Stats 133 Final Project
Group Name: Clif Bar
Group Members: Achilleas Ghinis, Adie Fridman, Peter Butler

Abstract

The goal of our study was to see whether or not there was a significant improvement in the Academic Performance Index (API) of California students after the No Child Left Behind Act (NCLB). To do this we looked at API base scores from the year 2000 (two years before the passing of NCLB) and the year 2004 (two years after NCLB). There seems to be a significant increase in API for all races and education levels ($p < 0.01$).

Introduction

The NCLB Act was the most recent update to the Elementary and Secondary Act of 1965. It was passed by Congress in 2001 and signed by President Bush in January, 2002. The purpose of the act was to improve academic performance in schools by mandating all states to administer standardized tests (1). Technically the NCLB is optional; schools do not have to participate but then they also forgo federal funding which is allocated as part of the act. The NCLB, although passed by a bipartisan majority in Congress, was met with some criticism. Some people worried that if standardized tests were instituted, teachers would be more concerned about their students' success on the tests rather than their actual academic growth. Another major critique of the NCLB was that students of different backgrounds are being tested and compared to a standard, which means that their success is not being measured in terms of growth, but in terms of how they compare to their peers. Also, since the standardized tests were administered for reading and mathematics, many schools began to focus more on these subjects and cut down on other areas of education (namely arts) because there was less incentive to devote time and resources to such programs. In fact, since 2007, nearly 71% of schools cut funding for elective courses such as music, art, and history (2).

The question remains whether or not the act actually had a significant impact on the academic performance of students. To investigate this question, we used API data from California schools in 2000 (two years before NCLB) and in 2004 (two years after NCLB). We got our data from the California Department of Education website: (http://www.cde.ca.gov/ta/ac/ap/apidatafiles.asp).

We investigated not only how the overall API changed for California but also examined separately changes for Elementary, Middle and High schools geographically by county. We also looked at how students of various backgrounds (African American, Asian, Hispanic, White, and Socioeconomically Disadvantaged) were affected by the act.

Methods and Data Collection

We downloaded .dbf versions of the 2000 and 2004 data from the website. We imported these into R using the read.dbf function. The original data set comes with 81 variables, of these we only looked at codes/names of schools, the type of school (elementary, middle, or high) and API's both overall and by the specific groups mentioned.

To create the API data maps, we used the country.fips item in the "foreign" package in R. The FIPS (Federal Information Processing Standard) contains information on the geographical territory of all the zip codes in our data and we made the plots using the "maps" package.

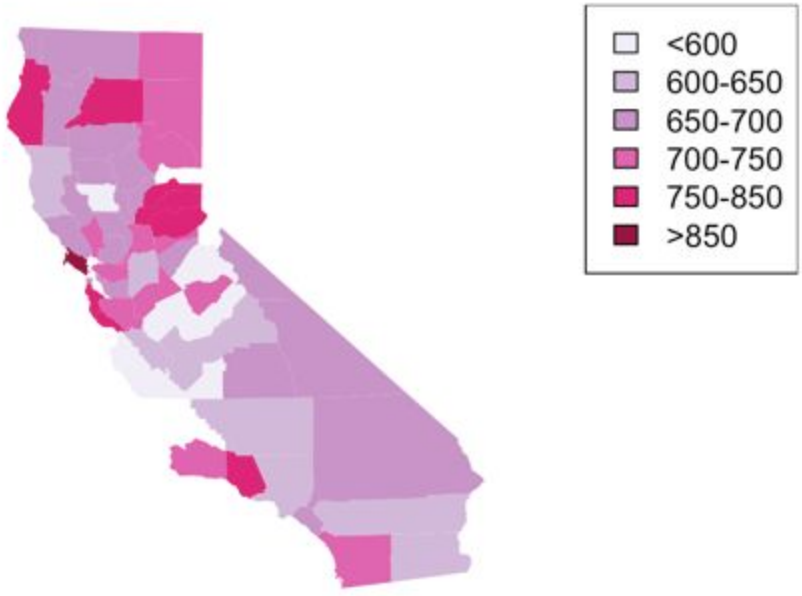
We looked at overall averages for both years and made a bar chart comparison of APIs in 2000 and 2004 both for California overall and for African American, Asian, Hispanic, White, and socioeconomically disadvantaged students. All of these data were analyzed by school type. School differences were tested using paired t-test for each of these groups.

Results and Data Visualization

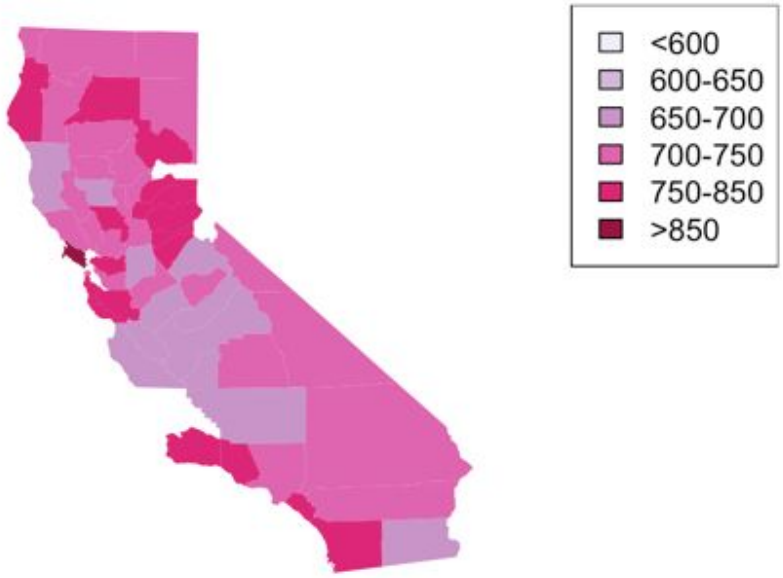
We found that for most counties there was an increase in average API at all school types; elementary, middle and high. The following maps show the 2000 API data by county, the 2004 API data by county, and the change in API (respectively) for the three school types.

Note: There are counties for which there was no data provided; these counties show up as blank on the map.

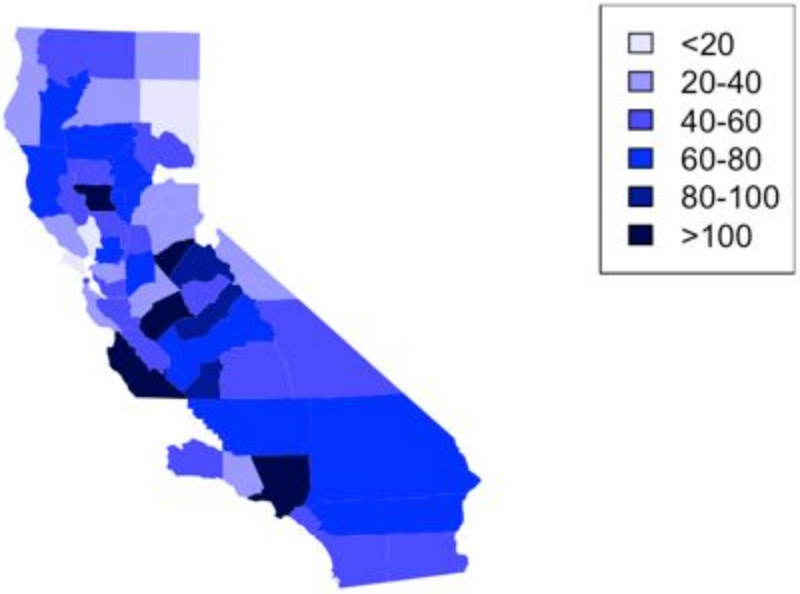
Elementary API per county, 2000



Elementary API per county, 2004

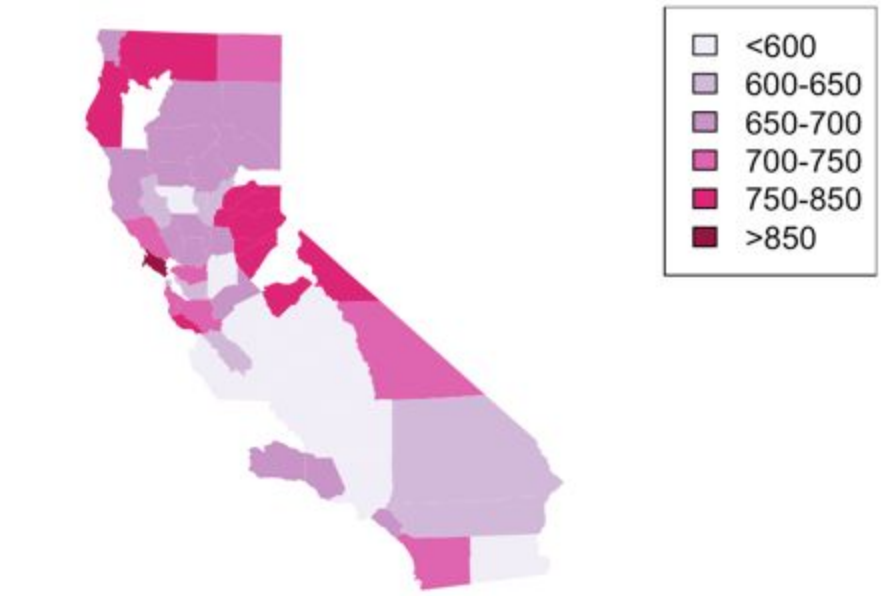


Elementary API change

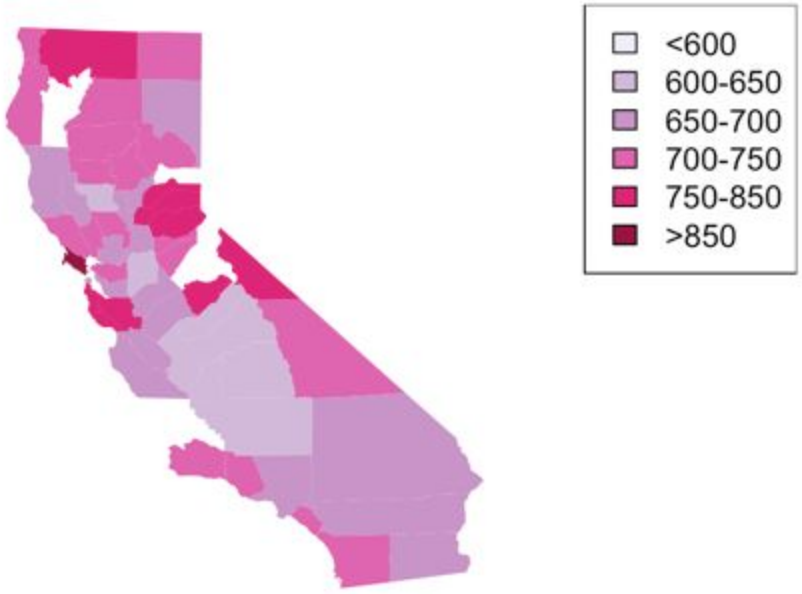


*We observe that with elementary level schools there were no negative changes in API score whereas with middle and high schools, we observe some counties with a decline in API score between the years 2000 and 2004.

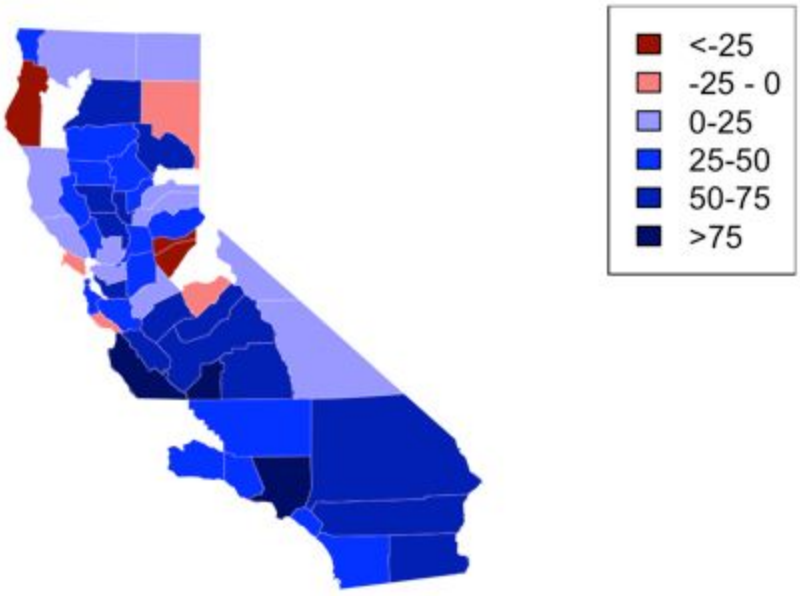
Middle School API per county, 2000



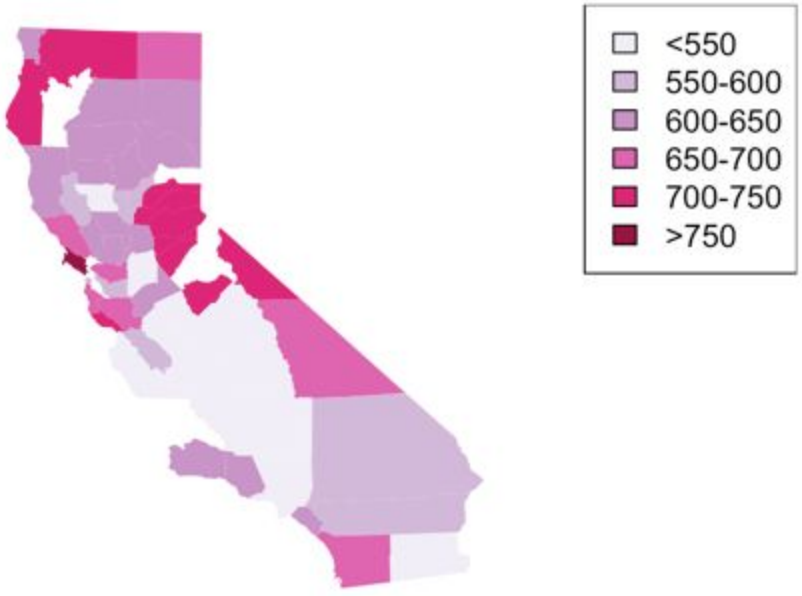
Middle School API per county, 2004



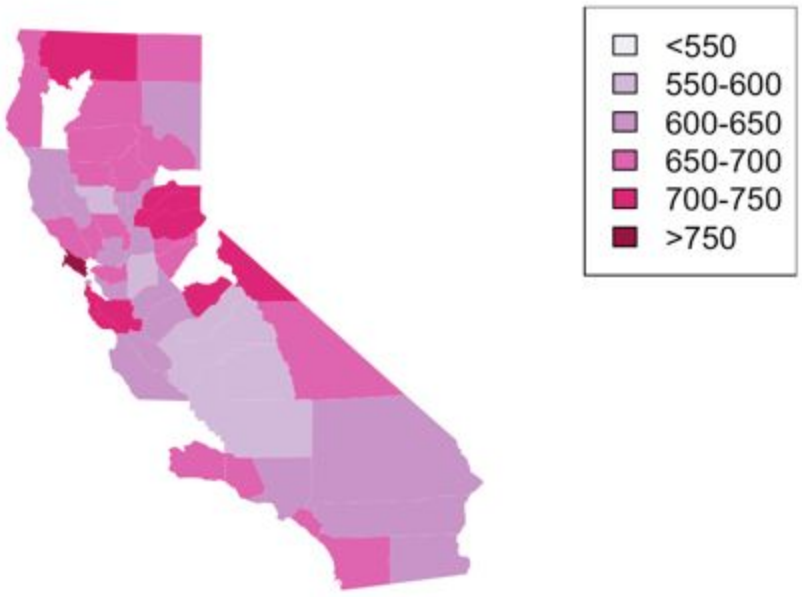
Middle School API change

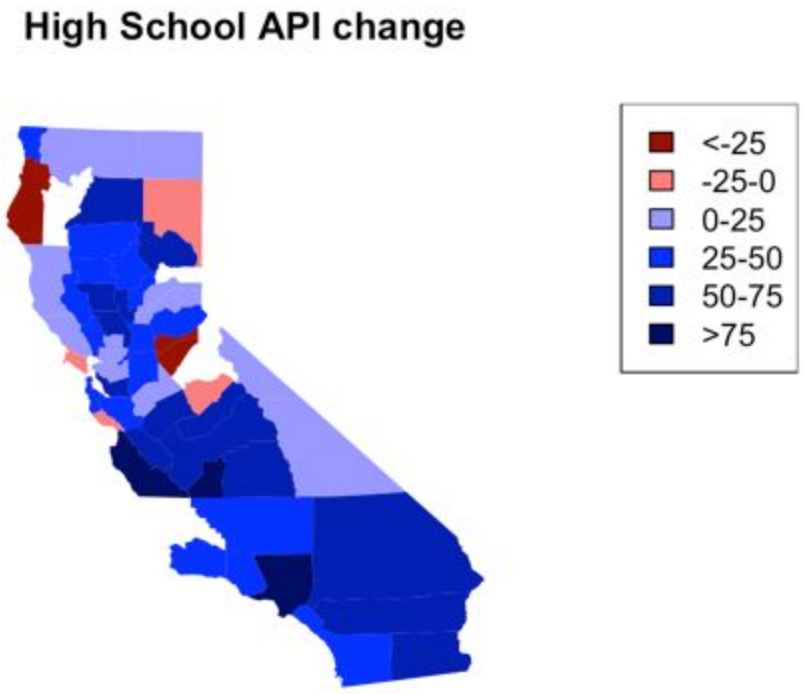


High School API per county, 2000



High School API per county, 2004

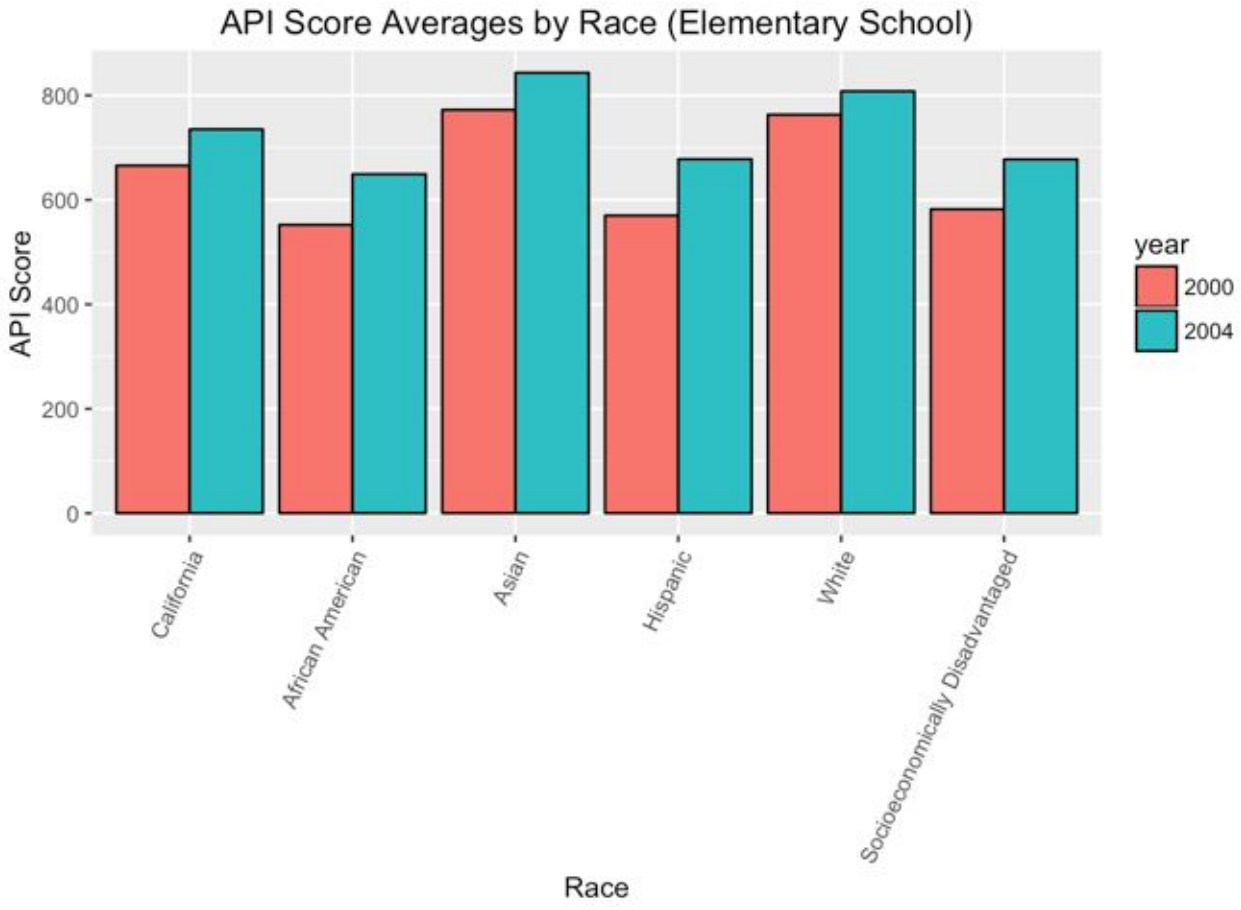


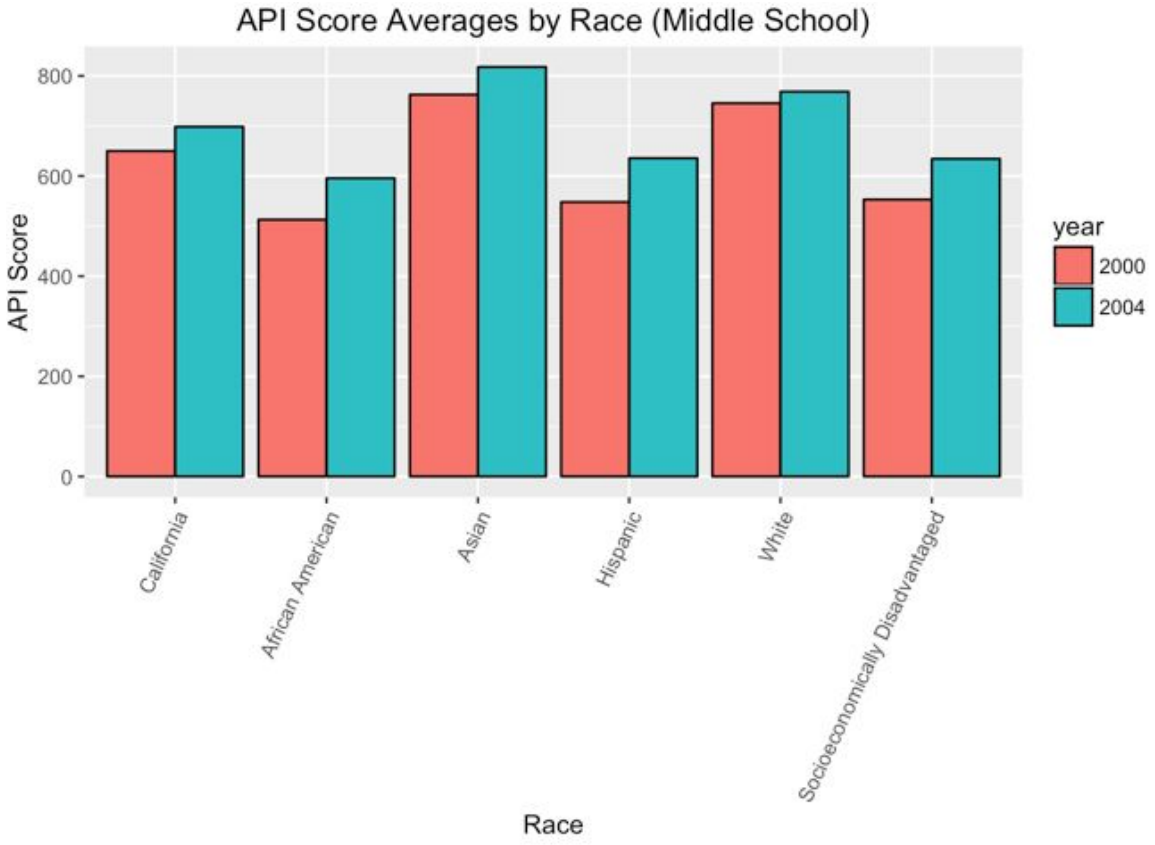


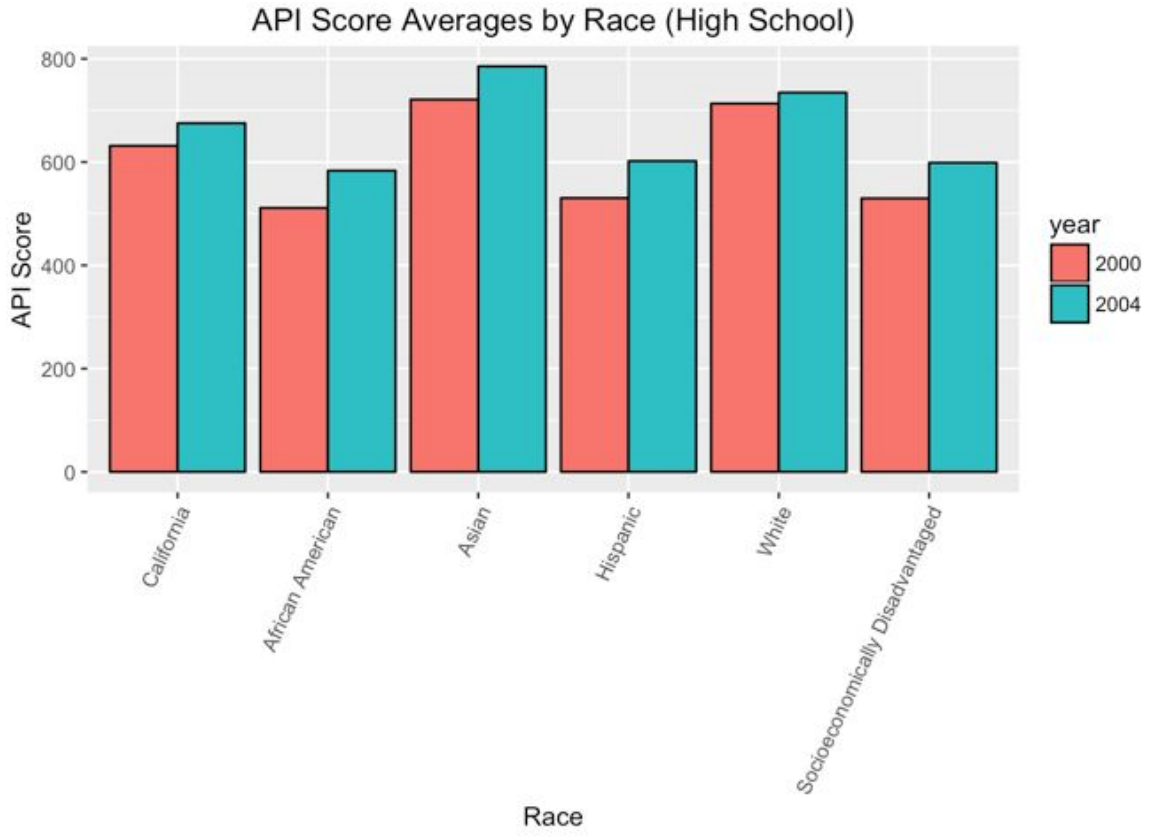
*We observe that with middle and high schools, there are seven counties which showed a decline in API score between the years 2000 and 2004. Additionally, the seven counties which showed a decline in API score with middle schools are the same counties which show a decline in API score with high schools. These counties were Humbolt (with the greatest decline in API score between the years, followed by El Dorado, Amador, Lassen, Mariposa, Santa Cruz, and Marin.

As can be seen from these graphs, API increased for most counties in California. It seems that for elementary schools, this increase was more widespread. Although middle and high schools also saw increases in API overall, these increases were less and for some counties there was even a decrease in average API for middle and high school students.

Now to see the improvement across different groups, the following graphs compare API average for California overall and for African American, Asian, Hispanic, White, and socioeconomically disadvantaged students. As can be seen here, minority races seemed to have the greatest overall improvement in API as a result of the NCLB although all racial groups saw improvement.


API Score Averages by Race (Elementary School)
API Score
800
600
400
200
0
California
African American
Asian
Hispanic
White
Socioeconomically Disadvantaged
Race
year
2000
2004


API Score Averages by Race (Middle School)
API Score
800
600
400
200
0
California
African American
Asian
Hispanic
White
Socioeconomically Disadvantaged
Race
year
2000
2004


API Score Averages by Race (High School)
API Score
800
600
400
200
0
California
African American
Asian
Hispanic
White
Socioeconomically Disadvantaged
Race
year
2000
2004

So there seems to be a consistent increase in API scores across the state and for students of every race, but to be certain that the increase we saw was significant, we tested school API differences (2000 vs 2004) using paired t-tests for each demographic group at each school type. From the calculated t statistics, it is evident that the groups that benefited the most over this period were Hispanics and socioeconomically disadvantaged students across all school types, and African-Americans in middle and high schools. It is possible that this is in part due to these students being targeted by the program. The following table summarizes the results:

	level	group	t	df	p
1	elementary	California	-26.79000	6662.90	0.000e+00
2	elementary	African American	-20.83800	1139.10	0.000e+00
3	elementary	Asian	-10.45500	1110.60	0.000e+00
4	elementary	White	-19.71800	4753.30	0.000e+00
5	elementary	Hispanic	-51.63400	4744.40	0.000e+00
6	elementary	S.D.	-49.88100	5395.40	0.000e+00
7	middle	California	-9.28350	1590.80	0.000e+00
8	middle	African American	-11.18300	368.23	0.000e+00
9	middle	Asian	-4.76150	402.93	2.685e-06
10	middle	White	-5.58870	1295.80	2.786e-08
11	middle	Hispanic	-94.74646	1207.70	0.000e+00
12	middle	S.D.	-24.04700	1359.70	0.000e+00
13	high	California	-8.41380	1365.00	0.000e+00
14	high	African American	-9.72340	361.31	0.000e+00
15	high	Asian	-6.86010	452.55	2.270e-11
16	high	White	-5.42920	1162.20	6.879e-08
17	high	Hispanic	-18.64700	1129.10	0.000e+00
18	high	S.D.	-18.14400	1203.20	0.000e+00

So it seems that this improvement in API is indeed significant. (Note: for the p-values that say 0, none were *actually* 0, but they were very low, <2.2e-16, so basically 0).

Conclusion

The data seems to indicate that there was indeed a significant improvement to API after the passage of the NCLB Act. It is interesting to see however the differences in improvement by racial/socio-economical groups.White and Asian students saw the least improvement while

minorities and socioeconomically disadvantaged student saw the most. This is possibly because the act targeted low-income schools and made special provisions for low socioeconomic status students. Also the two groups with least improvement had the highest scores at baseline, making it more difficult to improve.

Although these tests seem to indicate that NCLB achieved its goal of raising academic performance, that is actually a very tricky question. Certain critics of the act believe that the organizations which made the API tests for schools have made the tests easier so that schools can meet standards and secure funding. There also is the issue of academic dishonesty; perhaps the emphasis on standardized testing has inspired students and teachers to cheat. Another possible explanation is that other interventions and global factors (e.g., improvement in the economy freeing more money for education) are the real causes for the API improvement in California. Thus, the results of this data analysis, while consequential, are insufficient to prove that NCLB actually raised API. A more thorough analysis would be required to isolate the NCLB effect and make sure that the tests administered were indeed of a similar difficulty level as before and that there was not an increase in academic dishonesty.

There also is still the question of whether or not the NCLB was actually an improvement to school systems even if the increase seen in API was real. The fact remains that with the emphasis on standardized testing, schools only have the incentive to boost performance in reading and math to bring those subjects up to the statewide standard. This results in budget cuts to art, music, history, and gifted programs.

Appendix:

References:

1) Elementary and Secondary Education Act
(2) Pederson, P. (2007). What Is Measured Is Treasured: The Impact of the No Child Left Behind Act on Nonassessed Subjects. Clearing House, 80(6), 287–291. Retrieved from Education Research Complete database

R Code:

```
library(DataComputing)
library(foreign)

### Step 1: Loading the data

API2000 <- read.dbf("api00b.dbf", as.is = TRUE)
API2004 <- read.dbf("api04bdb.dbf", as.is= TRUE)
#head(API2000)
#head(API2004)

### Step 2: Cleaning up the data

API2000_small <-
  API2000 %>%
  select(CDS, STYPE, DNAME, CNAME, API00, AA_API, AS_API,HI_API,WH_API,
SD_API)

API2004_small <-
  API2004 %>%
  select(CDS, STYPE, DNAME, CNAME, API04, AA_API, AS_API,HI_API,WH_API,
SD_API)

# For 2000
API2000_small$API00 <- as.numeric(API2000_small$API00)
API2000_small$AA_API <- as.numeric(API2000_small$AA_API)
API2000_small$AS_API <- as.numeric(API2000_small$AS_API)
API2000_small$HI_API <- as.numeric(API2000_small$HI_API)
API2000_small$WH_API <- as.numeric(API2000_small$WH_API)
```

```
API2000_small$SD_API <- as.numeric(API2000_small$SD_API)

# For 2004
API2004_small$API04 <- as.numeric(API2004_small$API04)
API2004_small$AA_API <- as.numeric(API2004_small$AA_API)
API2004_small$AS_API <- as.numeric(API2004_small$AS_API)
API2004_small$HI_API <- as.numeric(API2004_small$HI_API)
API2004_small$WH_API <- as.numeric(API2004_small$WH_API)
API2004_small$SD_API <- as.numeric(API2004_small$SD_API)

#str(API2000_small)
#str(API2004_small)

#first change names to make it easier to read
colnames(API2000_small) <- c("CDS", "STYPE","DNAME", "CNAME", "API00",
"AA_API00", "AS_API00", "HI_API00", "WH_API00", "SD_API00")
colnames(API2004_small) <- c("CDS", "STYPE","DNAME", "CNAME", "API04",
"AA_API04", "AS_API04", "HI_API04", "WH_API04", "SD_API04")
#inner join to match the cases (in this case schools) from both tables
API_joined <- inner_join(API2000_small,API2004_small)
head(API_joined)


API_elementary <-
  API_joined %>%
  filter(STYPE == "E")
head(API_elementary)

API_middle <-
 API_joined %>%
 filter(STYPE == "M")
head(API_middle)
API_high  <-
 API_joined %>%
 filter(STYPE == "H")
head(API_high)


# B) Analysis
```

Elementary schools

```
CA_00 <- mean(API_elementary$API00, na.rm=TRUE)
AA_00 <- mean(API_elementary$AA_API00, na.rm=TRUE)
AS_00 <- mean(API_elementary$AS_API00, na.rm=TRUE)
HI_00 <- mean(API_elementary$HI_API00, na.rm=TRUE)
WH_00 <- mean(API_elementary$WH_API00, na.rm=TRUE)
SD_00 <- mean(API_elementary$SD_API00, na.rm=TRUE)

CA_04 <- mean(API_elementary$API04, na.rm=TRUE)
AA_04 <- mean(API_elementary$AA_API04, na.rm=TRUE)
AS_04 <- mean(API_elementary$AS_API04, na.rm=TRUE)
HI_04 <- mean(API_elementary$HI_API04, na.rm=TRUE)
WH_04 <- mean(API_elementary$WH_API04, na.rm=TRUE)
SD_04 <- mean(API_elementary$SD_API04, na.rm=TRUE)

elementary_averages_2000 <-
  c(CA_00,AA_00,AS_00,HI_00,WH_00,SD_00)
elementary_averages_2004 <-
  c(CA_04,AA_04,AS_04,HI_04,WH_04,SD_04)
elementary_change <- elementary_averages_2004 - elementary_averages_2000
elementary_averages <-
  data.frame(elementary_averages_2000,elementary_averages_2004, elementary_change)
rownames(elementary_averages) <- c("CA average", "AA average", "AS average", "HI
average","WH average", "SD average" )
colnames(elementary_averages) <- c("2000", "2004", "change")
elementary_averages



elementary_districts <-
  API_elementary %>%
  group_by(CNAME) %>%
  summarise(`2000` = mean(API00, na.rm=TRUE), `2004`=mean(API04, na.rm=TRUE)) %>%
  mutate(change= `2004`-`2000`)
#elementary_districts
average_elementary_change <- mean(elementary_districts$change)
average_elementary_change
```

```
require(maps)
require(ggmap)
library(stringi)
data(us.cities)

# doing some cleaning with the package to prepare a table which will allow us to create the map

pattern  <- "california\\,([[:alpha:]]+[[:space:]][[:alpha:]]+|[[:alpha:]]+)"
matches <-
  county.fips %>%
  filter(grepl(pattern,polyname)) %>%
  extractMatches(pattern,polyname,county=1) %>%
  select(fips,county)
matches$county <- stri_trans_totitle(matches$county)

colors = c("#F1EEF6", "#D4B9DA", "#C994C7", "#DF65B0", "#DD1C77", "#980043")
colors_change_elem = c("#e6e6ff", "#9999ff", "#4d4dff", "#0000ff", "#000099", "#00004d")



map_test <- inner_join(elementary_districts,matches, by=c("CNAME"="county"))
elementary_districts$colorBuckets2000 <- as.numeric(cut(elementary_districts$`2000`, c(550,
600, 650, 700, 750, 850, 900)))
elementary_districts$colorBuckets2004 <- as.numeric(cut(elementary_districts$`2004`, c(550,
600, 650, 700, 750, 850, 900)))
elementary_districts$colorBucketschange <- as.numeric(cut(elementary_districts$change, c(0,
20, 40, 60, 80, 100, 120)))

colorsmatched_elementary2000 <- elementary_districts$colorBuckets2000[match(matches$fips,
map_test$fips)]
colorsmatched_elementary2004 <- elementary_districts$colorBuckets2004[match(matches$fips,
map_test$fips)]
colorsmatched_elementarychange <-
elementary_districts$colorBucketschange[match(matches$fips, map_test$fips)]

map("county",regions="california", col = colors[colorsmatched_elementary2000], fill = TRUE,
resolution = 0, lty = 0, projection = "polyconic")
title("Elementary API per county, 2000")
```

```
leg.txt <- c("<600", "600-650", "650-700", "700-750", "750-850", ">850")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col = colors[colorsmatched_elementary2004], fill = TRUE,
resolution = 0, lty = 0, projection = "polyconic")
title("Elementary API per county, 2004")
leg.txt <- c("<600", "600-650", "650-700", "700-750", "750-850", ">850")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col =
colors_change_elem[colorsmatched_elementarychange], fill = TRUE, resolution = 0, lty = 0,
projection = "polyconic")
title("Elementary API change")
leg.txt <- c("<20", "20-40", "40-60", "60-80", "80-100", ">100")
legend("topright", leg.txt, fill = colors_change_elem)



### Analysis for middle school

CA_00M <- mean(API_middle$API00, na.rm=TRUE)
AA_00M <- mean(API_middle$AA_API00, na.rm=TRUE)
AS_00M <- mean(API_middle$AS_API00, na.rm=TRUE)
HI_00M <- mean(API_middle$HI_API00, na.rm=TRUE)
WH_00M <- mean(API_middle$WH_API00, na.rm=TRUE)
SD_00M <- mean(API_middle$SD_API00, na.rm=TRUE)

CA_04M <- mean(API_middle$API04, na.rm=TRUE)
AA_04M <- mean(API_middle$AA_API04, na.rm=TRUE)
AS_04M <- mean(API_middle$AS_API04, na.rm=TRUE)
HI_04M <- mean(API_middle$HI_API04, na.rm=TRUE)
WH_04M <- mean(API_middle$WH_API04, na.rm=TRUE)
SD_04M <- mean(API_middle$SD_API04, na.rm=TRUE)

middle_averages_2000 <-
  c(CA_00M,AA_00M,AS_00M,HI_00M,WH_00M,SD_00M)
middle_averages_2004 <-
  c(CA_04M,AA_04M,AS_04M,HI_04M,WH_04M,SD_04M)
middle_changes <- middle_averages_2004 - middle_averages_2000
```

```
middle_averages <-
  data.frame(middle_averages_2000,middle_averages_2004, middle_changes)



rownames(middle_averages) <- c("CA average", "AA average", "AS average", "HI
average","WH average", "SD average")
colnames(middle_averages) <- c("2000", "2004", "change")

middle_averages



middle_districts <-
  API_middle %>%
  group_by(CNAME) %>%
  summarise(`2000` = mean(API00, na.rm=TRUE), `2004`=mean(API04, na.rm=TRUE)) %>%
  mutate(change= `2004`-`2000`)
#middle_districts
middle_average_change <- mean(middle_districts$change)
middle_average_change



map_test_elementary <- left_join(middle_districts,matches, by=c("CNAME"="county"))

middle_districts$colorBuckets2000 <- as.numeric(cut(middle_districts$`2000`, c(546, 600, 650,
700, 750, 850, 900)))
middle_districts$colorBuckets2004 <- as.numeric(cut(middle_districts$`2004`, c(550, 600, 650,
700, 750, 850, 900)))
middle_districts$colorBucketschange <- as.numeric(cut(middle_districts$change, c(-50, -25,0,
25, 50, 75, 100)))



colors_change = c("#990000", "#ff8080", "#9999ff", "#0000ff", "#0000b3", "#000066")

colorsmatched_middle2000 <- middle_districts$colorBuckets2000[match(matches$fips,
map_test_elementary$fips)]
colorsmatched_middle2004 <- middle_districts$colorBuckets2004[match(matches$fips,
map_test_elementary$fips)]
```

```
colorsmatched_middlechange <- middle_districts$colorBucketschange[match(matches$fips,
map_test_elementary$fips)]

map("county",regions="california", col = colors[colorsmatched_middle2000], fill = TRUE,
resolution = 0, lty = 0, projection = "polyconic")
title("Middle School API per county, 2000")
leg.txt <- c("<600", "600-650", "650-700", "700-750", "750-850", ">850")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col = colors[colorsmatched_middle2004], fill = TRUE,
resolution = 0, lty = 0, projection = "polyconic")
title("Middle School API per county, 2004")
leg.txt <- c("<600", "600-650", "650-700", "700-750", "750-850", ">850")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col = colors_change[colorsmatched_middlechange], fill =
TRUE, resolution = 0, lty = 0, projection = "polyconic")
map("county", regions="california", col = "white", fill = FALSE, add = TRUE, lty = 1, lwd =
0.2,
          projection = "polyconic")
title("Middle School API change")
leg.txt <- c("<-25", "-25 - 0", "0-25", "25-50", "50-75", ">75")
legend("topright", leg.txt, fill = colors_change)

### High School

CA_00H <- mean(API_high$API00, na.rm=TRUE)
AA_00H <- mean(API_high$AA_API00, na.rm=TRUE)
AS_00H <- mean(API_high$AS_API00, na.rm=TRUE)
HI_00H <- mean(API_high$HI_API00, na.rm=TRUE)
WH_00H <- mean(API_high$WH_API00, na.rm=TRUE)
SD_00H <- mean(API_high$SD_API00, na.rm=TRUE)

CA_04H <- mean(API_high$API04, na.rm=TRUE)
AA_04H <- mean(API_high$AA_API04, na.rm=TRUE)
AS_04H <- mean(API_high$AS_API04, na.rm=TRUE)
HI_04H <- mean(API_high$HI_API04, na.rm=TRUE)
WH_04H <- mean(API_high$WH_API04, na.rm=TRUE)
SD_04H <- mean(API_high$SD_API04, na.rm=TRUE)
```

```
high_averages_2000 <-
  c(CA_00H,AA_00H,AS_00H,HI_00H,WH_00H,SD_00H)
high_averages_2004 <-
  c(CA_04H,AA_04H,AS_04H,HI_04H,WH_04H,SD_04H)
high_changes <- high_averages_2004 - high_averages_2000

high_averages <-
  data.frame(high_averages_2000,high_averages_2004, high_changes)


rownames(high_averages) <- c("CA average", "AA average", "AS average", "HI average","WH
average", "SD average")
colnames(high_averages) <- c("2000", "2004", "change")

high_averages

high_districts <-
  API_high %>%
  group_by(CNAME) %>%
  summarise(`2000` = mean(API00, na.rm=TRUE), `2004`=mean(API04, na.rm=TRUE)) %>%
  mutate(change= `2004`-`2000`)
#high_districts
high_average_change <- mean(high_districts$change)
high_average_change

high_districts %>%
  ggplot(aes(CNAME,`2000`)) + geom_bar(stat="identity")

##Making the maps

map_test_high <- left_join(high_districts,matches, by=c("CNAME"="county"))


high_districts$colorBuckets2000 <- as.numeric(cut(high_districts$`2000`, c(500, 550, 600, 650,
700, 750, 800)))
high_districts$colorBuckets2004 <- as.numeric(cut(high_districts$`2004`, c(500, 550, 600, 650,
700, 750, 800)))
```

```
high_districts$colorBucketschange <- as.numeric(cut(high_districts$change, c(-50, -25,0, 25, 50, 75, 111)))

colorsmatched_high2000 <- high_districts$colorBuckets2000[match(matches$fips, map_test_high$fips)]
colorsmatched_high2004 <- high_districts$colorBuckets2004[match(matches$fips, map_test_high$fips)]
colorsmatched_highchange <- high_districts$colorBucketschange[match(matches$fips, map_test_high$fips)]


map("county",regions="california", col = colors[colorsmatched_middle2000], fill = TRUE, resolution = 0, lty = 0, projection = "polyconic")
title("High School API per county, 2000")
leg.txt <- c("<550", "550-600", "600-650", "650-700", "700-750", ">750")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col = colors[colorsmatched_middle2004], fill = TRUE, resolution = 0, lty = 0, projection = "polyconic")
title("High School API per county, 2004")
leg.txt <- c("<550", "550-600", "600-650", "650-700", "700-750", ">750")
legend("topright", leg.txt, fill = colors)

map("county",regions="california", col = colors_change[colorsmatched_middlechange], fill = TRUE, resolution = 0, lty = 0, projection = "polyconic")
title("High School API change")
leg.txt <- c("<-25", "-25-0", "0-25", "25-50", "50-75", ">75")
legend("topright", leg.txt, fill = colors_change)

### Basic plots

elementary_averages %>%
 mutate(categories = row.names(elementary_averages),
        categories = sub("CA average", "California", categories),
        categories = sub("AA average", "African American", categories),
        categories = sub("AS average", "Asian", categories),
        categories = sub("HI average", "Hispanic", categories),
        categories = sub("WH average", "White", categories),
        categories = sub("SD average", "Socioeconomically Disadvantaged", categories),
```

```
        categories = factor(categories,  c("California", "African American", "Asian", "Hispanic",
"White", "Socioeconomically Disadvantaged"))) %>%
select(`2000`, `2004`, categories) %>%
 gather("year", "API scores", `2000`:`2004`) %>%
 ggplot(aes(x=categories, y=`API scores`, fill=year)) +
   geom_bar(stat="identity", position=position_dodge(), colour="black") + theme(axis.text.x =
element_text(angle =65, hjust = 1)) + labs(title= "API Score Averages by Race (Elementary
School)", x= "Race", y="API Score")

middle_averages %>%
 mutate(categories = row.names(elementary_averages),
        categories = sub("CA average", "California", categories),
        categories = sub("AA average", "African American", categories),
        categories = sub("AS average", "Asian", categories),
        categories = sub("HI average", "Hispanic", categories),
        categories = sub("WH average", "White", categories),
        categories = sub("SD average", "Socioeconomically Disadvantaged", categories),
        categories = factor(categories,  c("California", "African American", "Asian", "Hispanic",
"White", "Socioeconomically Disadvantaged"))) %>%
select(`2000`, `2004`, categories) %>%
 gather("year", "API scores", `2000`:`2004`) %>%
 ggplot(aes(x=categories, y=`API scores`, fill=year)) +
   geom_bar(stat="identity", position=position_dodge(), colour="black") + theme(axis.text.x =
element_text(angle =65, hjust = 1)) + labs(title= "API Score Averages by Race (Middle
School)", x= "Race", y="API Score")


high_averages %>%
 mutate(categories = row.names(elementary_averages),
        categories = sub("CA average", "California", categories),
        categories = sub("AA average", "African American", categories),
        categories = sub("AS average", "Asian", categories),
        categories = sub("HI average", "Hispanic", categories),
        categories = sub("WH average", "White", categories),
        categories = sub("SD average", "Socioeconomically Disadvantaged", categories),
        categories = factor(categories,  c("California", "African American", "Asian", "Hispanic",
"White", "Socioeconomically Disadvantaged"))) %>%
select(`2000`, `2004`, categories) %>%
 gather("year", "API scores", `2000`:`2004`) %>%
```

```
 ggplot(aes(x=categories, y=`API scores`, fill=year)) +
   geom_bar(stat="identity", position=position_dodge(), colour="black") + theme(axis.text.x =
element_text(angle =65, hjust = 1)) + labs(title= "API Score Averages by Race (High School)",
x= "Race", y="API Score")

### testing for significance

# for districts

t.test(high_districts$`2000`,high_districts$`2004`)
t.test(middle_districts$`2000`, middle_districts$`2004`)
t.test(elementary_districts$`2000`,elementary_districts$`2004`)

# for race in elementary

t.test(API_elementary$API00,API_elementary$API04)
t.test(API_elementary$AA_API00,API_elementary$AA_API04)
t.test(API_elementary$AS_API00,API_elementary$AS_API04)
t.test(API_elementary$WH_API00,API_elementary$WH_API04)
t.test(API_elementary$HI_API00,API_elementary$HI_API04)
t.test(API_elementary$SD_API00,API_elementary$SD_API04)

# for race in middle
t.test(API_middle$API00,API_middle$API04)
t.test(API_middle$AA_API00,API_middle$AA_API04)
t.test(API_middle$AS_API00,API_middle$AS_API04)
t.test(API_middle$WH_API00,API_middle$WH_API04)
t.test(API_middle$HI_API00,API_middle$HI_API04)
t.test(API_middle$SD_API00,API_middle$SD_API04)

# for race in high

t.test(API_high$API00,API_high$API04)
t.test(API_high$AA_API00,API_high$AA_API04)
t.test(API_high$AS_API00,API_high$AS_API04)
t.test(API_high$WH_API00,API_high$WH_API04)
t.test(API_high$HI_API00,API_high$HI_API04)
t.test(API_high$SD_API00,API_high$SD_API04)
```