

Math130 Exploratory Data Analysis Project – High School and Beyond

Kaylee Henderson

2025-06-27

Introduction:

The High School and Beyond is a study to shows educational, vocational, and personal development of young people, starting with their elementary or high school years and following them over time as they take on adult responsibilities. For this project I wanted to look how race is compared to their science scores. I compared their performance in science from their test scores and see how each race has been affected.

```
hsb2 <- read.delim("hsb2.txt")  
dim(hsb2)
```

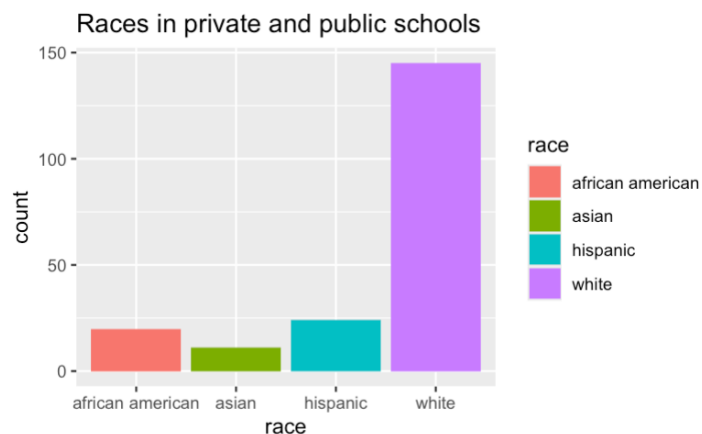
```
[1] 200 11
```

Grouping of Race

african american	asian
20	11
hispanic	white
24	145

This is a table of the race amount of African American, Asian, Hispanic, and white people in the study, with white people taking the majority.

```
ggplot(hsb2, aes(x=race, fill=race)) + geom_bar() + xlab("race") +  
ggtitle("Races in private schools")
```



This is more of a visual representation of the table above to show how white is dominating any other race

Science Scores

```
table(hsb2$science)
```

```
26 29 31 33 34 35 36 39 40 42 44 45 46
 1  1  2  2  5  1  4 13  2 12 11  1  1
47 48 49 50 51 53 54 55 56 57 58 59 61
11  2  2 21  2 15  3 18  2  1 19  1 14
63 64 65 66 67 69 72 74
12  1  1  9  1  6  2  1
```

The table shows the science scores test, that range from 26 to 74

Summary

Min.	1st Qu.	Median	Mean
26.00	44.00	53.00	51.85
3rd Qu.	Max.		
58.00	74.00		

The students average for the science test was 51.85% with a median of 53%.

Bivariate Exploration

```
      26  29  31
african american 0.000 0.050 0.100
asian            0.000 0.000 0.000
hispanic        0.042 0.000 0.000
white           0.000 0.000 0.000

      33  34  35
african american 0.050 0.100 0.050
asian            0.000 0.000 0.000
hispanic         0.000 0.000 0.000
white            0.007 0.014 0.000

      36  39  40
african american 0.000 0.100 0.000
asian            0.000 0.000 0.000
hispanic        0.042 0.167 0.042
white           0.021 0.048 0.007

      42  44  45
african american 0.050 0.100 0.000
asian            0.182 0.000 0.000
hispanic        0.083 0.208 0.042
white           0.048 0.028 0.000

      46  47  48
african american 0.000 0.000 0.000
asian            0.000 0.091 0.000
hispanic        0.000 0.083 0.000
white           0.007 0.055 0.014

      49  50  51
african american 0.050 0.150 0.000
asian            0.000 0.091 0.000
hispanic        0.000 0.083 0.000
white           0.007 0.103 0.014

      53  54  55
african american 0.100 0.000 0.050
asian            0.000 0.091 0.091
hispanic        0.042 0.000 0.042
white           0.083 0.014 0.103

      56  57  58
african american 0.000 0.000 0.000
asian            0.000 0.091 0.091
hispanic        0.042 0.000 0.000
white           0.007 0.000 0.124

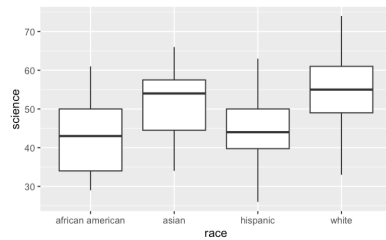
      59  61  63
african american 0.000 0.050 0.000
asian            0.000 0.091 0.000
hispanic        0.000 0.042 0.042
white           0.007 0.076 0.076

      64  65  66
african american 0.000 0.000 0.000
asian            0.000 0.000 0.091
hispanic        0.000 0.000 0.000
white           0.007 0.007 0.055

      67  69  72
african american 0.000 0.000 0.000
asian            0.000 0.000 0.000
hispanic        0.000 0.000 0.000
white           0.007 0.041 0.014

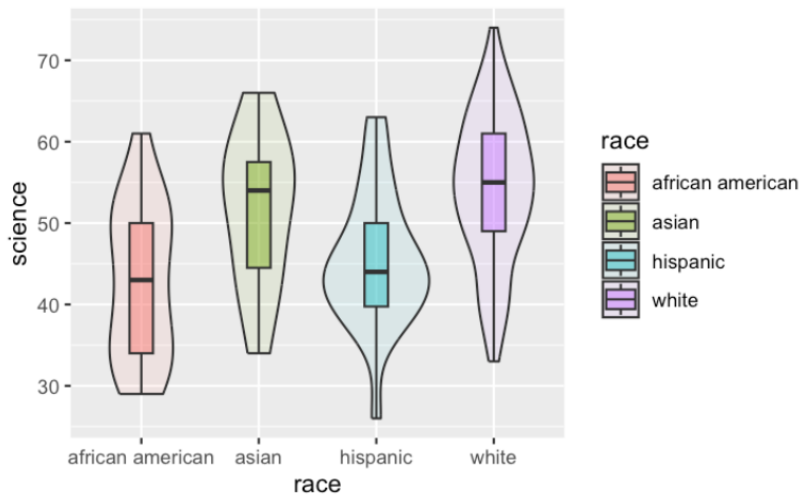
      74
african american 0.000
asian            0.000
hispanic        0.000
white           0.007
```

```
ggplot(hsb2, aes(x=race, y=science)) + geom_boxplot()
```



In the box plot it showed that African Americans and Hispanic tend to do worse in science tests, and white people and Asians have done better with white people doing the best.

```
ggplot(hsb2, aes(x=race, y=science, fill=race)) +  
geom_violin(alpha=.1) + geom_boxplot(alpha=.5, width=.2)
```



White people tend to do the best, with Asian people coming right behind them. One of the lowest being Hispanic people, while African American people take the lowest scores for the science tests.

Conclusion:

The High School and Beyond study's data shows that racial groups average science test scores range significantly from one another. African American and Hispanic students typically scored worse on average in science, whereas white and Asian students typically performed better. My prior hypothesis assumed that Asian people would have the best science test scores. While it is very close white people do have the best scores out of them all.