

# Final Project

Kimberly Lor

9/20/2023

## Introduction

For my final project, I decided to choose High School and Beyond as my data analysis. The High School and Beyond was the second study directed as a part of the National Longitudinal Studies Program. The program was set up to study the educational, vocational, and personal development of young people who was beginning their elementary or high school years. This data set covers 200 observations and 11 variables. I will be mainly focusing on these three data variables: gender, race, and program. This analysis aims to compare which gender and race got into which programs. Are there more females in each program more than males? Which program had more African Americans?

---

## Read in the “High School and Beyond” data set

```
hsb2 <- read.table("/Users/kimberlylor/Desktop/math130/Data/hsb2.csv", header=TRUE, sep="\t")
```

```
library('dplyr')
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library('ggplot2')
```

```
library('gridExtra')
```

```
##
```

```
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   combine
```

```
str(hsb2)
```

```
## 'data.frame':   200 obs. of  11 variables:  
## $ id      : int  70 121 86 141 172 113 50 11 84 48 ...  
## $ gender  : chr  "male" "female" "male" "male" ...  
## $ race    : chr  "white" "white" "white" "white" ...  
## $ ses     : chr  "low" "middle" "high" "high" ...  
## $ schtyp  : chr  "public" "public" "public" "public" ...  
## $ prog    : chr  "general" "vocational" "general" "vocational" ...  
## $ read    : int  57 68 44 63 47 44 50 34 63 57 ...  
## $ write   : int  52 59 33 44 52 52 59 46 57 55 ...  
## $ math    : int  41 53 54 47 57 51 42 45 54 52 ...  
## $ science: int  47 63 58 53 53 63 53 39 58 50 ...  
## $ socst   : int  57 61 31 56 61 61 61 36 51 51 ...
```

## Table of Gender

```
table(hsb2$gender)
```

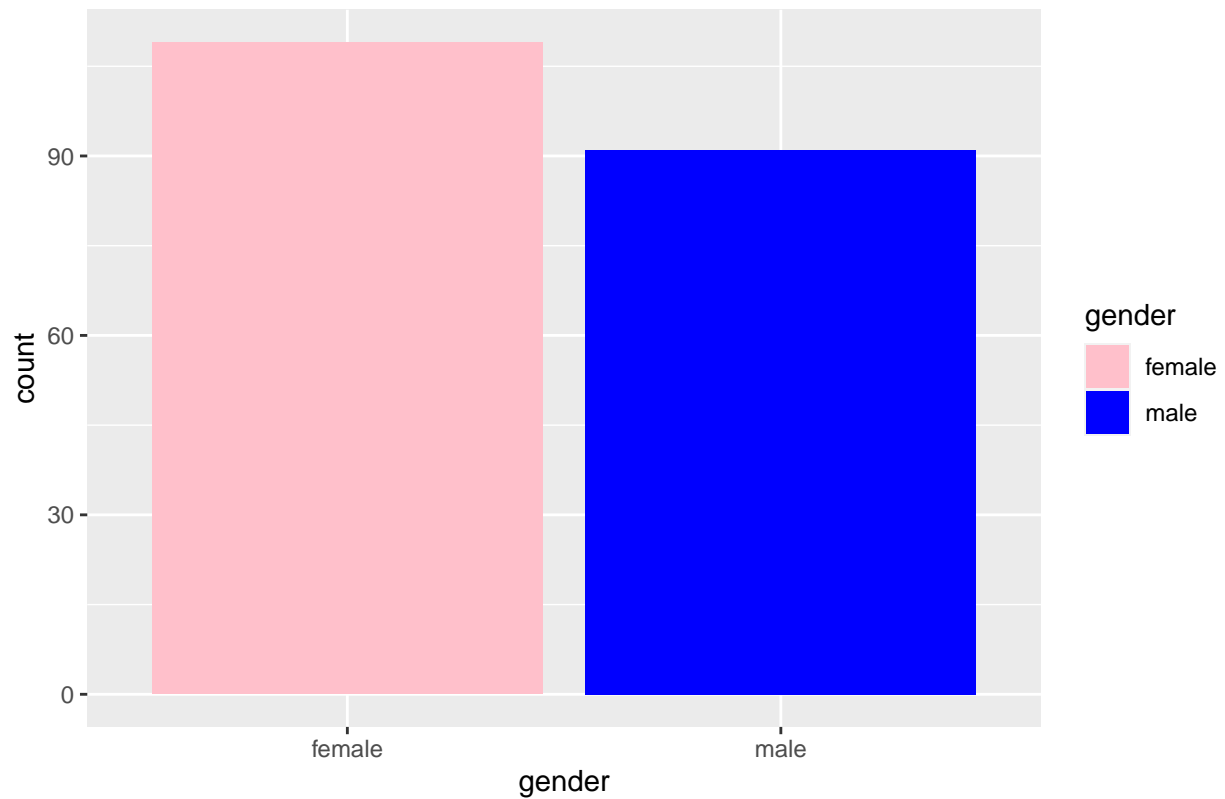
```
##  
## female  male  
##   109    91
```

```
prop.table(table(hsb2$gender))*100
```

```
##  
## female  male  
##   54.5   45.5
```

```
ggplot(hsb2, aes(x=gender, fill=gender)) + geom_bar() + scale_fill_manual(values=c("pink", "blue")) + g
```

## Gender Comparison



## Table of Race

```
table(hsb2$race)
```

```
##  
## african american      asian      hispanic      white  
##           20           11           24           145
```

```
prop.table(table(hsb2$race))*100
```

```
##  
## african american      asian      hispanic      white  
##           10.0           5.5           12.0           72.5
```

```
ggplot(hsb2, aes(race))+geom_bar()+ggtitle("Number of Individuals in Race Group")
```

Number of Individuals in Race Group

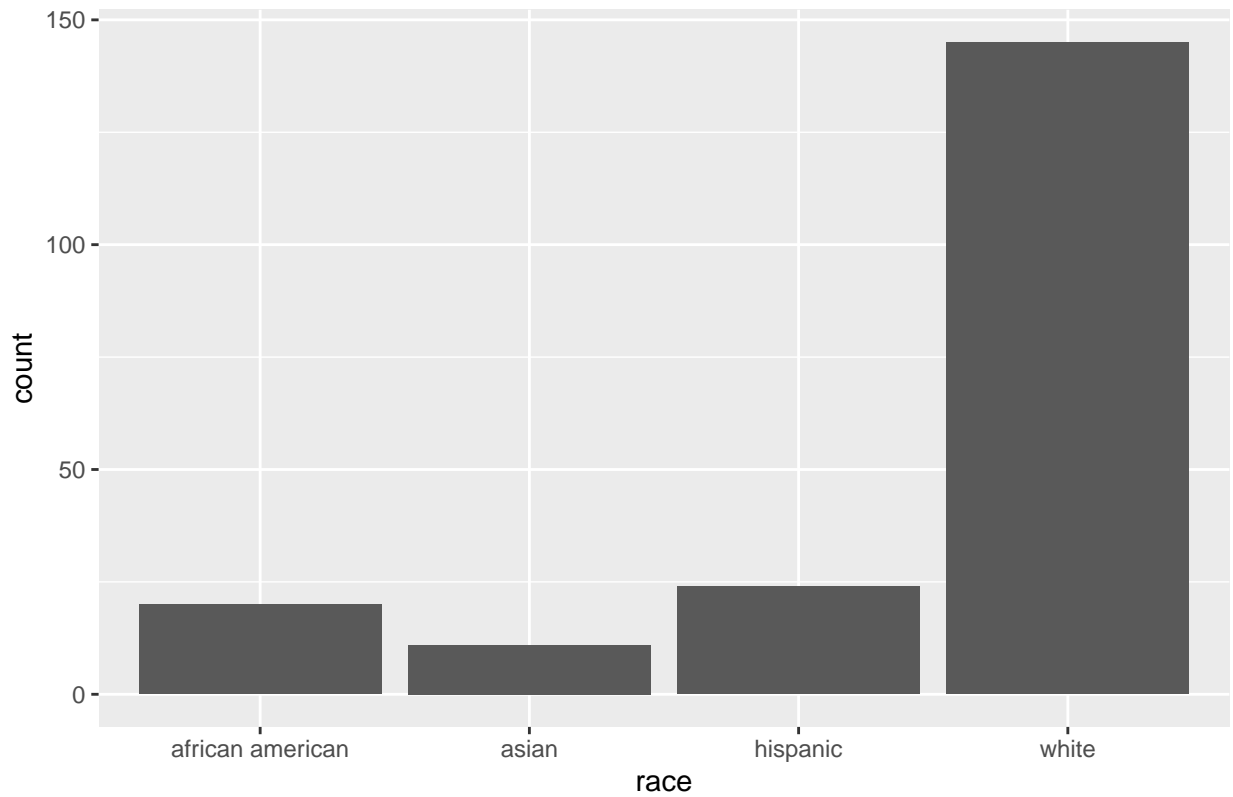


Table of Programs

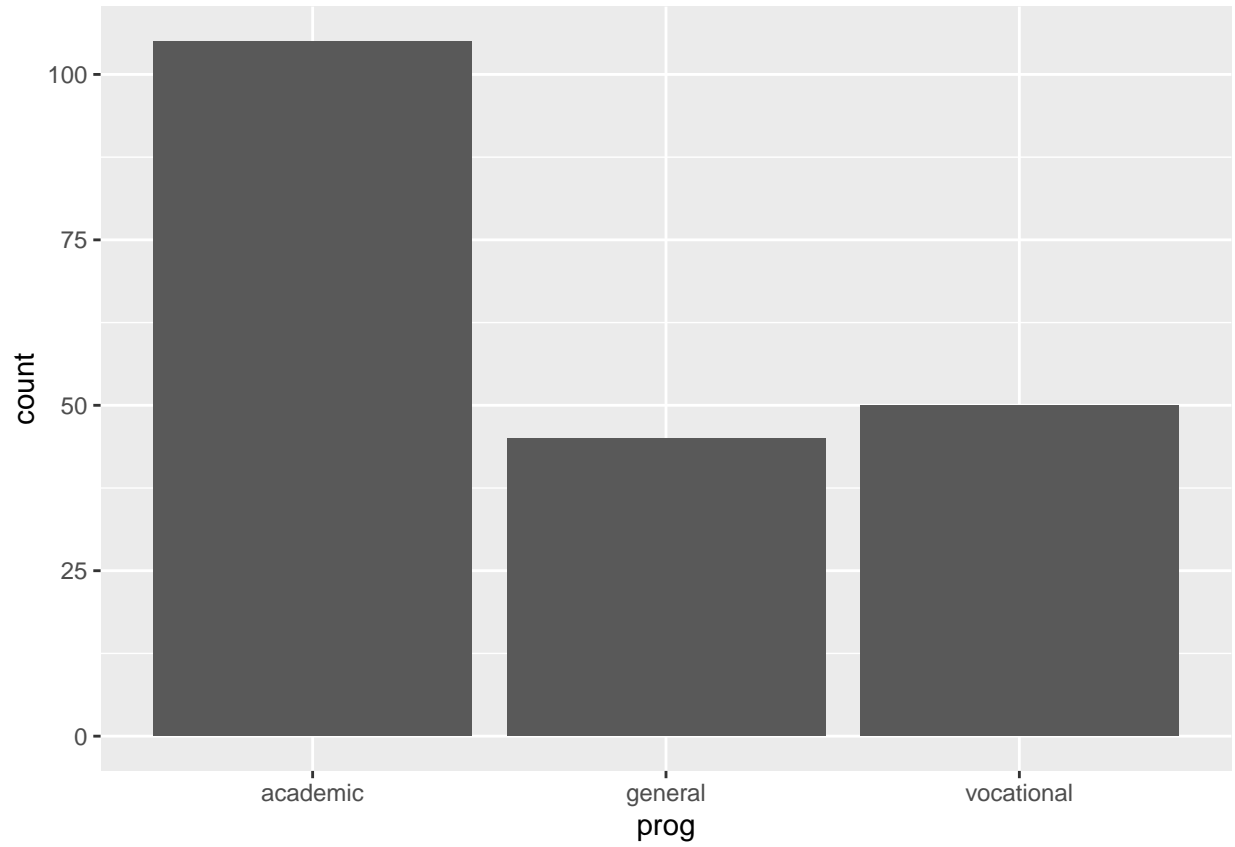
```
table(hsb2$prog)
```

```
##  
##  academic    general vocational  
##      105         45         50
```

```
prop.table(table(hsb2$prog))*100
```

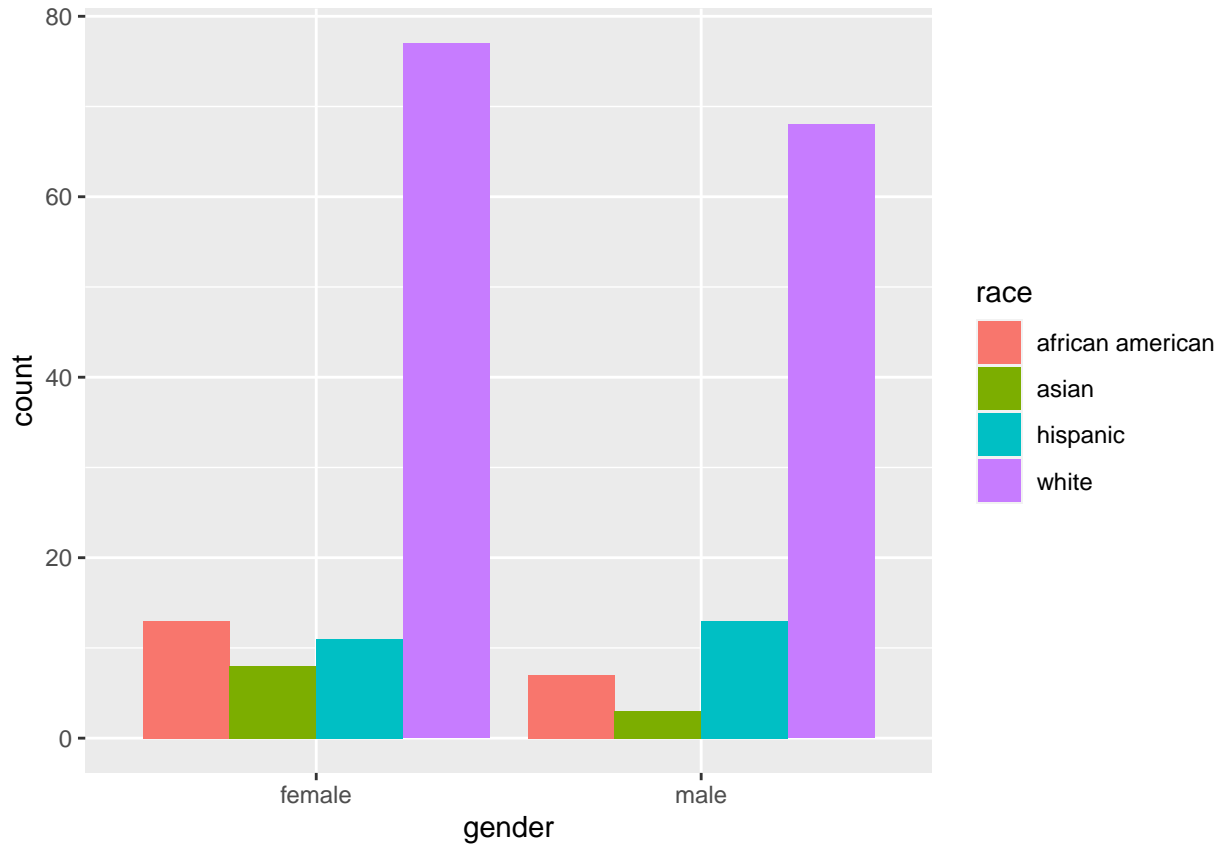
```
##  
##  academic    general vocational  
##      52.5     22.5     25.0
```

```
ggplot(hsb2, aes(prog))+geom_bar()
```

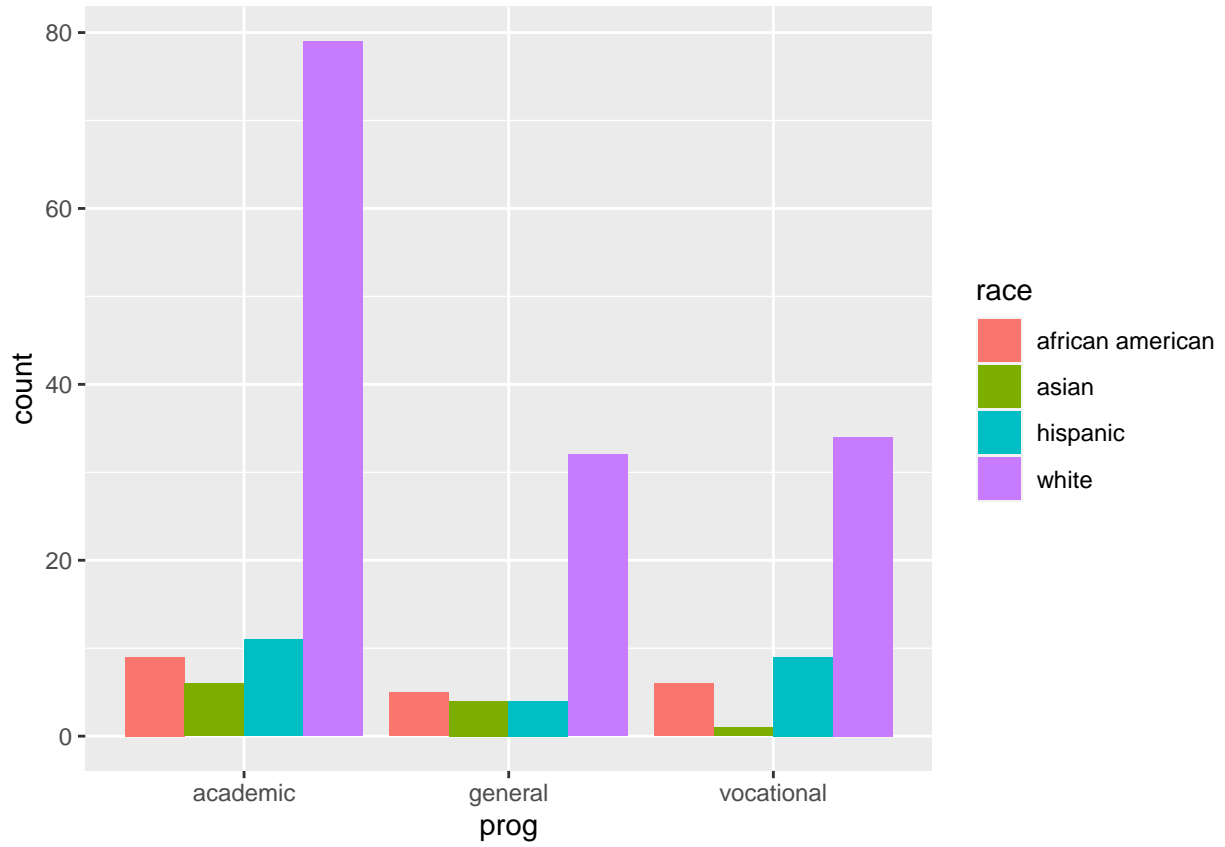


## Plots

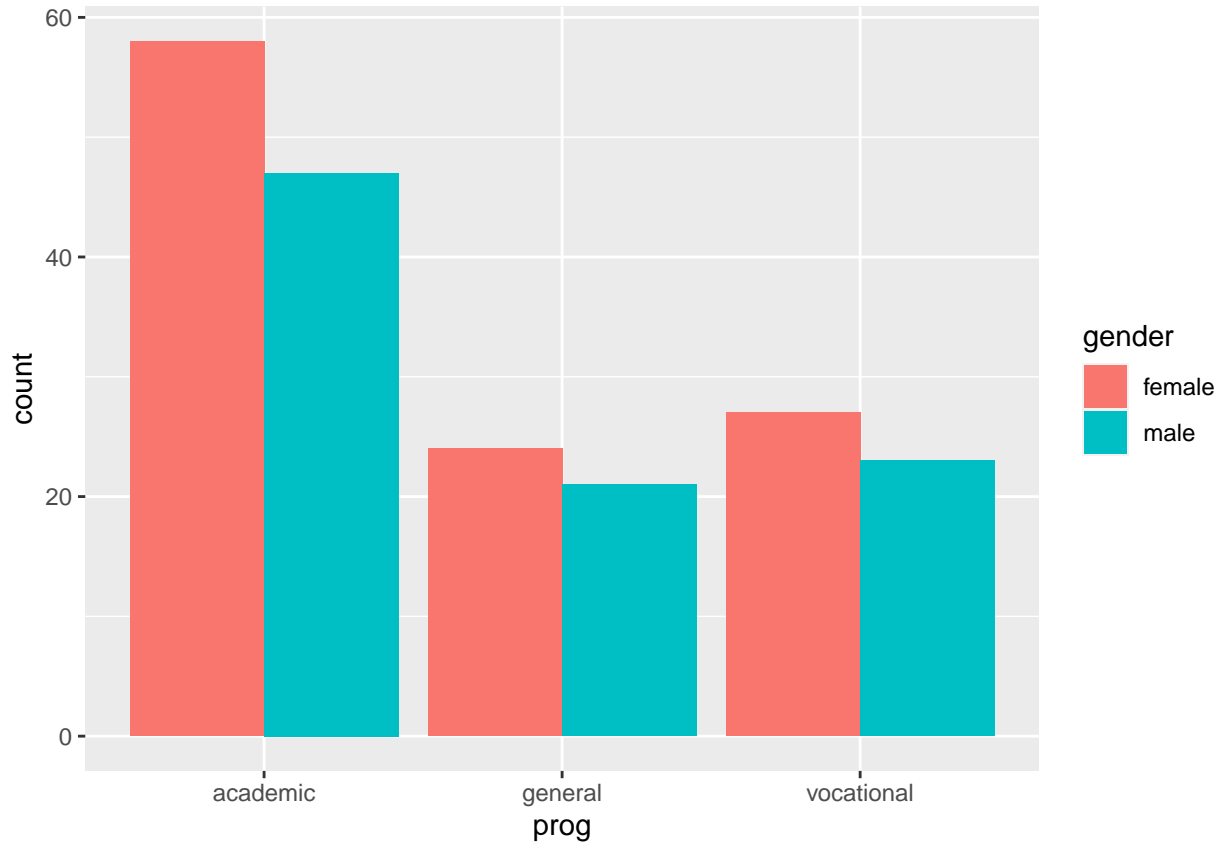
```
ggplot(hsb2, aes(x=gender, fill=race))+geom_bar(position="dodge")
```



```
ggplot(hsb2, aes(x=prog, fill=race))+geom_bar(position="dodge")
```



```
ggplot(hsb2, aes(x=prog, fill=gender))+geom_bar(position="dodge")
```



### Conclusion

In conclusion, the data shows that there were more female whites in each programs than any other. The academic program had more african american than any other programs. However I think that my questions were pretty on point and that my prediction was correct. We can see that there were more whites in each programs than any other ones.