,enrollment_avg), y=enrollment_avg))+
        geom_bar(stat="Identity", fill= rainbow_hcl(8), color="black")+
   geom_text(aes(label=round(enrollment_avg *100, 0)), position=position_dodge(width=.75), vjust=-0.25)
        labs(x= "Race", y = "Enrollment",
        title = "College Enrollment by Race")+
        scale_y_continuous(labels = scales::percent) +
         theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggsumrace
```



```r
 college<-college%>%
      mutate(X2SEX_new2=recode(X2SEX, `1`="Male", `2`="Female"))


gggender<-ggplot(college,
   aes(x = X2SEX_new2,
       y = ..count.. / sum(..count..))) +
  geom_bar(fill = rainbow_hcl(2), color= "black") +
  theme_minimal() +
  labs(x = "Gender",
       y = "Percent",
  title  = "Participant's Gender") +
  scale_y_continuous(labels = scales::percent)
gggender
```




```r
college_sumgender<-college%>%
  group_by(X2SEX_new2)%>%
  summarise(enrollment_avg=mean(S3CLGFT_new_full_part))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
college_sumgender
```

```
## # A tibble: 2 x 2
```

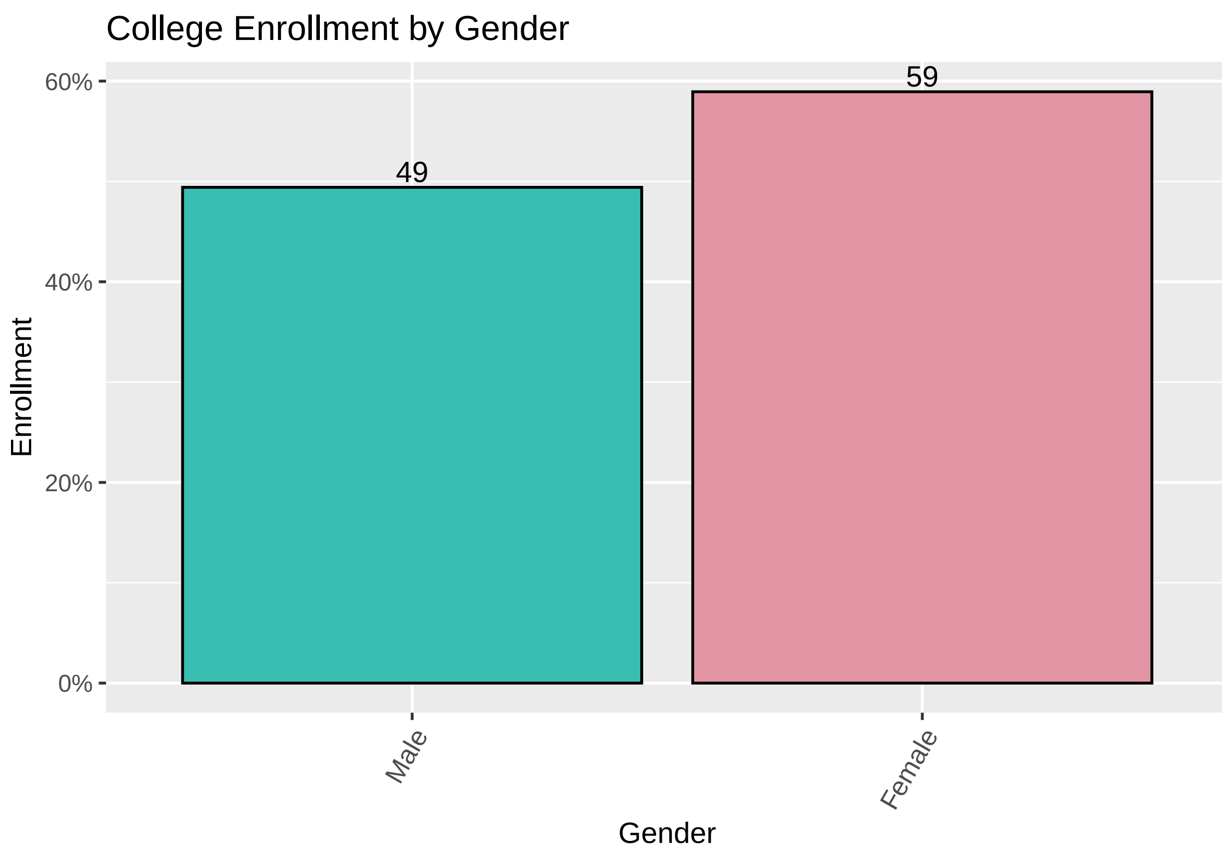
```
##   X2SEX_new2 enrollment_avg
##   <chr>               <dbl>
## 1 Female              0.589
## 2 Male                0.494
```

```r
college %>%
  count(X2SEX_new2,S3CLGFT_new_full_part) %>%
  group_by(X2SEX_new2)%>%
  mutate(prop = prop.table(n)) %>%
  select(-n) %>%
  spread(S3CLGFT_new_full_part, prop)%>%kable()
```

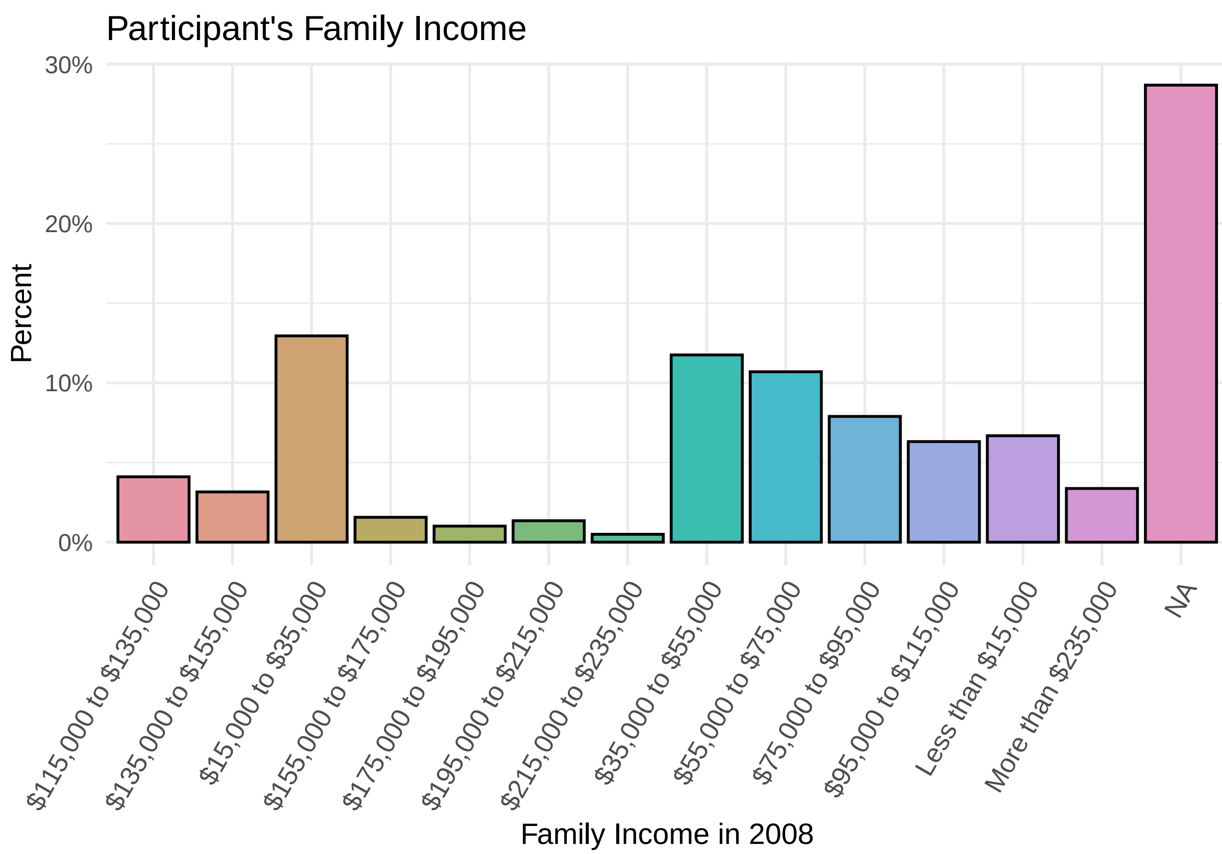
X2SEX_new2	0	1
Female	0.4106523	0.5893477
Male	0.5058873	0.4941127

```r
ggsumgender<-ggplot(college_sumgender, aes(x=fct_reorder (X2SEX_new2,enrollment_avg), y=enrollment_avg)
        geom_bar(stat="Identity", fill= rainbow_hcl(2), color="black")+
  geom_text(aes(label=round(enrollment_avg *100, 0)), position=position_dodge(width=.75), vjust=-0.25)+
        labs(x= "Gender", y = "Enrollment",
        title = "College Enrollment by Gender")+
        scale_y_continuous(labels = scales::percent) +
         theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggsumgender
```




```
ggincome<-ggplot(college,
   aes(x = X1FAMINCOME_new,
       y = ..count.. / sum(..count..))) +
  geom_bar(fill = rainbow_hcl(14), color= "black") +
  theme_minimal() +
  labs(x = "Family Income in 2008",
       y = "Percent",
  title  = "Participant's Family Income") +
  scale_y_continuous(labels = scales::percent)+
   theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))
ggincome
```




```
college_sumincome<-college%>%
  group_by(X1FAMINCOME_new)%>%
  summarise(enrollment_avg=mean(S3CLGFT_new_full_part))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
college_sumincome
```

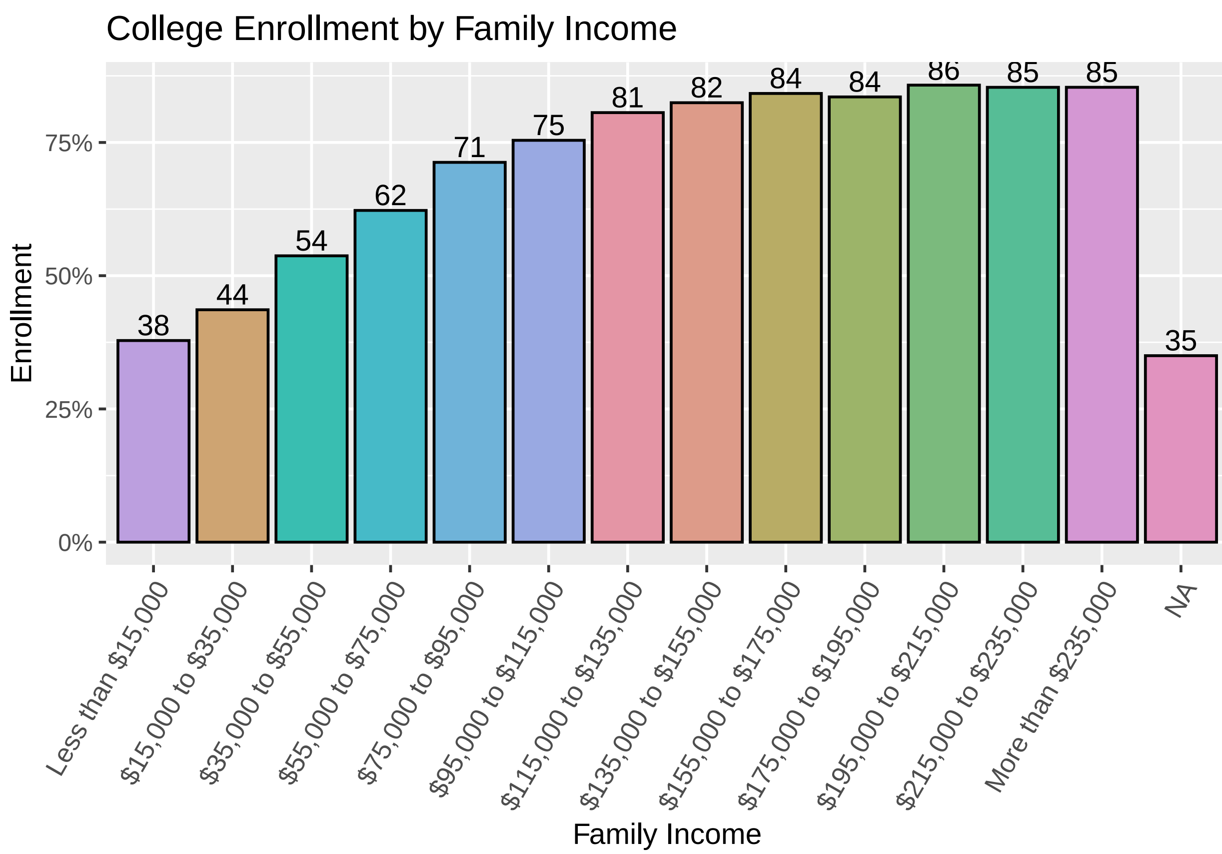
```
## # A tibble: 14 x 2
##    X1FAMINCOME_new      enrollment_avg
##    <chr>                         <dbl>
##  1 $115,000 to $135,000          0.806
##  2 $135,000 to $155,000          0.825
##  3 $15,000 to $35,000            0.436
##  4 $155,000 to $175,000          0.842
##  5 $175,000 to $195,000          0.835
##  6 $195,000 to $215,000          0.858
##  7 $215,000 to $235,000          0.853
##  8 $35,000 to $55,000            0.537
##  9 $55,000 to $75,000            0.623
## 10 $75,000 to $95,000            0.713
## 11 $95,000 to $115,000           0.754
## 12 Less than $15,000             0.378
## 13 More than $235,000            0.854
## 14 <NA>                          0.350
```

```r
college %>%
  count(X1FAMINCOME_new,S3CLGFT_new_full_part) %>%
  group_by(X1FAMINCOME_new)%>%
  mutate(prop = prop.table(n)) %>%
  select(-n) %>%
  spread(S3CLGFT_new_full_part, prop)%>%kable()
```

X1FAMINCOME_new	0	1
$115,000 to $135,000	0.1939834	0.8060166
$135,000 to $155,000	0.1754386	0.8245614
$15,000 to $35,000	0.5639172	0.4360828
$155,000 to $175,000	0.1580381	0.8419619
$175,000 to $195,000	0.1645570	0.8354430
$195,000 to $215,000	0.1424051	0.8575949
$215,000 to $235,000	0.1465517	0.8534483
$35,000 to $55,000	0.4627082	0.5372918
$55,000 to $75,000	0.3774861	0.6225139
$75,000 to $95,000	0.2873315	0.7126685
$95,000 to $115,000	0.2459569	0.7540431
Less than $15,000	0.6216561	0.3783439
More than $235,000	0.1464646	0.8535354
NA	0.6501038	0.3498962

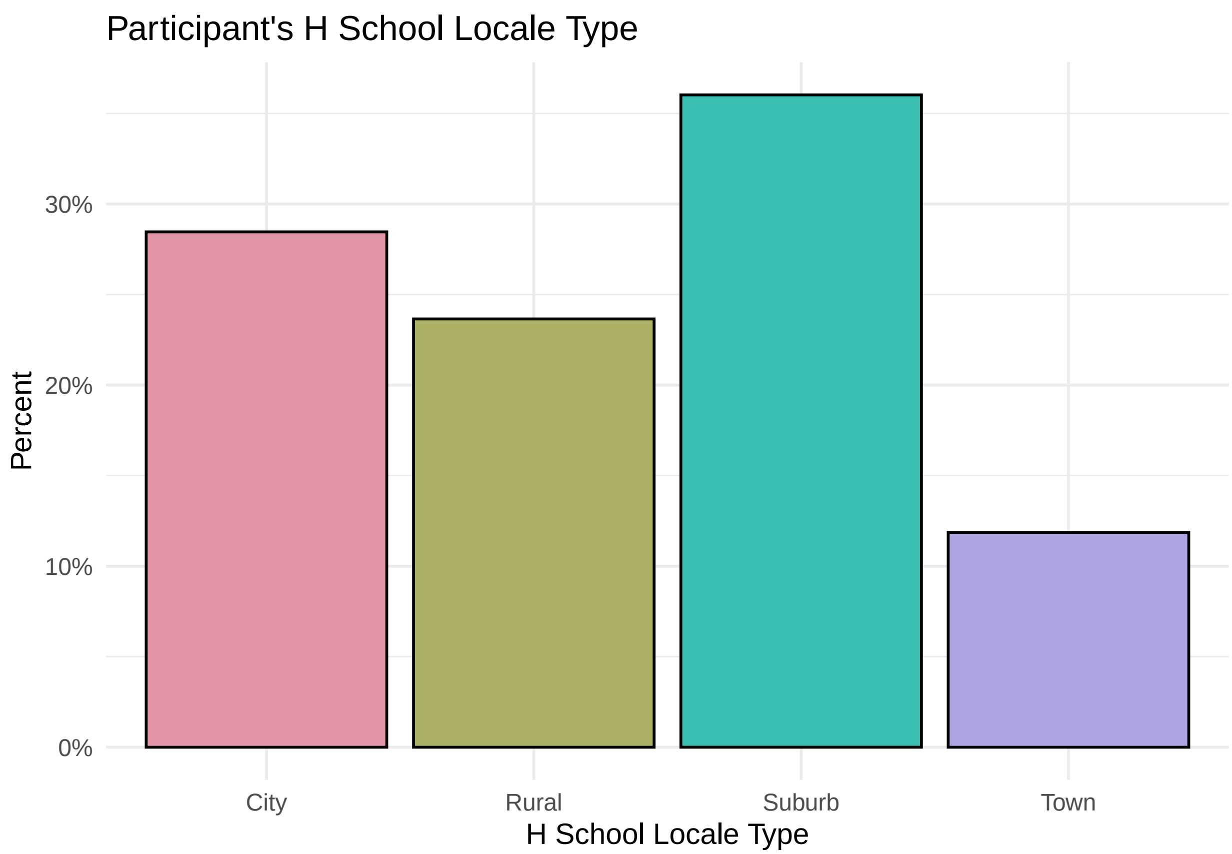
```r
ggsumincome<-college_sumincome%>%
            mutate(X1FAMINCOME_new = fct_relevel(X1FAMINCOME_new, "Less than $15,000", "$15,000 to $35,
        "$35,000 to $55,000", "$55,000 to $75,000",  "$75,000 to $95,000","$95,000 to $115,000", "$115,
  ggplot (aes(x=X1FAMINCOME_new, y=enrollment_avg))+
        geom_bar(stat="Identity", fill= rainbow_hcl(14), color="black")+
  geom_text(aes(label=round(enrollment_avg *100, 0)), position=position_dodge(width=.75), vjust=-0.25)+
        labs(x= "Family Income", y = "Enrollment",
        title = "College Enrollment by Family Income")+
        scale_y_continuous(labels = scales::percent) +
         theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggsumincome
```



Through our data exploration, we were interested in where our sample population's schools were located and the college enrollment based on these factors. Although the states and school ids for our sample population was unavailable, we were able to explore the school location based on four regions of the country and the school location type.

```
gglocale<-ggplot(college,
   aes(x = X1LOCALE_new,
       y = ..count.. / sum(..count..))) +
  geom_bar(fill = rainbow_hcl(4), color= "black") +
  theme_minimal() +
  labs(x = "H School Locale Type",
       y = "Percent",
  title  = "Participant's H School Locale Type") +
  scale_y_continuous(labels = scales::percent)
gglocale
```




```r
college_sumlocale<-college%>%
  group_by(X1LOCALE_new)%>%
  summarise(enrollment_avg=mean(S3CLGFT_new_full_part))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
college_sumlocale
```

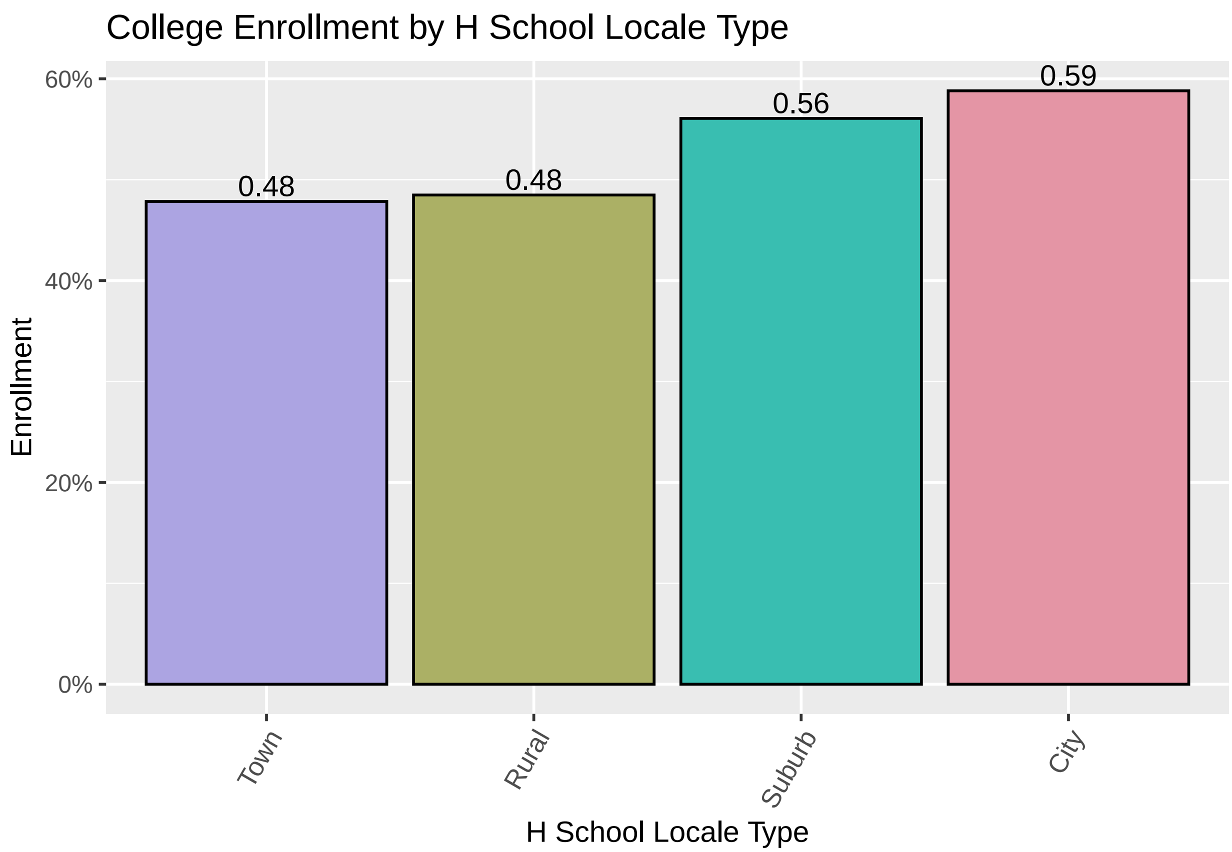
```
## # A tibble: 4 x 2
##   X1LOCALE_new enrollment_avg
##   <chr>                 <dbl>
## 1 City                  0.588
## 2 Rural                 0.485
## 3 Suburb                0.561
## 4 Town                  0.478
```

```r
college %>%
  count(X1LOCALE_new,S3CLGFT_new_full_part) %>%
  group_by(X1LOCALE_new)%>%
  mutate(prop = prop.table(n)) %>%
  select(-n) %>%
  spread(S3CLGFT_new_full_part, prop)%>%kable()
```

X1LOCALE_new	0	1
City	0.4118702	0.5881298
Rural	0.5152006	0.4847994
Suburb	0.4392347	0.5607653
Town	0.5215208	0.4784792

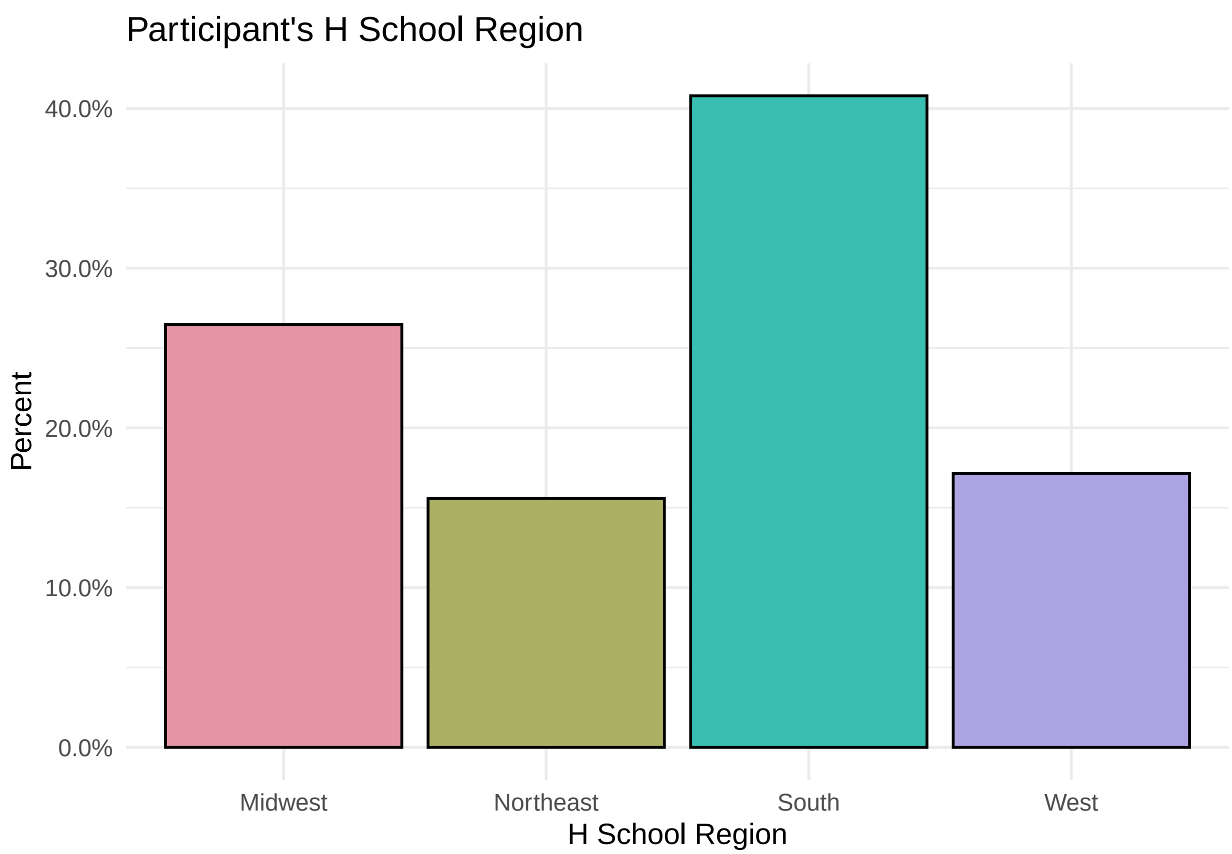
```
ggsumlocale<-ggplot(college_sumlocale, aes(x=fct_reorder (X1LOCALE_new,enrollment_avg), y=enrollment_avg
        geom_bar(stat="Identity", fill= rainbow_hcl(4), color="black")+
        geom_text(aes(label=round(enrollment_avg, 2)), position=position_dodge(width=.75), vjust=-0.25)
        labs(x= "H School Locale Type", y = "Enrollment",
        title = "College Enrollment by H School Locale Type")+
        scale_y_continuous(labels = scales::percent) +
         theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggsumlocale
```




```
ggregion<-ggplot(college,
   aes(x = X1REGION_new,
       y = ..count.. / sum(..count..))) +
  geom_bar(fill = rainbow_hcl(4), color= "black") +
  theme_minimal() +
  labs(x = "H School Region",
       y = "Percent",
```

```
  title  = "Participant's H School Region") +
  scale_y_continuous(labels = scales::percent)
ggregion
```




```
college_sumregion<-college%>%
  group_by(X1REGION_new)%>%
  summarise(enrollment_avg=mean(S3CLGFT_new_full_part))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
college_sumregion
```

```
## # A tibble: 4 x 2
##   X1REGION_new enrollment_avg
##   <chr>                 <dbl>
## 1 Midwest               0.567
## 2 Northeast             0.582
## 3 South                 0.523
## 4 West                  0.505
```

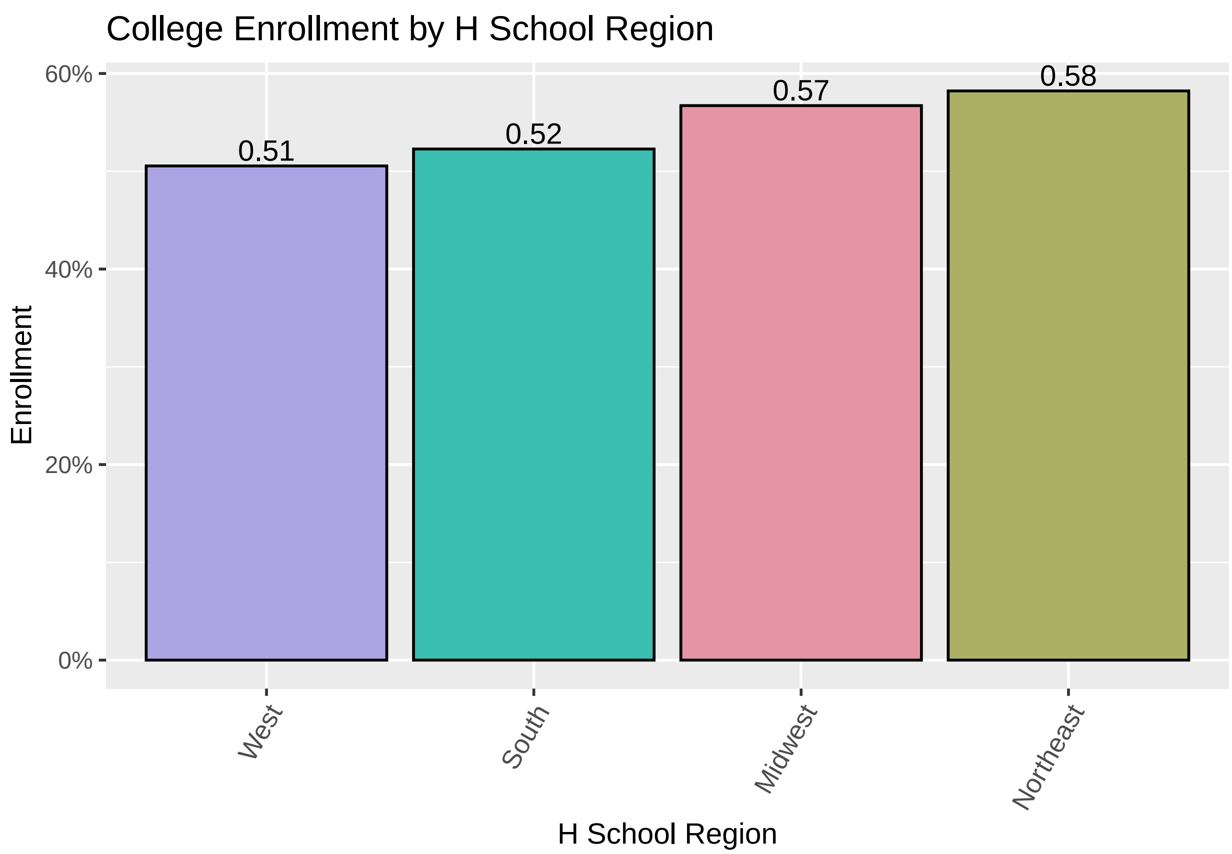
```
college %>%
  count(X1REGION_new,S3CLGFT_new_full_part) %>%
  group_by(X1REGION_new)%>%
```

```
  mutate(prop = prop.table(n)) %>%
  select(-n) %>%
  spread(S3CLGFT_new_full_part, prop)%>%kable()
```

X1REGION_new	0	1
Midwest	0.4328406	0.5671594
Northeast	0.4178045	0.5821955
South	0.4772087	0.5227913
West	0.4945409	0.5054591

```
ggsumregion<-ggplot(college_sumregion, aes(x=fct_reorder (X1REGION_new,enrollment_avg), y=enrollment_avg
        geom_bar(stat="Identity", fill= rainbow_hcl(4), color="black")+
  geom_text(aes(label=round(enrollment_avg, 2)), position=position_dodge(width=.75), vjust=-0.25)+
  labs(x= "H School Region", y = "Enrollment",
        title = "College Enrollment by H School Region")+
        scale_y_continuous(labels = scales::percent) +
         theme(axis.text.x = element_text(size=10,angle = 60, hjust = 1))


ggsumregion
```



Indices

We created 3 Indices - 1) College Prep Resources, 2) Parent & Family Support and 3) College Courses. Each of these was created simply by adding the binary variables together.

College Prep resources included five variables: 1) School has a college counselor, 2) School organizes student visits to colleges, 3) School participates in college fairs, 4) School consults with post secondary reps and 5) Student has met with high school counselor about college admissions. The max score was 5. The Family Support Index included four variables: 1) School assists students with finding financial aid for college, 2) School takes other steps to assist with HS to college transition, 3) School holds info session on transition to college, 4) Student has met with high school counselor about financial aid. The max score was 4.

The College Course Index included four variables: 1) the student has taken course for college credit while in high school, 2) the student has taken AP Courses, 3) the student has taken IB Courses and 4) the student has taken dual enrollment courses. However, since most students do not take both AP and IP courses, the max score was effectively 3.

```
college <- college%>%
              mutate (college_prep_index =C1CLGPREP_new + C1VISITCLG_new + C1CLGFAIR_new + C1POSTSECREQ_
college <- college%>%
              mutate (family_student_support_index = C1FINANCEAID_new + C1ASSISTOTH_new + C1INFOSESSN_ne
college <- college%>%
              mutate (courses_prep_index =S3ANYCLGCRED_new + S3AP_new  + S3IB_new + S3DUAL_new)
```

Our first logit model took into consideration our 3 indices without factoring for any other variable. We were not surprised that we had significant findings for all three indices due to the size of our sample. We used this information to refine our model and look at our effect size.

```
college_logit_mod1<-glm(S3CLGFT_new_full_part~
                       college_prep_index +
                       family_student_support_index +
                       courses_prep_index,
                   data=college,
                  family=binomial(link="logit"),
                     y=TRUE)

summary(college_logit_mod1)
```

```
##
## Call:
## glm(formula = S3CLGFT_new_full_part ~ college_prep_index + family_student_support_index +
##     courses_prep_index, family = binomial(link = "logit"), data = college,
##     y = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9923  -0.9428   0.2950   0.7159   1.9112
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.65082    0.04726 -34.932   <2e-16 ***
## college_prep_index            0.17654    0.01525  11.579   <2e-16 ***
## family_student_support_index  0.18205    0.02046   8.897   <2e-16 ***
## courses_prep_index            1.17049    0.01765  66.331   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 32425  on 23502  degrees of freedom
## Residual deviance: 24947  on 23499  degrees of freedom
## AIC: 24955
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(college_logit_mod1))
```

```
##                 (Intercept)          college_prep_index
##                   0.1918931                   1.1930824
## family_student_support_index          courses_prep_index
##                   1.1996791                   3.2235799
```

The index measures all had different ranges of values. In order to make sure that we were measuring their impact equally, we converted each index score to a z-score using the scale function. We also then converted the SES score to a percentile rank in order to make it more understandable to a general audience.

```
college <- college%>%
              mutate(college_prep_index_z=scale(college_prep_index,center = TRUE,scale=TRUE))

college <- college%>%
              mutate(family_student_support_index_z=scale(family_student_support_index,center = TRUE,sca

college <- college%>%
              mutate(courses_prep_index_z=scale(courses_prep_index,center = TRUE,scale=TRUE))
college <- college%>%
              mutate(X2SES_percent_rank=percent_rank(X2SES_new)*100)
college %>%
  select(STU_ID, college_prep_index_z, family_student_support_index_z, courses_prep_index_z, X2SES_perc
```

```
## # A tibble: 23,503 x 5
##    STU_ID college_prep_ind~ family_student_su~ courses_prep_in~ X2SES_percent_r~
##     <dbl>             <dbl>              <dbl>            <dbl>            <dbl>
##  1  10001             0.489              1.61             1.05              97.5
##  2  10002             1.23               0.618           -0.752             33.1
##  3  10003             1.23               1.61            -0.752             87.4
##  4  10004            -2.47              -2.37            -0.752             32.9
##  5  10005             0.489             -0.377            1.05              86.0
##  6  10006            -0.251              0.618            1.05              63.6
##  7  10007             0.489              0.618           -0.752             62.3
##  8  10008             0.489              0.618            1.95              72.2
##  9  10009             1.23               0.618            1.95              59.4
## 10  10010             0.489              0.618            1.05              61.2
## # ... with 23,493 more rows
```

Comparison Regressions

For the sake of creating comparisons between racial groups and to gain insight into the possible effects of our index measures, we ran a series of simpler regressions that then added in additional controls. First we ran a model where college enrollment was only analyzed as a function of race and sex. Second we then added a control for SES, and then we could compare these racial results against Model 3, where we have our index measures, controlled for Race, Sex, SES. Model 2 is also an intermediary model where we did not include SES. In each model, we also exponentiated the coefficients in order to get the odds ratios for each variable.

```
college_logit_mod2race<-glm(S3CLGFT_new_full_part~
                       as.character(X2RACE_new) +
                        as.character (X2SEX),
                   data=college,
                  family=binomial(link="logit"),
                     y=TRUE)

summary(college_logit_mod2race)
```

```
##
## Call:
## glm(formula = S3CLGFT_new_full_part ~ as.character(X2RACE_new) +
##     as.character(X2SEX), family = binomial(link = "logit"), data = college,
##     y = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5763  -1.2128   0.9685   1.1424   1.5708
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.08288    0.02186   3.791  0.00015 ***
## as.character(X2RACE_new)2  0.43064    0.05180   8.313  < 2e-16 ***
## as.character(X2RACE_new)3 -0.30492    0.04436  -6.874 6.22e-12 ***
## as.character(X2RACE_new)4 -0.97247    0.13644  -7.128 1.02e-12 ***
## as.character(X2RACE_new)5 -0.43932    0.03799 -11.564  < 2e-16 ***
## as.character(X2RACE_new)6 -0.25920    0.04812  -5.387 7.18e-08 ***
## as.character(X2RACE_new)7 -0.47963    0.18679  -2.568  0.01024 *
## as.character(X2RACE_new)9 -0.86305    0.15715  -5.492 3.98e-08 ***
## as.character(X2SEX)2       0.38809    0.02653  14.627  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 32425  on 23502  degrees of freedom
## Residual deviance: 31839  on 23494  degrees of freedom
## AIC: 31857
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(college_logit_mod2race))
```

```
##               (Intercept) as.character(X2RACE_new)2 as.character(X2RACE_new)3
```

```
##                  1.0864091                  1.5382392                  0.7371831
## as.character(X2RACE_new)4 as.character(X2RACE_new)5 as.character(X2RACE_new)6
##                  0.3781487                  0.6444755                  0.7716662
## as.character(X2RACE_new)7 as.character(X2RACE_new)9      as.character(X2SEX)2
##                  0.6190122                  0.4218745                  1.4741657
```

```r
college_logit_mod2raceSES<-glm(S3CLGFT_new_full_part~
                       X2SES_percent_rank +
                       as.character(X2RACE_new) +
                        as.character (X2SEX),
                   data=college,
                  family=binomial(link="logit"),
                     y=TRUE)

summary(college_logit_mod2raceSES)
```

```
##
## Call:
## glm(formula = S3CLGFT_new_full_part ~ X2SES_percent_rank + as.character(X2RACE_new) +
##     as.character(X2SEX), family = binomial(link = "logit"), data = college,
##     y = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1829  -1.0964   0.6658   0.9723   1.8719
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.0599394  0.0385009 -27.530  < 2e-16 ***
## X2SES_percent_rank        0.0241363  0.0005611  43.017  < 2e-16 ***
## as.character(X2RACE_new)2  0.5082778  0.0607598   8.365  < 2e-16 ***
## as.character(X2RACE_new)3 -0.0299794  0.0501589  -0.598  0.55005
## as.character(X2RACE_new)4 -0.1608029  0.1588925  -1.012  0.31153
## as.character(X2RACE_new)5  0.0013298  0.0437270   0.030  0.97574
## as.character(X2RACE_new)6 -0.1410659  0.0539849  -2.613  0.00897 **
## as.character(X2RACE_new)7 -0.2442735  0.2094960  -1.166  0.24361
## as.character(X2RACE_new)9 -0.5300633  0.1785981  -2.968  0.00300 **
## as.character(X2SEX)2      0.4266443  0.0299230  14.258  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28433  on 20918  degrees of freedom
## Residual deviance: 25887  on 20909  degrees of freedom
##   (2584 observations deleted due to missingness)
## AIC: 25907
##
## Number of Fisher Scoring iterations: 4
```

```r
exp(coef(college_logit_mod2raceSES))
```

```
##               (Intercept)        X2SES_percent_rank as.character(X2RACE_new)2
```

```
##                 0.3464768                 1.0244299                 1.6624258
## as.character(X2RACE_new)3 as.character(X2RACE_new)4 as.character(X2RACE_new)5
##                 0.9704655                 0.8514599                 1.0013306
## as.character(X2RACE_new)6 as.character(X2RACE_new)7 as.character(X2RACE_new)9
##                 0.8684321                 0.7832734                 0.5885677
##      as.character(X2SEX)2
##                 1.5321076
```

```r
college_logit_mod2<-glm(S3CLGFT_new_full_part~
                       college_prep_index_z +
                       family_student_support_index_z +
                       courses_prep_index_z+
                       as.character(X2RACE_new) +
                        as.character (X2SEX),
                   data=college,
                  family=binomial(link="logit"),
                     y=TRUE)

summary(college_logit_mod2)
```

```
##
## Call:
## glm(formula = S3CLGFT_new_full_part ~ college_prep_index_z +
##     family_student_support_index_z + courses_prep_index_z + as.character(X2RACE_new) +
##     as.character(X2SEX), family = binomial(link = "logit"), data = college,
##     y = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8745  -0.9172   0.2982   0.7059   2.1691
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     0.30562    0.02646  11.548  < 2e-16 ***
## college_prep_index_z            0.23920    0.02069  11.560  < 2e-16 ***
## family_student_support_index_z  0.18385    0.02067   8.894  < 2e-16 ***
## courses_prep_index_z            1.27993    0.01977  64.725  < 2e-16 ***
## as.character(X2RACE_new)2       0.13826    0.06242   2.215 0.026755 *
## as.character(X2RACE_new)3      -0.02200    0.05044  -0.436 0.662762
## as.character(X2RACE_new)4      -0.56894    0.15676  -3.629 0.000284 ***
## as.character(X2RACE_new)5      -0.31616    0.04456  -7.095 1.29e-12 ***
## as.character(X2RACE_new)6      -0.25030    0.05684  -4.403 1.07e-05 ***
## as.character(X2RACE_new)7      -0.43602    0.21923  -1.989 0.046712 *
## as.character(X2RACE_new)9      -0.72551    0.18078  -4.013 5.99e-05 ***
## as.character(X2SEX)2            0.24358    0.03110   7.833 4.76e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 32425  on 23502  degrees of freedom
## Residual deviance: 24784  on 23491  degrees of freedom
## AIC: 24808
##
```

```
## Number of Fisher Scoring iterations: 4
```

```r
exp(coef(college_logit_mod2))
```

```
##                 (Intercept)          college_prep_index_z
##                   1.3574679                     1.2702350
## family_student_support_index_z          courses_prep_index_z
##                   1.2018359                     3.5963785
##      as.character(X2RACE_new)2      as.character(X2RACE_new)3
##                   1.1482705                     0.9782450
##      as.character(X2RACE_new)4      as.character(X2RACE_new)5
##                   0.5661250                     0.7289445
##      as.character(X2RACE_new)6      as.character(X2RACE_new)7
##                   0.7785683                     0.6466058
##      as.character(X2RACE_new)9          as.character(X2SEX)2
##                   0.4840765                     1.2758140
```

```r
college_logit_mod3<-glm(S3CLGFT_new_full_part~
                       college_prep_index_z +
                       family_student_support_index_z +
                       courses_prep_index_z+ X2SES_percent_rank +
                       as.character(X2RACE_new) +
                        as.character (X2SEX),
                   data=college,
                  family=binomial(link="logit"),
                     y=TRUE)

summary(college_logit_mod3)
```

```
##
## Call:
## glm(formula = S3CLGFT_new_full_part ~ college_prep_index_z +
##     family_student_support_index_z + courses_prep_index_z + X2SES_percent_rank +
##     as.character(X2RACE_new) + as.character(X2SEX), family = binomial(link = "logit"),
##     data = college, y = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9653  -0.8761   0.3447   0.7791   2.3098
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -0.6271555  0.0439164 -14.281  < 2e-16 ***
## college_prep_index_z            0.2144285  0.0222755   9.626  < 2e-16 ***
## family_student_support_index_z  0.1837420  0.0221855   8.282  < 2e-16 ***
## courses_prep_index_z            1.1062987  0.0206420  53.594  < 2e-16 ***
## X2SES_percent_rank              0.0190302  0.0006363  29.909  < 2e-16 ***
## as.character(X2RACE_new)2       0.1905763  0.0691393   2.756  0.00584 **
## as.character(X2RACE_new)3       0.1560888  0.0548688   2.845  0.00444 **
## as.character(X2RACE_new)4      -0.0609732  0.1737976  -0.351  0.72572
## as.character(X2RACE_new)5       0.0122566  0.0491765   0.249  0.80318
## as.character(X2RACE_new)6      -0.1610198  0.0610830  -2.636  0.00839 **
## as.character(X2RACE_new)7      -0.2843424  0.2375749  -1.197  0.23136
```

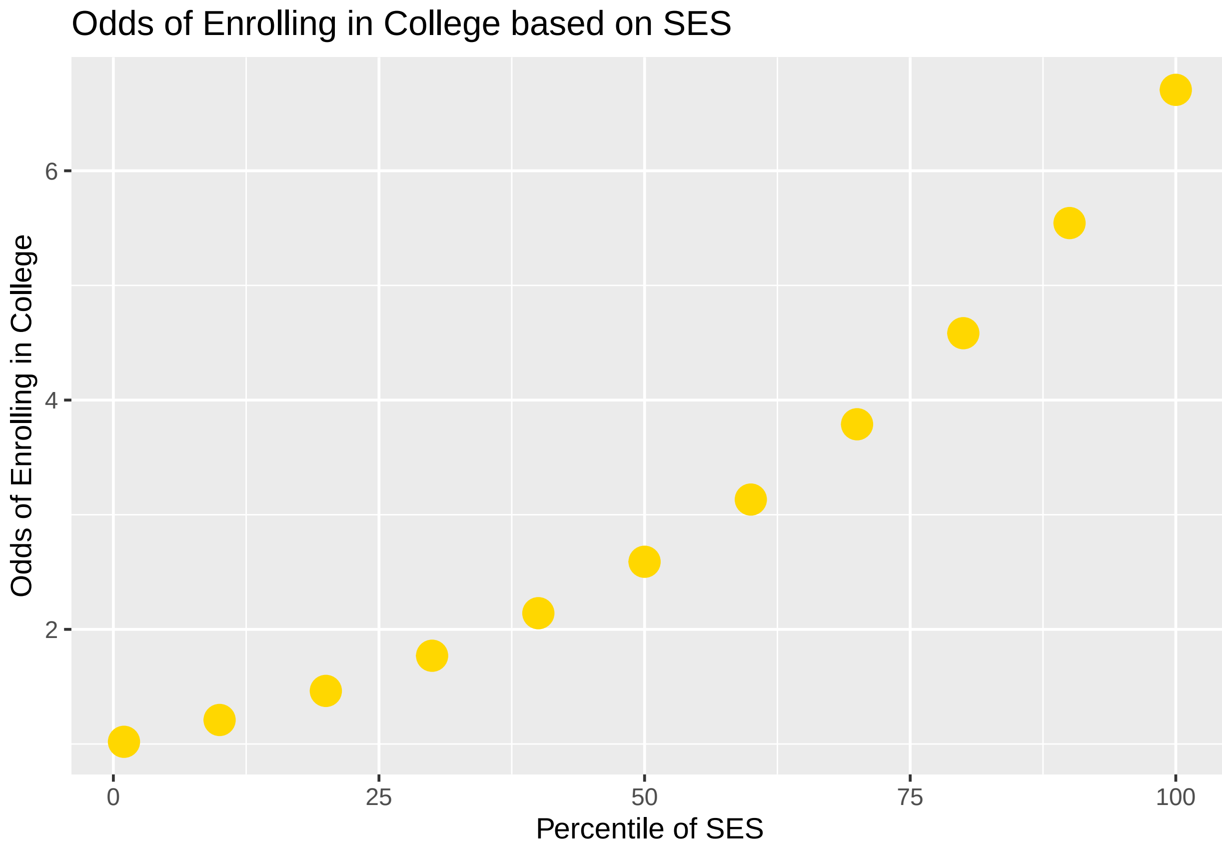
```
## as.character(X2RACE_new)9      -0.4693282  0.1971316  -2.381  0.01728 *
## as.character(X2SEX)2             0.2816867  0.0336097   8.381  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28433  on 20918  degrees of freedom
## Residual deviance: 21370  on 20906  degrees of freedom
##   (2584 observations deleted due to missingness)
## AIC: 21396
##
## Number of Fisher Scoring iterations: 5
```

```
exp(coef(college_logit_mod3))
```

```
##                    (Intercept)           college_prep_index_z
##                      0.5341089                      1.2391536
## family_student_support_index_z           courses_prep_index_z
##                      1.2017057                      3.0231481
##              X2SES_percent_rank      as.character(X2RACE_new)2
##                      1.0192124                      1.2099466
##      as.character(X2RACE_new)3      as.character(X2RACE_new)4
##                      1.1689301                      0.9408485
##      as.character(X2RACE_new)5      as.character(X2RACE_new)6
##                      1.0123320                      0.8512752
##      as.character(X2RACE_new)7      as.character(X2RACE_new)9
##                      0.7525089                      0.6254223
##           as.character(X2SEX)2
##                      1.3253634
```

This graphic illustrates the Odds of enrolling in college based solely on SES. The odds for each percentile were calculated by raising to that power and we graphed this curve.

```
ggses<-ggplot(percentile,aes(x=Percentile,y=Ofe))
ggses<-ggses+geom_point(color="gold",size=5)
ggses<-ggses+xlab("Percentile of SES")+ylab("Odds of Enrolling in College")+ ggtitle("Odds of Enrolling
ggses
```



In order to show the effects of our comparison models, we created vectors from the exponentiated coefficients of each model and then joined them in a data frame. This includes a gather function so that we could change the shape of the data and graph the results in a multi-line line chart.

```
expoutput2race=c(exp(coef(college_logit_mod2race)))
expoutput2raceSES=c(exp(coef(college_logit_mod2raceSES)))
expoutput3=c(exp(coef(college_logit_mod3)))

df1=data.frame(expoutput2race)
df2=data.frame(expoutput2raceSES)
df3=data.frame(expoutput3)


df4=df1[2:8,]
df5=df2[3:9,]
df6=df3[6:12,]

df7=data.frame(df4,df5,df6)

Group<-c("Asian", "Black", "Hispanic no race", "Hispanic race", "More than one race", "Pacific Islander

df8=data.frame(Group, df7)
df8long=gather(df8,condition,measurement,df4:df6,factor_key=TRUE)
df8long=df8long%>%

            mutate (condition_new=recode(condition, ‘df4‘=’Race’, ‘df5‘ =’Race+SES’, ‘df6‘ =’Race+SES+P
```

```r
df8long
```

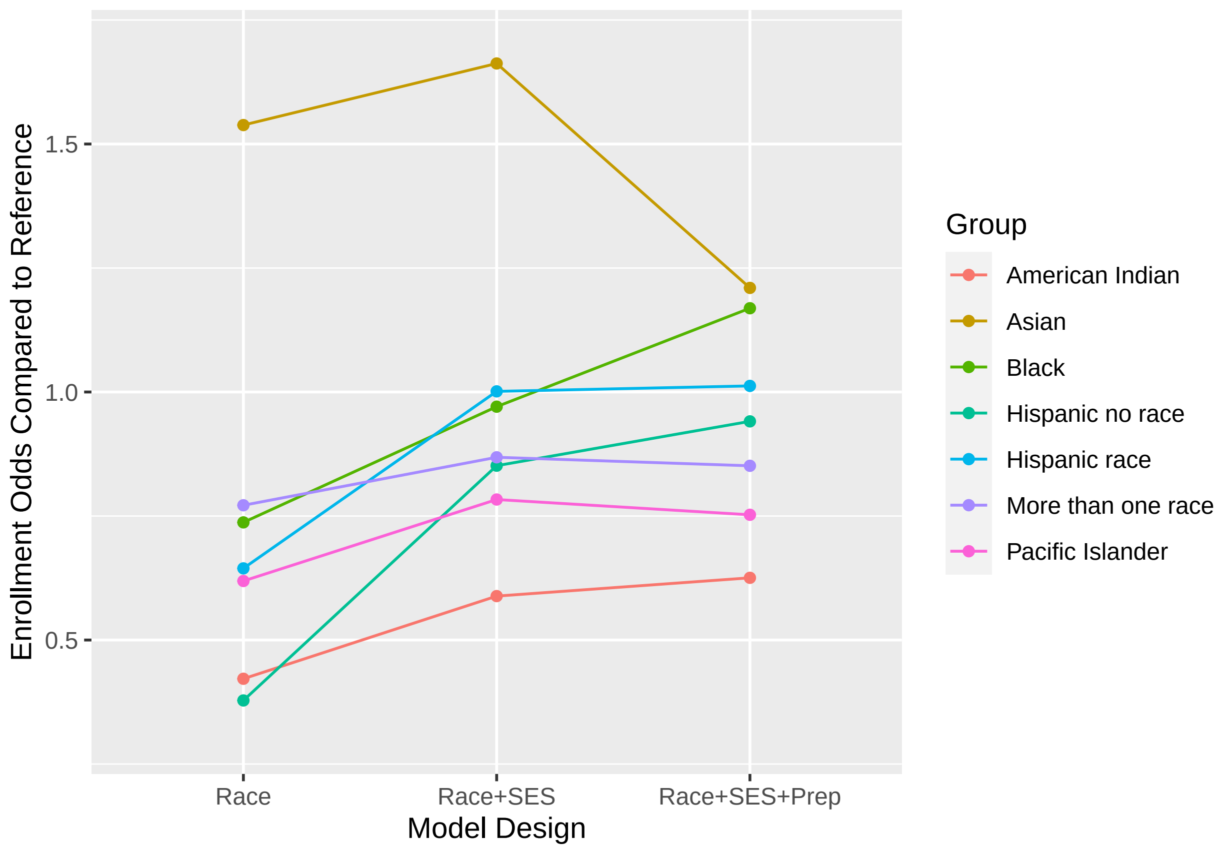
```
##                 Group condition measurement condition_new
## 1               Asian       df4   1.5382392          Race
## 2               Black       df4   0.7371831          Race
## 3    Hispanic no race       df4   0.3781487          Race
## 4       Hispanic race       df4   0.6444755          Race
## 5  More than one race       df4   0.7716662          Race
## 6    Pacific Islander       df4   0.6190122          Race
## 7     American Indian       df4   0.4218745          Race
## 8               Asian       df5   1.6624258      Race+SES
## 9               Black       df5   0.9704655      Race+SES
## 10   Hispanic no race       df5   0.8514599      Race+SES
## 11      Hispanic race       df5   1.0013306      Race+SES
## 12 More than one race       df5   0.8684321      Race+SES
## 13   Pacific Islander       df5   0.7832734      Race+SES
## 14    American Indian       df5   0.5885677      Race+SES
## 15              Asian       df6   1.2099466 Race+SES+Prep
## 16              Black       df6   1.1689301 Race+SES+Prep
## 17   Hispanic no race       df6   0.9408485 Race+SES+Prep
## 18      Hispanic race       df6   1.0123320 Race+SES+Prep
## 19 More than one race       df6   0.8512752 Race+SES+Prep
## 20   Pacific Islander       df6   0.7525089 Race+SES+Prep
## 21    American Indian       df6   0.6254223 Race+SES+Prep
```

```r
ggacc<-ggplot(df8long,aes(x=condition_new,y= measurement, group=Group, colour=Group))+

        geom_line()+

        geom_point()+ylim(0.3,1.7)+xlab("Model Design")+ylab("Enrollment Odds Compared to Reference")

ggacc
```



Conclusion & Recommendations

Through our exploration and logit model, we found that students had 3 times greater odds of enrolling in college if they took AP, IB, Dual Credit courses while in high school. These odds remained even when controlling for Race, Sex, and Socio-economic status, our findings remained significant. Students at the top of the SES spectrum had 6 times greater odds of enrolling in college than those at the bottom. When you control for SES African Am students have higher odds than other races for enrolling at slightly higher rates. Groups 2 (Asian) and 3 (African Americans) had slightly higher odds than whites for enrolling in college. Our data shows that access to and success in these types of courses is a leveling factor in student success and college enrollment. Preparing students to take these types of courses, and facilitating the actual taking of the course, is a key lever that schools can take to increase the number of students that enroll in college.

For educators contemplating funding/strategic direction in high school programs, the data provides food for thought. This data set provides availability for additional studies into college entrance and perseverance.

References

National Center for Education Statistics. (2001).Paving the Way to Postsecondary (NCES 2001-205).U.S. Department of Education.https://nces.ed.gov/pubs2001/2001205.pdf

National Center for Education Statistics (NCES) Home Page, part of the U.S. Department of Education. (2019). Retrieved October 12, 2020, from Ed.gov website: https://nces.ed.gov/