

# cal\_schools analysis

2024-04-10

## Overview

According to National Center of Education Statistics, “Higher levels of educational attainment are positively correlated with many outcomes, such as employment, earnings, and health.” (National Center for Education Statistics). The purpose of this analysis is to examine data publicly available through California Department of Education’s website (cde.ca.gov) to analyze where educational attainment gaps may exist, among different demographic groups.

```
# import libraries

library(DBI)
library(RODBC)
library(odbc)
library(RPostgres)
suppressPackageStartupMessages(library(tidyverse))

# set theme
theme_set(theme_minimal())
```

## Part I

### Establish Connection with Database

In order to carry out the analysis, a connection will be established with Postgres cal\_schools database. DBI, RPostgres, RODBC, odbc packages will be used for this process.

```
# create connection object / establish connection with Postgres

con <- dbConnect(RPostgres::Postgres(),
                 dbname = 'cal_schools',
                 port = 5433,
                 user = 'postgres',
                 password = 'erikchris')

# query database/ create table with high level stats (college going rate,
## enrollment, etc), along with school description info

school_stats <- dbGetQuery(con,
"WITH
  enr_school (cds_code, enr_total) AS
  (SELECT cds_code, SUM(enr_total) FROM enrollment
```

```

GROUP BY cds_code),

frmp_schools (cds_code, frmp_el) AS
(SELECT cds_code, AVG(frmp_count::numeric/enr::numeric * 100)
FROM fr_lunch
GROUP BY cds_code),

ell_schools(cds_code, ell_count) AS
(SELECT cds_code, SUM(total_ell)
FROM ell
GROUP BY cds_code),

cgr_schools (cds_code,hs_completers, enr_college, cgr) AS
(SELECT cds_code, SUM(hs_completers), SUM(enr_college),
AVG(enr_college::numeric/ hs_completers::numeric * 100)
FROM college_gr
WHERE reporting_cat = 'Total' AND completer_type = 'TA'
GROUP BY cds_code),

grad_schools (cds_code, grad_count) AS
(SELECT cds_code, SUM(grad_count)
FROM graduates
WHERE reporting_cat = 'Total'
GROUP BY cds_code),

abs_schools (cds_code, unexc_abs_pct) AS
(SELECT cds_code, AVG(unexc_absences_pct)
FROM absents
WHERE reporting_cat = 'Total'
GROUP BY cds_code
)

SELECT counties.county_code, county_name, schools.cds_code, district_name,
school_name, status_type, street, city, zip, phone, email, website, open_date,
close_date, charter, soc, soc_type, ed_ops_name, eil_name, gs_offered, virtual,
year_round, latitude, longitude, adm_email, last_update, district_code, enr_total,
frmp_el, ell_count, (ell_count::numeric/enr_total::numeric * 100) AS ell_pct,
unexc_abs_pct, grad_count, cgr
FROM schools
LEFT JOIN enr_school enr
USING(cds_code)
LEFT JOIN frmp_schools fr
ON fr.cds_code = enr.cds_code
LEFT JOIN ell_schools ell
ON fr.cds_code = ell.cds_code
LEFT JOIN cgr_schools cgr
ON ell.cds_code = cgr.cds_code
LEFT JOIN grad_schools gr
ON cgr.cds_code = gr.cds_code
LEFT JOIN abs_schools abs
ON gr.cds_code = abs.cds_code
JOIN counties
ON schools.county_code = counties.county_code;"

```

```
# query database/ create table that has detailed information about college going
## rates
```

```
college_gr <- dbGetQuery(con, "SELECT ac_year, cds_code, co.county_code,
                                county_name, school_name, reporting_cat, completer_type,
                                hs_completers, enr_college, enr_instate, enr_oos,
                                not_enr_college, uc_enr, csu_enr, ccc_enr,
                                enr_instate_private, enr_oos_4yr, enr_oos_2yr
                                FROM college_gr cgr
                                LEFT JOIN schools s
                                USING(cds_code)
                                LEFT JOIN counties co
                                ON cgr.county_code = co.county_code
                                WHERE cds_code NOT IN (SELECT cds_code FROM schools
                                WHERE soc_type = 'Continuation High Schools') ;")
```

```
# create new column in college stats ; cgr
```

```
college_gr <- college_gr |>
  mutate(cgr = (enr_college/hs_completers*100))
```

```
# disconnect from database; move forward to analysis
```

```
dbDisconnect(con)
```

## Part II

### Exploratory Analysis

In this phase, high level summary statistics will be acquired to get information about tables and their respective variables. Additionally, histograms, density plots, and box plots will be created to visualize how distributions of college going rates differ among varying demographic groups. Additionally, scatter plots will be created to visualize the relationship between college going rates and other variables/ characteristics.

```
# summary statistics of school_stats
```

```
summary(school_stats)
```

```
## county_code      county_name      cds_code      district_name
## Length:18335     Length:18335     Length:18335     Length:18335
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## school_name      status_type      street      city
## Length:18335     Length:18335     Length:18335     Length:18335
## Class :character  Class :pq_sts     Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
```

```

##
##
##
##      zip                phone                email                website
## Length:18335          Length:18335          Length:18335          Length:18335
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      open_date          close_date          charter          soc
## Min.   :1850-07-01      Min.   :1980-07-01      Mode :logical          Length:18335
## 1st Qu.:1980-07-01      1st Qu.:1989-06-30      FALSE:14934            Class :character
## Median :1980-07-01      Median :1994-06-03      TRUE :2007              Mode  :character
## Mean   :1992-02-15      Mean   :1999-05-10      NA's :1394
## 3rd Qu.:2003-08-28      3rd Qu.:2010-06-17
## Max.   :2069-09-01      Max.   :2024-01-05
## NA's   :1394            NA's   :12048
##      soc_type          ed_ops_name          eil_name          gs_offered
## Length:18335          Length:18335          Length:18335          Length:18335
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      virtual          year_round          latitude          longitude
## Length:18335          Mode :logical          Min.   :32.55          Min.   : -124.3
## Class :character      FALSE:15831          1st Qu.:33.99          1st Qu.: -121.7
## Mode  :character      TRUE :1110           Median :35.63          Median : -119.3
##                                     NA's :1394           Mean   :36.00          Mean   : -119.7
##                                     3rd Qu.:37.83          3rd Qu.: -118.0
##                                     Max.   :41.99          Max.   : -114.1
##                                     NA's   :4808          NA's   :4808
##      adm_email          last_update          district_code          enr_total
## Length:18335          Min.   :1999-06-24          Length:18335          Min.   :    1
## Class :character      1st Qu.:2011-12-12          Class :character      1st Qu.:   847
## Mode  :character      Median :2023-02-09          Mode  :character      Median : 1508
##                                     Mean   :2016-12-31          Mean   : 1849
##                                     3rd Qu.:2023-02-09          3rd Qu.: 2186
##                                     Max.   :2024-03-18          Max.   :53132
##                                     NA's   :7883
##      frmp_el          ell_count          ell_pct          unexc_abs_pct
## Min.   : 0.00          Min.   : 1          Min.   : 0.033          Min.   : 0.00
## 1st Qu.: 40.44          1st Qu.: 84          1st Qu.: 7.496          1st Qu.: 32.43
## Median : 69.01          Median : 246          Median : 16.525          Median : 52.27
## Mean   : 62.24          Mean   : 349          Mean   : 20.684          Mean   : 49.95
## 3rd Qu.: 86.16          3rd Qu.: 513          3rd Qu.: 30.326          3rd Qu.: 72.10
## Max.   :100.00          Max.   :4738          Max.   :100.000          Max.   :100.00
## NA's   :8429          NA's   :8642          NA's   :8642          NA's   :16145
##      grad_count          cgr
## Min.   : 9          Min.   : 0.00
## 1st Qu.: 143          1st Qu.:32.05

```

```
## Median : 399 Median :56.79
## Mean : 769 Mean :52.80
## 3rd Qu.: 1326 3rd Qu.:71.97
## Max. : 4028 Max. :96.04
## NA's :16143 NA's :16142
```

```
# summary statistics of school_stats
```

```
summary(college_gr)
```

```
## ac_year cds_code county_code county_name
## Length:413135 Length:413135 Length:413135 Length:413135
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## school_name reporting_cat completer_type hs_completers
## Length:413135 Length:413135 Length:413135 Min. : 11.00
## Class :character Class :character Class :character 1st Qu.: 23.00
## Mode :character Mode :character Mode :character Median : 51.00
## Mean : 93.02
## 3rd Qu.: 123.00
## Max. :1134.00
## enr_college enr_instate enr_oos not_enr_college
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 0.00
## 1st Qu.: 12.00 1st Qu.: 11.00 1st Qu.: 0.000 1st Qu.: 8.00
## Median : 30.00 Median : 28.00 Median : 2.000 Median : 17.00
## Mean : 61.14 Mean : 55.21 Mean : 5.928 Mean : 31.88
## 3rd Qu.: 79.00 3rd Qu.: 73.00 3rd Qu.: 5.000 3rd Qu.: 39.00
## Max. :882.00 Max. :825.00 Max. :262.000 Max. :647.00
## uc_enr csu_enr ccc_enr enr_instate_private
## Min. : 0.000 Min. : 0.00 Min. : 0.00 Min. : 0.000
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 7.00 1st Qu.: 0.000
## Median : 1.000 Median : 3.00 Median : 17.00 Median : 1.000
## Mean : 6.815 Mean : 11.74 Mean : 33.86 Mean : 2.787
## 3rd Qu.: 8.000 3rd Qu.: 15.00 3rd Qu.: 45.00 3rd Qu.: 3.000
## Max. :286.000 Max. :270.00 Max. :459.00 Max. :93.000
## enr_oos_4yr enr_oos_2yr cgr
## Min. : 0.000 Min. : 0.0000 Min. : 0.00
## 1st Qu.: 0.000 1st Qu.: 0.0000 1st Qu.: 45.24
## Median : 1.000 Median : 0.0000 Median : 64.06
## Mean : 5.465 Mean : 0.4624 Mean : 61.54
## 3rd Qu.: 4.000 3rd Qu.: 1.0000 3rd Qu.: 80.65
## Max. :259.000 Max. :16.0000 Max. :100.00
```

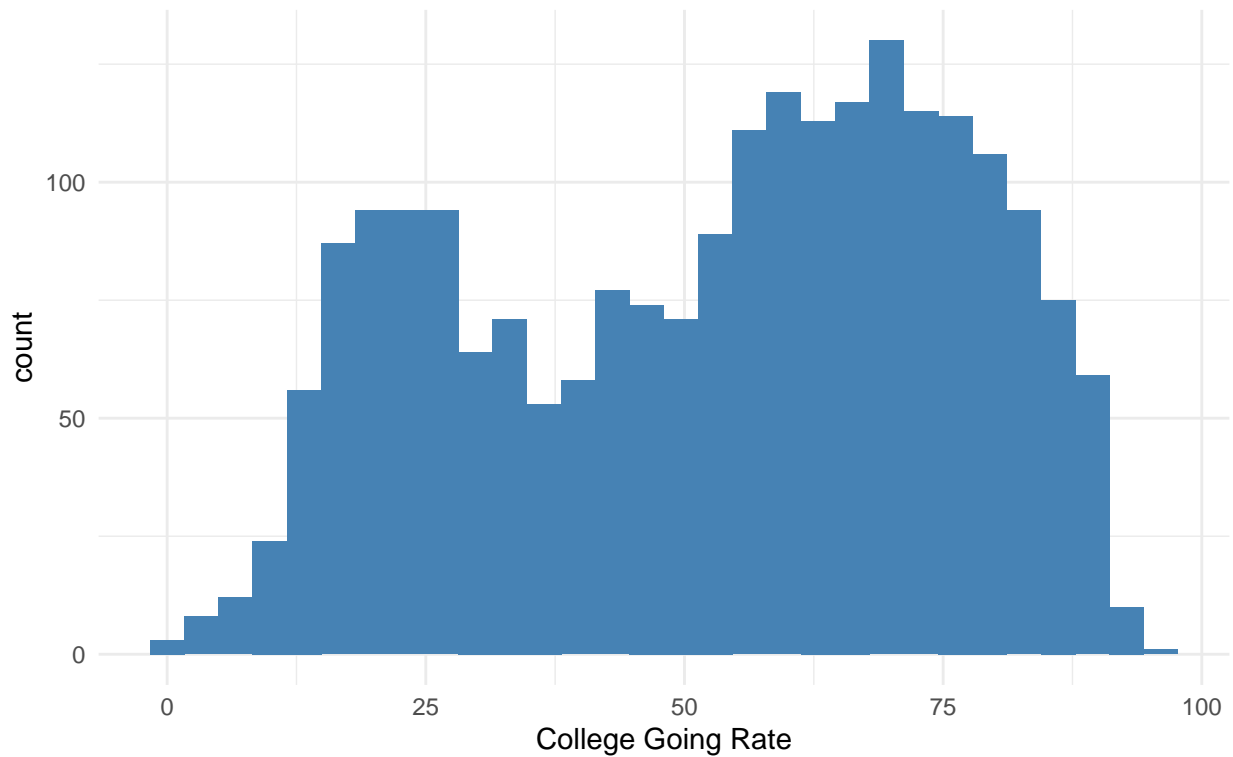
```
# histogram of college going rate
```

```
ggplot(school_stats, aes(cgr))+
  geom_histogram(fill = 'steelblue')+
  labs(x = 'College Going Rate',
       title = 'Histogram',
       subtitle = 'College Going Rates in California')
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 16142 rows containing non-finite values ('stat_bin()').
```

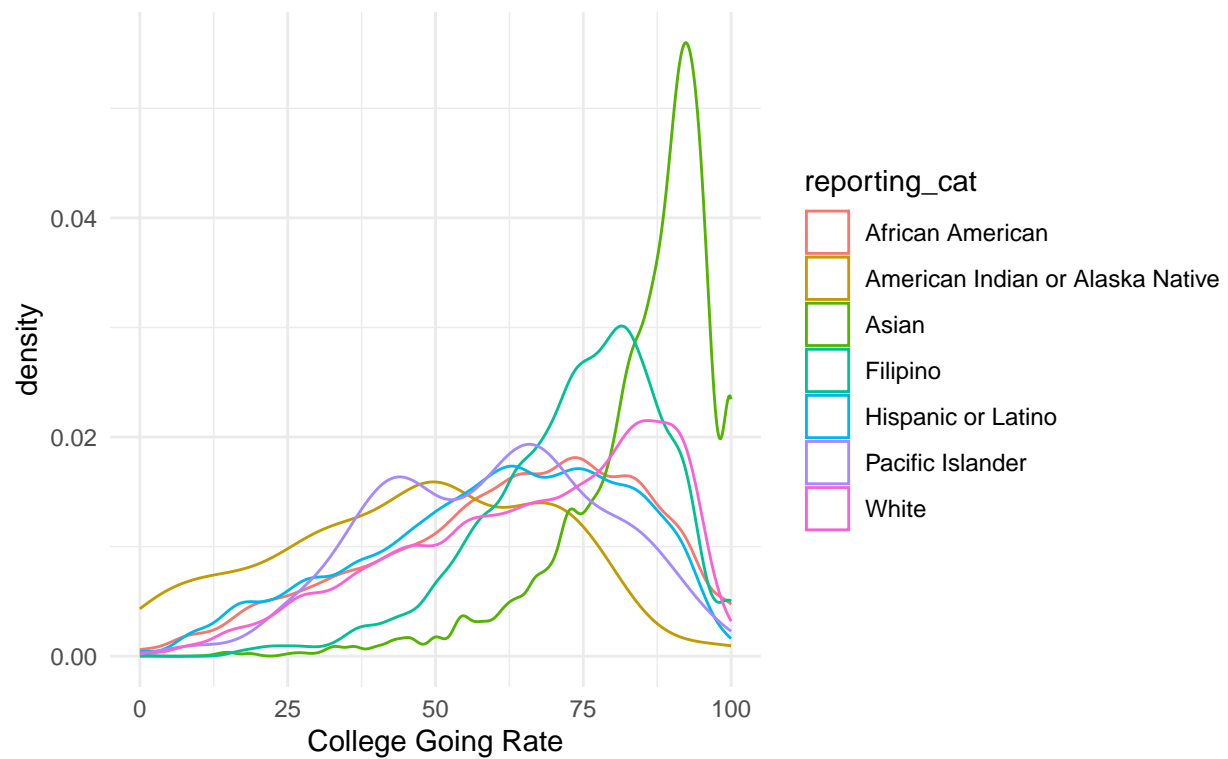
Histogram  
College Going Rates in California



```
# density plots of college going rates grouped by demographic info
```

```
college_gr |>
  filter(reporting_cat %in% c('African American', 'Filipino', 'Pacific Islander', 'American Indian or A
                              'Asian', 'Hispanic or Latino', 'White'))|>
  ggplot(aes(cgr, color = reporting_cat ))+
  geom_density()+
  labs(x = 'College Going Rate',
       title = "Density Plot",
       subtitle = 'College Going Rates by Race')
```

Density Plot  
College Going Rates by Race

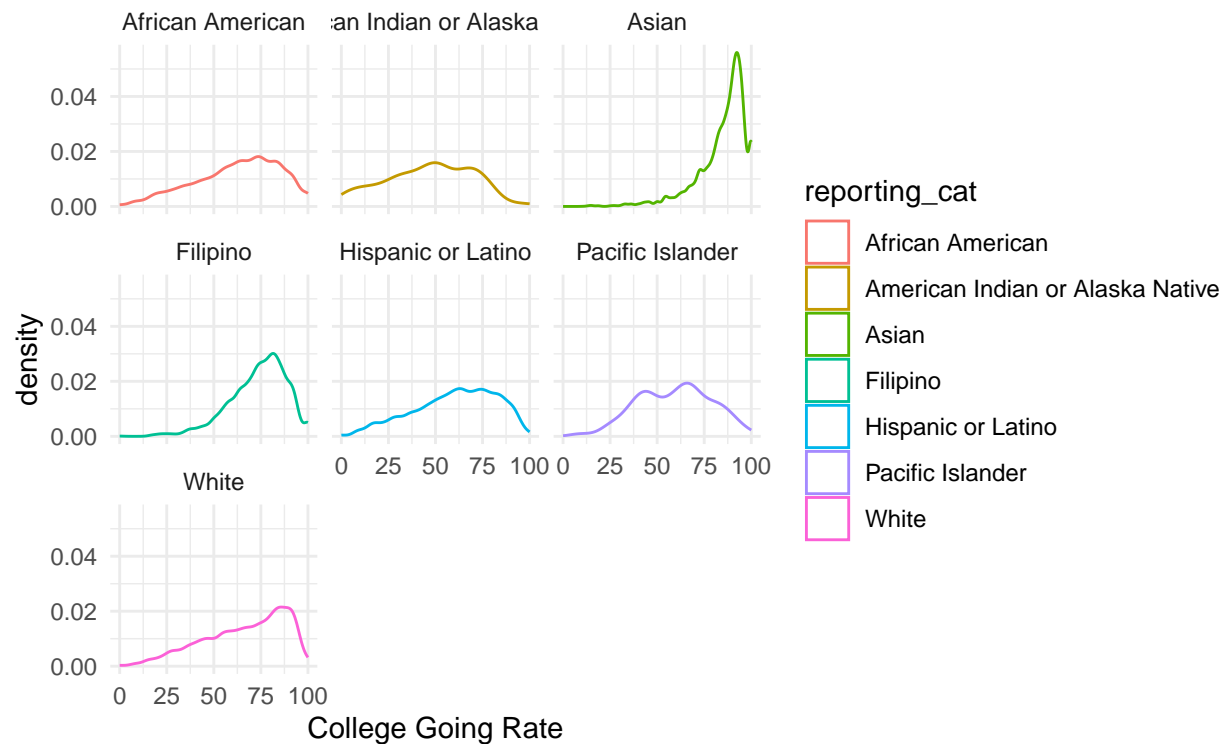


*## multi-facet density plot to see each distribution more clearly*

```
college_gr |>
  filter(reporting_cat %in% c('African American', 'Filipino', 'Pacific Islander', 'American Indian or A
  ggplot(aes(cgr, color = reporting_cat ))+
  geom_density()+
  facet_wrap(~reporting_cat)+
  labs(x = 'College Going Rate',
       title = 'Density Plots',
       subtitle = 'College Going Rates by Race')
```

## Density Plots

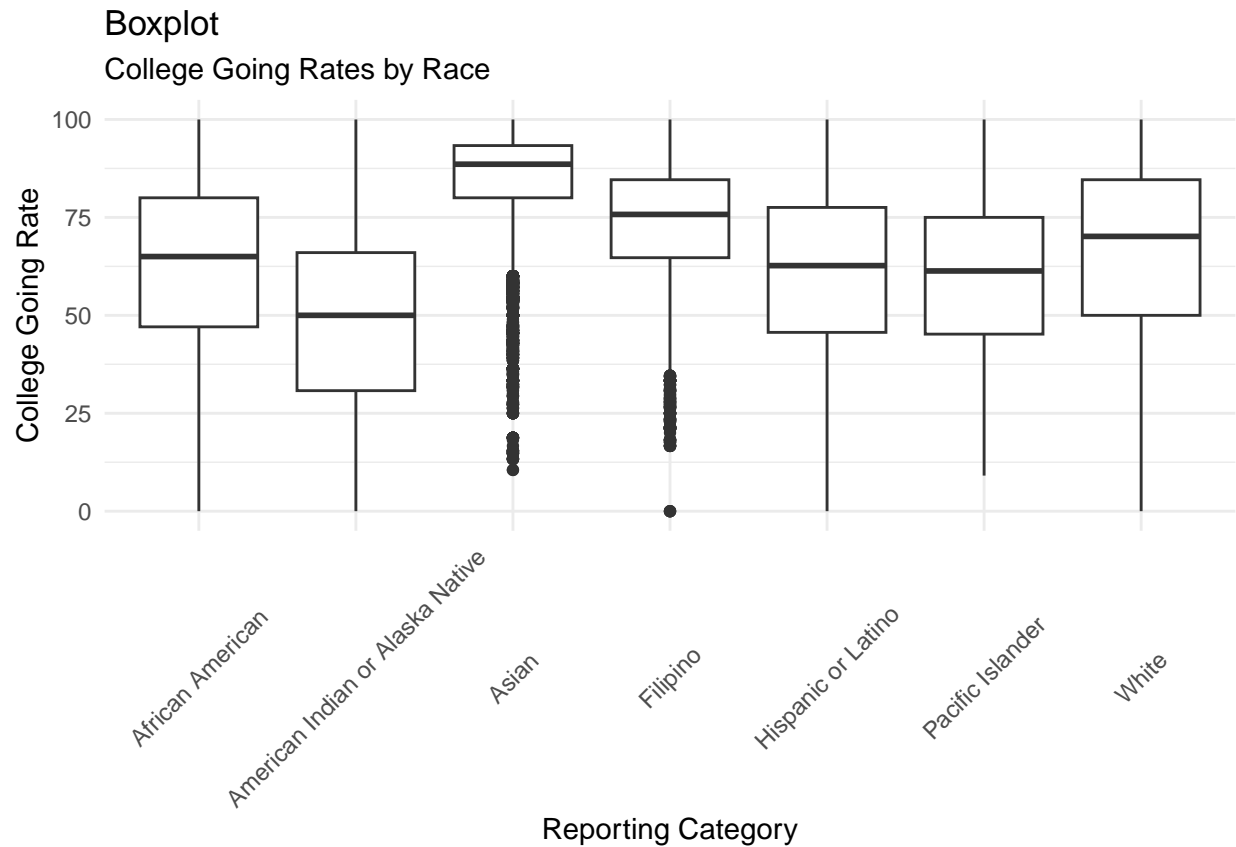
### College Going Rates by Race



*# box plots of college going rate grouped by demographic info*

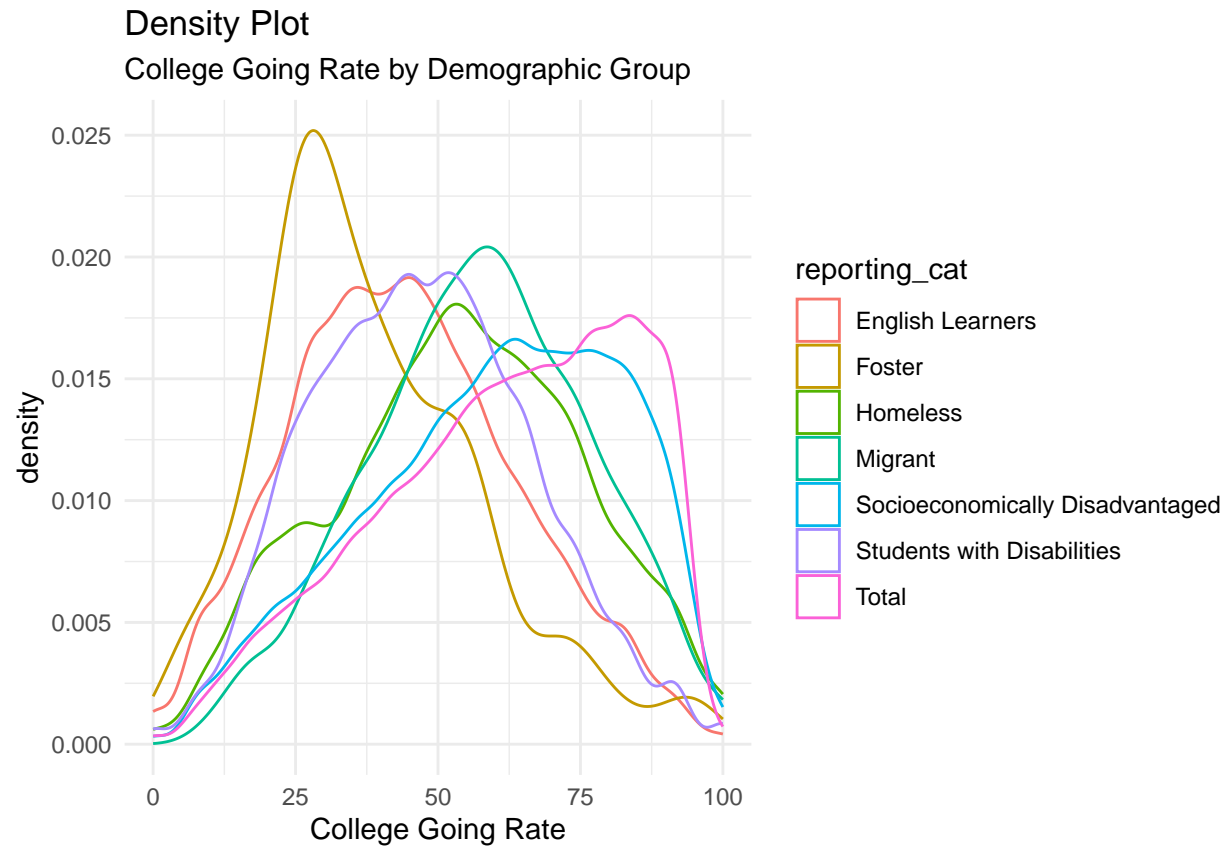
```
college_gr |>
  filter(reporting_cat %in% c('African American', 'Filipino', 'Pacific Islander', 'American Indian or Alaska Native',
                             'Asian', 'Hispanic or Latino', 'White'))|>
  ggplot( aes(reporting_cat, cgr))+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5))+
  labs(x = 'Reporting Category',
       y = 'College Going Rate',
       title = 'Boxplot',
       subtitle = 'College Going Rates by Race')
```





*# density plots of college going rates grouped by additional demographic characteristics (ell, socioeco*

```
college_gr |>
  filter(reporting_cat %in% c('Students with Disabilities', 'Total', 'Homeless',
    'Socioeconomically Disadvantaged', 'Migrant', 'Foster', 'English Learners', 'Total'))|>
  ggplot(aes(cgr, color = reporting_cat))+
  geom_density()+
  labs(x = 'College Going Rate',
    title = 'Density Plot',
    subtitle = 'College Going Rate by Demographic Group')
```

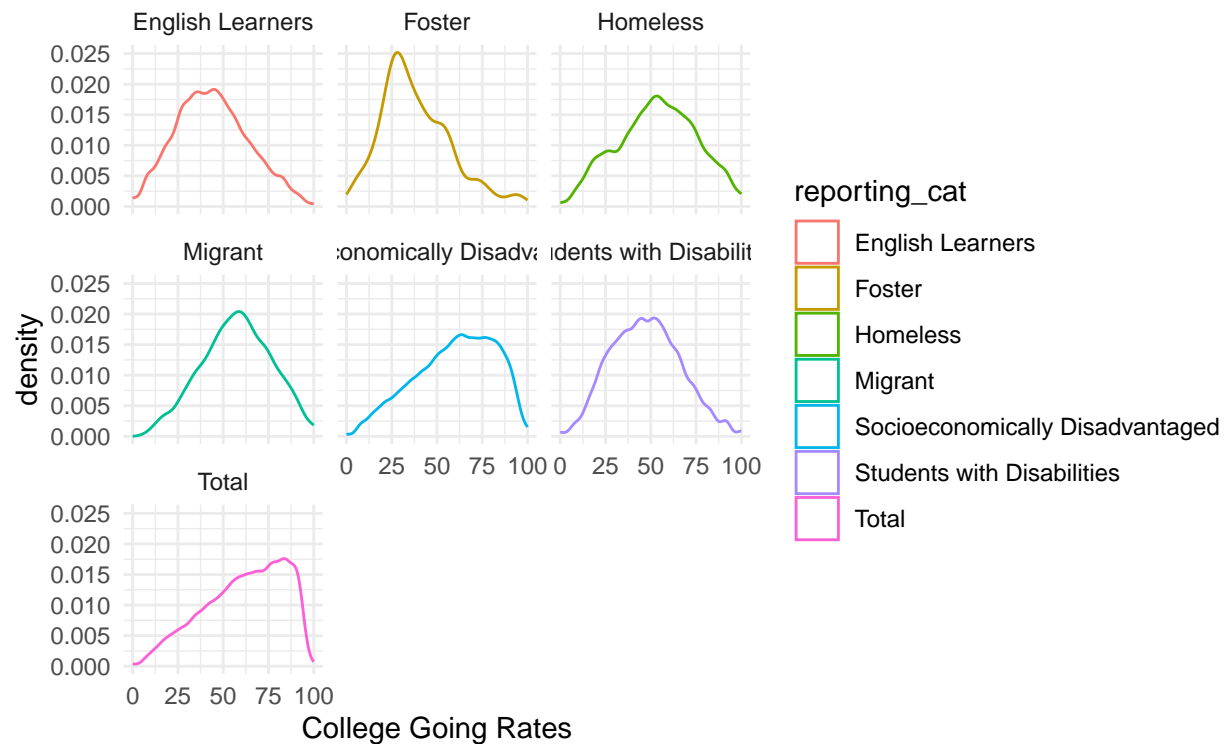


```
## multi-facet density plots to see each distribution more clearly

college_gr |>
  filter(reporting_cat %in% c('Students with Disabilities', 'Total', 'Homeless',
    'Socioeconomically Disadvantaged', 'Migrant', 'Foster', 'English Learners', 'Total')) |>
  ggplot(aes(cgr, color = reporting_cat ))+
  geom_density()+
  facet_wrap(~reporting_cat)+
  labs(x= 'College Going Rates',
    title = 'Density Plot',
    subtitle = 'College Going Rate by Demographic Group')
```

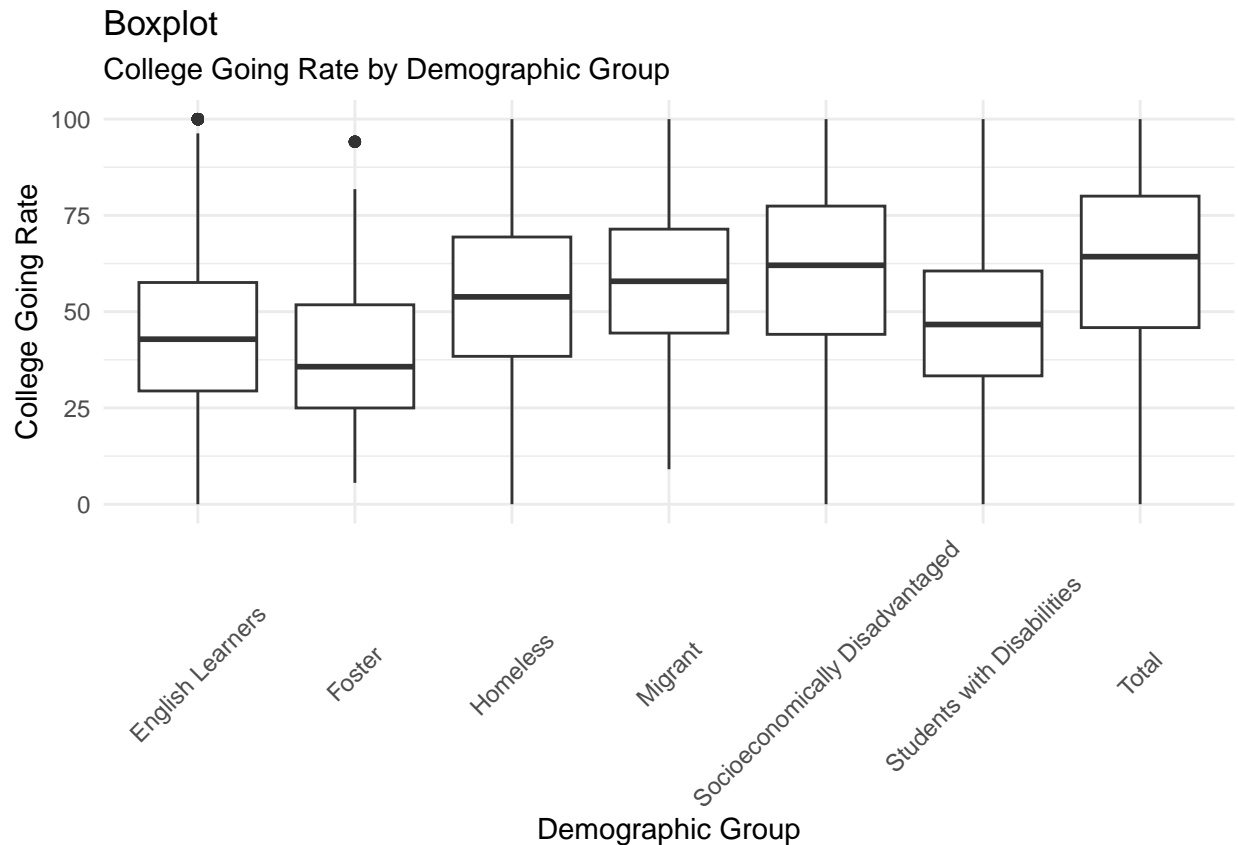
## Density Plot

### College Going Rate by Demographic Group



*# box plots grouped by additional demographic info*

```
college_gr |>
  filter(reporting_cat %in% c('Students with Disabilities', 'Total', 'Homeless',
    'Socioeconomically Disadvantaged', 'Migrant', 'Foster', 'English Learners', 'Total'))|>
  ggplot(aes(reporting_cat, cgr))+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5))+
  labs(x = 'Demographic Group',
    y = 'College Going Rate',
    title = 'Boxplot',
    subtitle = 'College Going Rate by Demographic Group')
```



```
# create df to check correlations
```

```
cgr_cor <- school_stats |>
  select(frmpl_el, unexc_abs_pct, ell_pct, cgr)
```

```
## correlations for frmpl_el, unexc_abs_pct, ell_pct, and cgr
```

```
cor(cgr_cor, use='complete.obs')
```

```
##           frmpl_el unexc_abs_pct  ell_pct      cgr
## frmpl_el      1.0000000      0.3996658  0.5304031 -0.4867759
## unexc_abs_pct  0.3996658      1.0000000  0.3407536 -0.1674253
## ell_pct        0.5304031      0.3407536  1.0000000 -0.2950742
## cgr            -0.4867759     -0.1674253 -0.2950742  1.0000000
```

```
# scatter plots
```

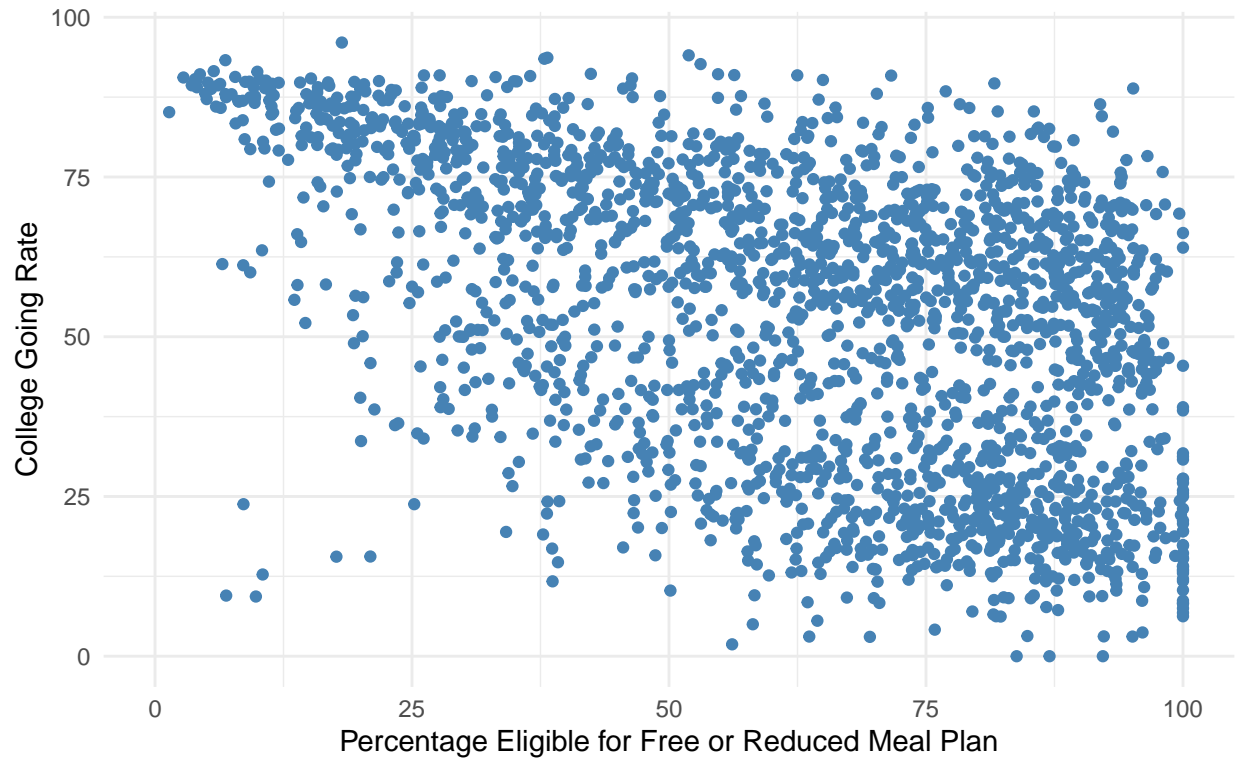
```
## college going rate and reduced lunch percent eligible
```

```
ggplot(school_stats, aes(frmpl_el, cgr))+
  geom_point(color = 'steelblue')+
  labs(x = 'Percentage Eligible for Free or Reduced Meal Plan',
       y = 'College Going Rate',
       title = 'Scatterplot',
       subtitle = 'Percentage Eligible for Free or Reduced Meal Plan')
```

```
## Warning: Removed 16142 rows containing missing values ('geom_point()').
```

## Scatterplot

Percentage Eligible for Free or Reduced Meal Plan

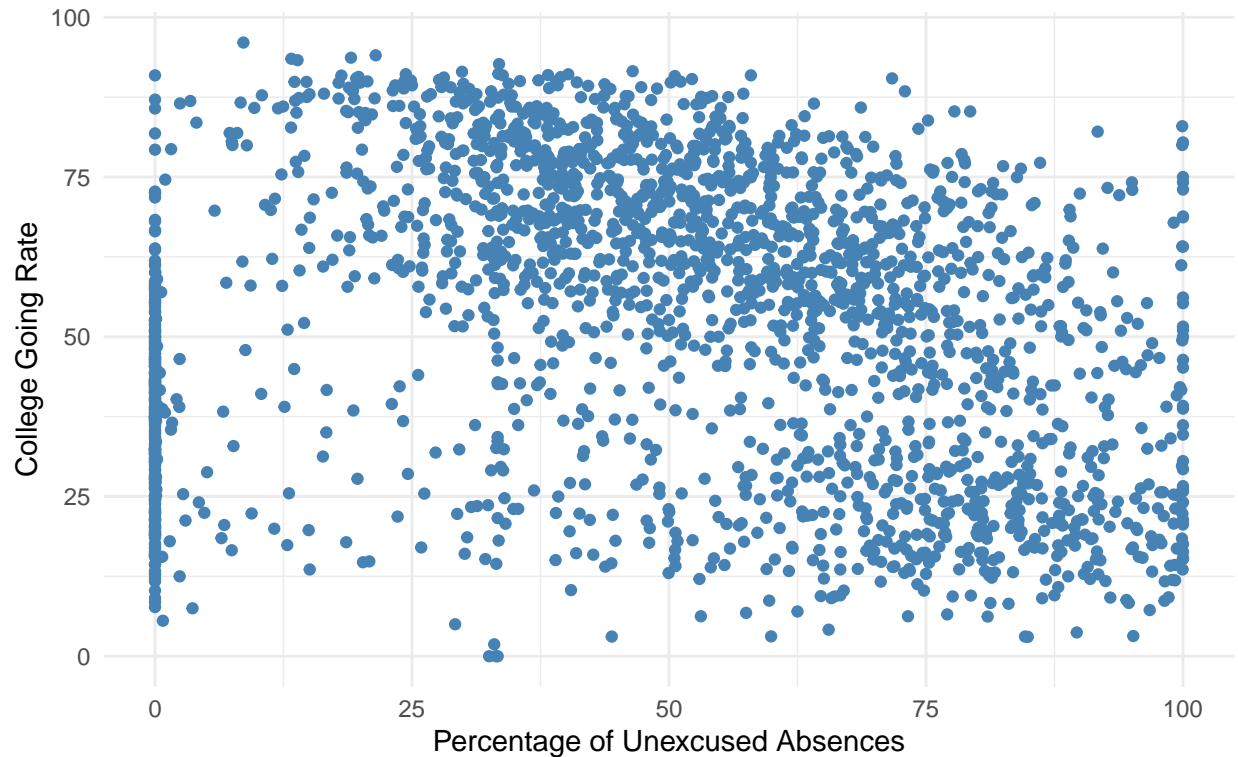


```
## college going rate and unexcused absence percent
ggplot(school_stats, aes(unexc_abs_pct, cgr))+
  geom_point(color = 'steelblue')+
  labs(x = 'Percentage of Unexcused Absences',
       y = 'College Going Rate',
       title = 'Scatterplot',
       subtitle = 'Percentage of Unexcused Absences vs College Going Rate per School')
```

```
## Warning: Removed 16145 rows containing missing values (‘geom_point()’).
```

## Scatterplot

Percentage of Unexcused Absences vs College Going Rate per School

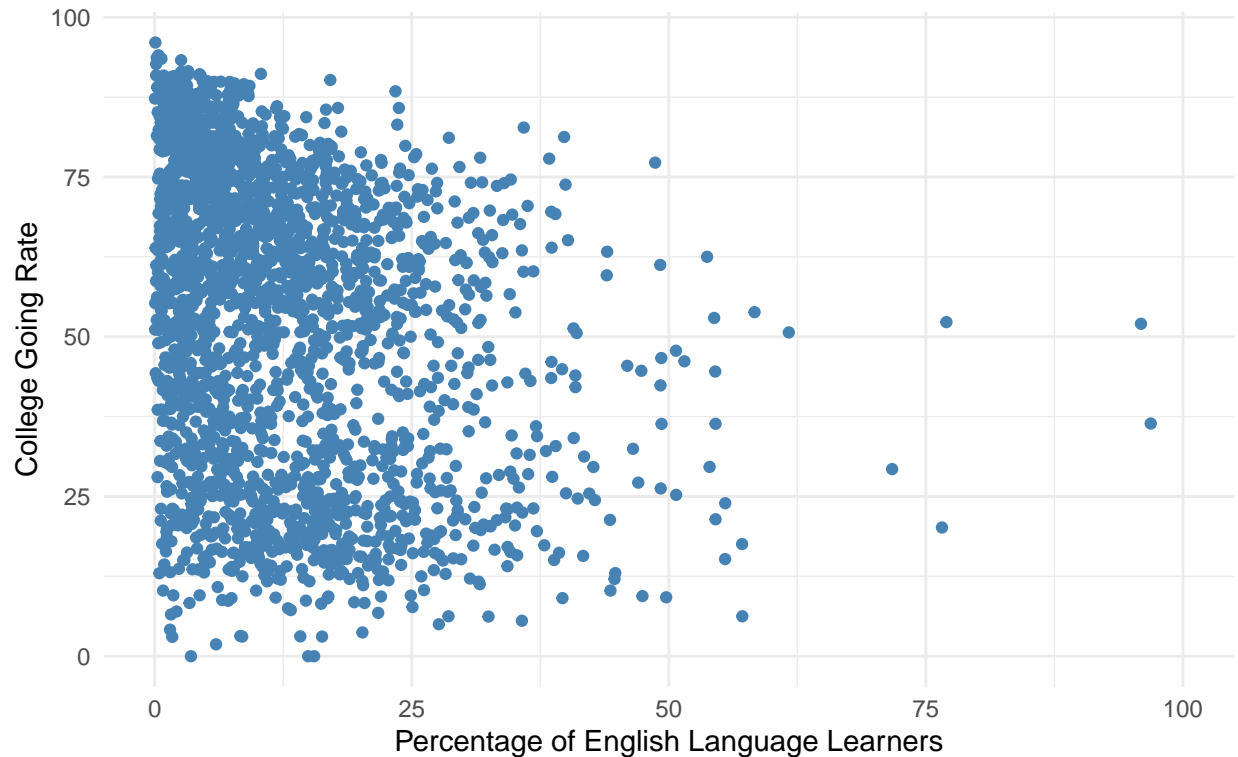


```
## college going rate and percent english language learners
ggplot(school_stats, aes(ell_pct, cgr))+
  geom_point(color = 'steelblue')+
  labs(x = 'Percentage of English Language Learners',
       y = 'College Going Rate',
       title = 'Scatterplot',
       subtitle = 'Percentage of English Language Learners vs College Going Rate per School')
```

```
## Warning: Removed 16142 rows containing missing values (‘geom_point()’).
```

## Scatterplot

Percentage of English Language Learners vs College Going Rate per School



## Part III

### Analysis and Visualizations

In the exploratory analysis, it was found that college going rates differ among different demographic groups. Additionally, a number of different factors are correlated with college going rates. This part of the analysis will utilize tidyverse functions to identify groups that have lower than average college going rates. ggplot2 will also be utilized to visualize instances where educational attainment gaps exist.

```
# lowest to highest college going rate by county

school_stats |>
  group_by(county_name) |>
  summarise(cgr = round(mean(cgr, na.rm = TRUE), 2)) |>
  arrange(cgr)
```

```
## # A tibble: 58 x 2
##   county_name  cgr
##   <chr>      <dbl>
## 1 Del Norte   27.5
## 2 Inyo       31.7
## 3 Mariposa   34.3
## 4 Lassen     35.5
## 5 Mendocino  36.0
```

```
## 6 Calaveras      37.8
## 7 Lake           38.4
## 8 Tehama         38.8
## 9 San Benito     39.6
## 10 Kern          40.1
## # i 48 more rows
```

```
# lowest to highest college going rate by county/district
```

```
school_stats |>
  group_by(county_name, district_name) |>
  summarise(cgr = round(mean(cgr, na.rm = TRUE), 2), .groups = 'drop') |>
  arrange(cgr) |>
  head(10)
```

```
## # A tibble: 10 x 3
##   county_name    district_name      cgr
##   <chr>         <chr>          <dbl>
## 1 Santa Barbara Santa Barbara County Office of Education 14.1
## 2 Lassen        Susanville Elementary                14.4
## 3 Amador         Amador County Office of Education    14.7
## 4 Mendocino      Round Valley Unified                 16.5
## 5 San Bernardino Baker Valley Unified                 16.7
## 6 Stanislaus     Stanislaus County Office of Education 16.9
## 7 Solano         Solano County Office of Education    17.4
## 8 Mono           Mono County Office of Education      17.6
## 9 San Benito     San Benito High                     19.6
## 10 San Bernardino Adelanto Elementary        19.7
```

```
# prepare data for visualization
```

```
## get location data for California and its counties
```

```
cal_counties <- map_data('county')
cal_counties <- cal_counties|>
  filter(region == 'california')
```

```
## get college going rates by county
```

```
cgr_counties <- school_stats |>
  group_by(county_name) |>
  summarise(cgr = round(mean(cgr, na.rm = TRUE), 2)) |>
  arrange(desc(cgr))
```

```
## change county_name to lowercase for merging purposes
```

```
cgr_counties$county_name <- tolower(cgr_counties$county_name)
```

```
# merge county location data with college going rate for plotting
```

```
cgr_counties <- left_join(cal_counties, cgr_counties, by = c('subregion' = 'county_name'))
```

```
# get 5 counties with the lowest college going rates
```

```
## will be used as labels for visualization
```

```
cgr_2 <- cgr_counties |>
```

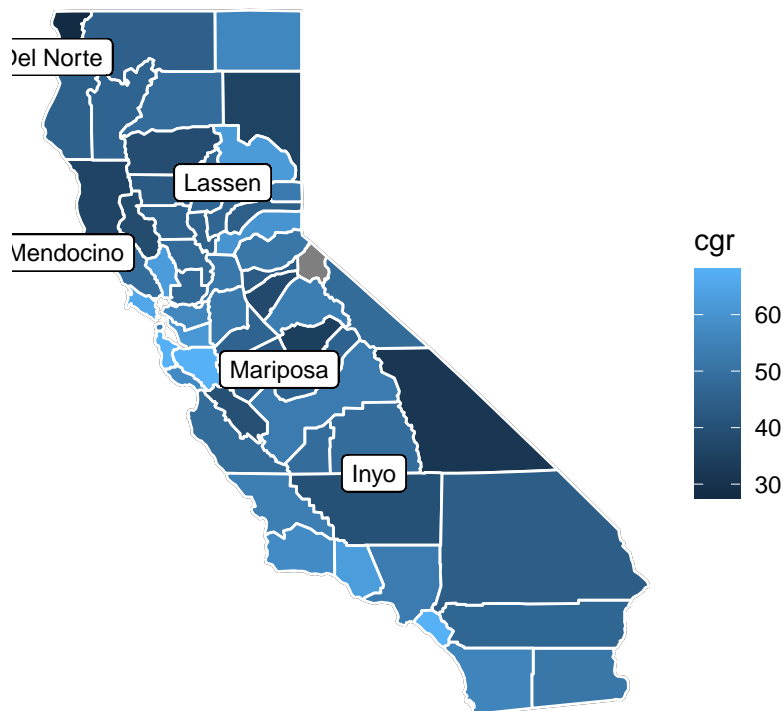


```
group_by(subregion) |>
summarise(long = min(long),
           lat = min(lat),
           cgr = mean(cgr))|>
arrange(cgr)|>
head(5)
```

*# visualize map of California with color density to reflect college going rate by county*

```
ggplot(cal_counties, mapping = aes(x=long, y=lat, group=group)) +
  coord_quickmap() +
  geom_polygon(color = 'black', fill='white') +
  geom_polygon(data = cgr_counties, aes(fill = cgr), color = 'white') +
  theme_void() +
  # label 5 counties with lowest college going rates
  geom_label(aes(x = -124.2344, y=41.39047, label = 'Del Norte'), size = 3) +
  geom_label(aes(x = -118.7799, y=35.78695, label = 'Inyo'), size = 3) +
  geom_label(aes(x = -121.3410, y=39.70024, label = 'Lassen'), size = 3) +
  geom_label(aes(x = -120.3842, y=37.18496, label = 'Mariposa'), size = 3) +
  geom_label(aes(x = -123.9995, y=38.75486, label = 'Mendocino'), size = 3) +
  labs(title = 'College Going Rate by County',
       subtitle = 'Labeled 5 Counties with Lowest College Going Rates')
```

College Going Rate by County  
Labeled 5 Counties with Lowest College Going Rates



```
# get college going rates grouped by demographic info
```

```
college_gr |>
  filter(reporting_cat %in% c('Asian', 'White', 'Filipino', 'Hispanic or Latino',
                             'African American', 'American Indian or Alaska Native',
                             'Pacific Islander')) |>
  group_by(ac_year, reporting_cat) |>
  summarise(cgr = round(mean(cgr, na.rm = TRUE), 2), .groups = 'drop') |>
  arrange(cgr)
```

```
## # A tibble: 28 x 3
##   ac_year reporting_cat      cgr
##   <chr>    <chr>         <dbl>
## 1 2020-21 American Indian or Alaska Native 38.2
## 2 2019-20 American Indian or Alaska Native 43.3
## 3 2017-18 American Indian or Alaska Native 44.8
## 4 2018-19 American Indian or Alaska Native 52.4
## 5 2019-20 Pacific Islander              54.2
## 6 2020-21 Pacific Islander              55.0
## 7 2020-21 Hispanic or Latino            56.9
## 8 2020-21 African American              57.2
## 9 2019-20 African American              58.4
## 10 2019-20 Hispanic or Latino           58.4
## # i 18 more rows
```

```
# create separate data frame for cgr/demographic group for plotting
```

```
cgr_dems <- college_gr |>
  group_by(ac_year, reporting_cat) |>
  summarise(cgr = round(mean(cgr), 2), .groups = 'drop') |>
  filter(reporting_cat %in% c('Asian', 'White', 'Filipino', 'Hispanic or Latino',
                             'African American', 'American Indian or Alaska Native',
                             'Pacific Islander'))
```

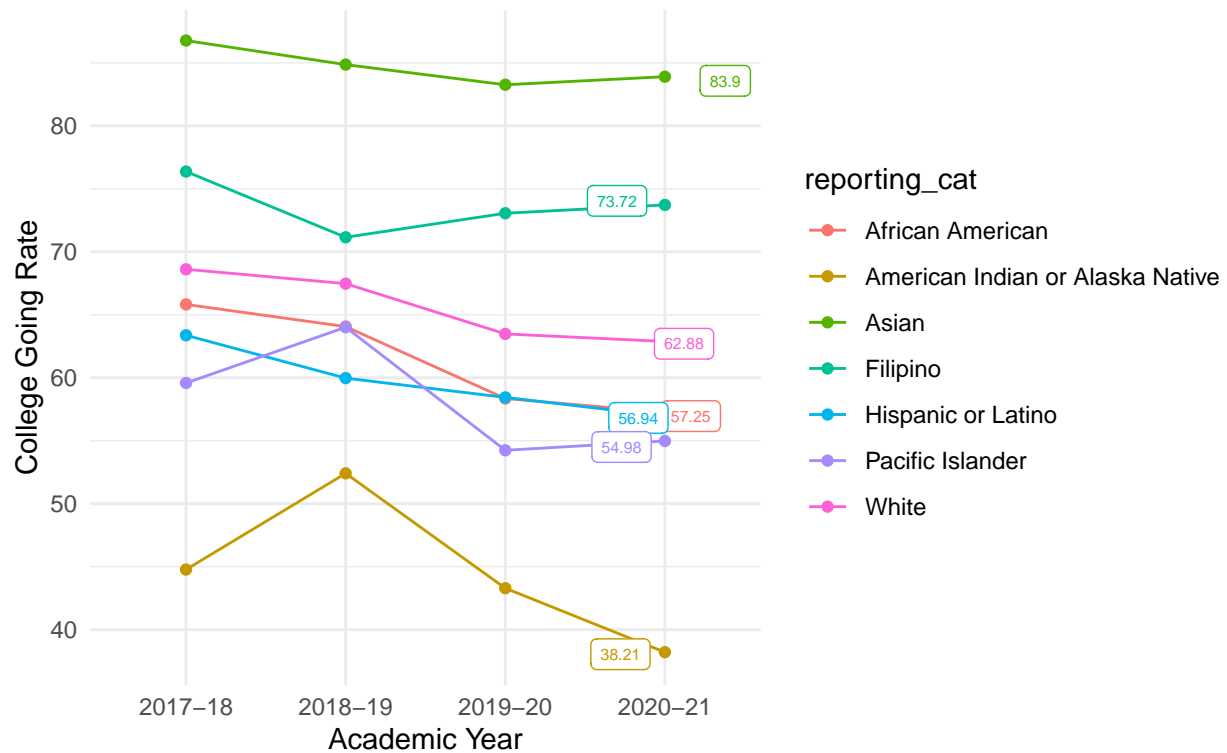
```
# create a separate object for labels
```

```
cgr_ends <- cgr_dems |>
  filter(ac_year == '2020-21')
```

```
# plot line graph
```

```
ggplot(cgr_dems, aes(ac_year, cgr, group= reporting_cat, color = reporting_cat))+
  geom_line()+
  geom_point()+
  geom_label(data = cgr_ends, aes(label = cgr), size = 2, position = position_jitter(width = 0.5, height = 0.5),
             show.legend = F)+
  labs(x = 'Academic Year',
       y = 'College Going Rate',
       title = "College Going Rate by Demographic Group",
       subtitle = '2017 - 2021')
```

## College Going Rate by Demographic Group 2017 – 2021



*# college going rate grouped by county & demographic group*

```
college_gr |>
  group_by(county_name, reporting_cat) |>
  summarise (cgr = mean(cgr), .groups = 'drop') |>
  arrange(cgr)
```

```
## # A tibble: 715 x 3
##   county_name reporting_cat    cgr
##   <chr>         <chr>         <dbl>
## 1 Del Norte    Two or More Races    0
## 2 Colusa       English Learners    9.09
## 3 Mono         Homeless            9.09
## 4 Nevada       English Learners   11.6
## 5 Inyo         English Learners   16.5
## 6 Inyo         Foster              16.7
## 7 Mono         English Learners   18.5
## 8 Glenn        English Learners   18.7
## 9 Nevada       African American   19.3
## 10 Humboldt    Homeless           20.3
## # i 705 more rows
```

*# filter df for needed demographic groups*

```
cgr_dems2 <- college_gr |>
  filter(reporting_cat %in% c('English Learners', 'Foster', 'Homeless', 'Migrant', 'Students with Disab
```

```

group_by(ac_year, reporting_cat) |>
summarise(total = sum(enr_college), .groups = 'drop')

# create df with aggregate info/ total college enrollment per year
cgr_years <- college_gr |>
  filter(reporting_cat == 'Total') |>
  group_by(ac_year) |>
  summarise(acc = sum(enr_college))

# join dfs for plotting
cgr_pcts <- left_join(cgr_dems2, cgr_years, by = join_by(ac_year))

# get percentage of demographic group enrolling in college
cgr_pcts <- cgr_pcts |>
  mutate (pct = total/acc * 100)

```

Out of all student enrolled in college, what percentage were in the following demographic groups?:

```

# plot percentage of certain demographics enrolled in college out of all students enrolled in college

ggplot(cgr_pcts, aes(ac_year, pct, fill = reporting_cat))+
  geom_bar(position = 'stack', stat = 'identity')+
  scale_y_continuous(name = 'pct', limits = c(0, 15))+
  scale_fill_brewer()+
  labs(x = 'Academic Year',
       title = 'Percentage of Students Enrolled in College by Demographic Info',
       subtitle = 'by Academic year')+
  coord_flip()

```

Percentage of Students Enrolled in College by Demographic Info  
by Academic year

