High School and Beyond

samantha

9/21/2022

```
hsb2<-read.delim("/Users/samanthareyes/Desktop/math130/data/hsb2.txt", sep="\t")
library(ggplot2)
```

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

```
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(sjPlot)
```

1. introduction/Description: High School and Beyond

I chose this data set High School and Beyond for my Exploratory data analysis. This data set is to study the educational, vocational, and personal development of young adults from elementary to high school years. Throughout time as they become adults and take on roles along with responsibilities. There are 200 observations with 11 variables in this date set. I will strictly scale my data set down by only sticking to Race, Gender, And school type(schtyp).

```
pol <- select(hsb2,race, gender, schtyp)
```

2. Univariate Exploration

Tables of Race Summary Statistics

```
table(hsb2$race)
```
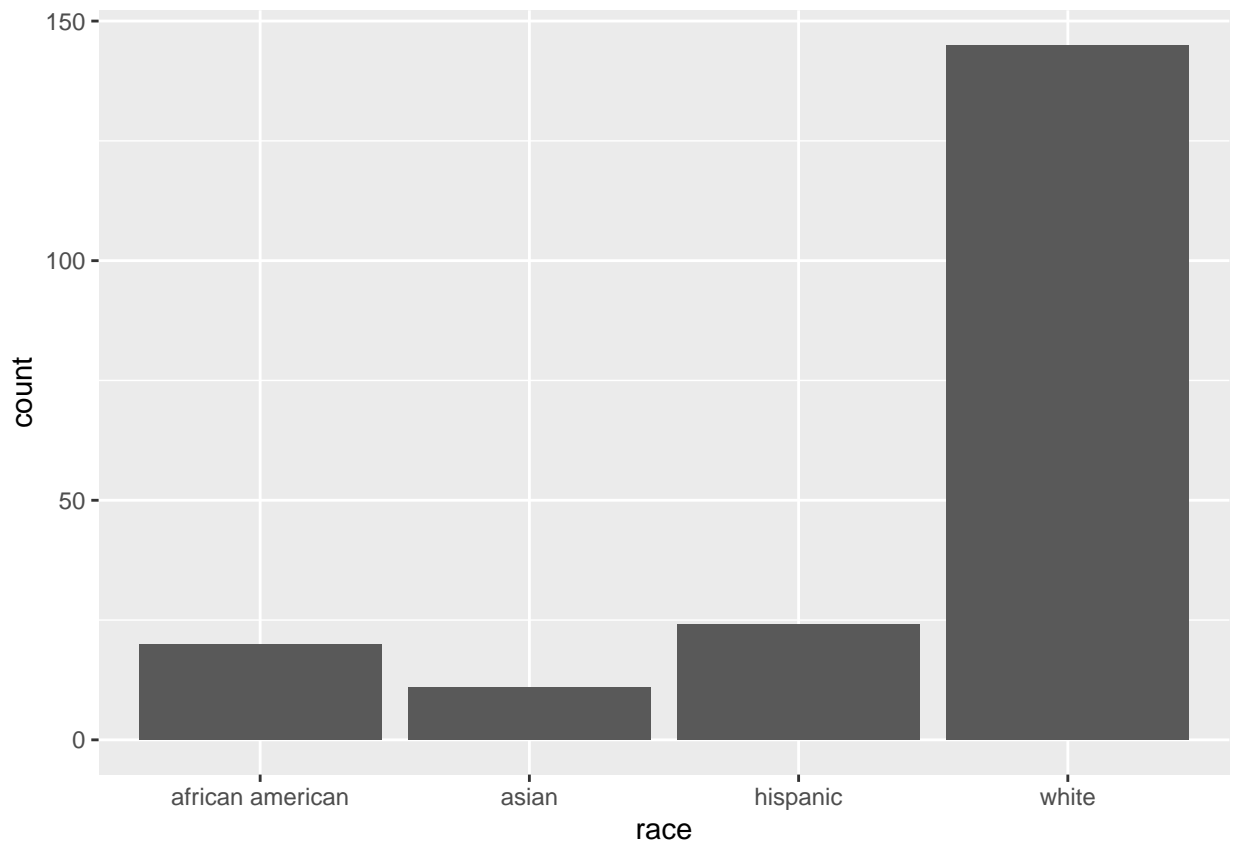
```
##
## african american          asian       hispanic          white
##              20             11             24            145
```

Above is the summary statistics for the Race variable from the 200 participants. This shows us the numbers of each race within the 200 being observed. We can see that 20 are African American, !! are Asian, 24 are Hispanic and 145 are White.

Bar graph for the the Race variable

```
ggplot(pol, aes(race))+geom_bar()
```


Now that we have the variable race visualized on our x-axis with the corresponding counts on the bar graph chart, we can better see how our data could be skewed toward one race.
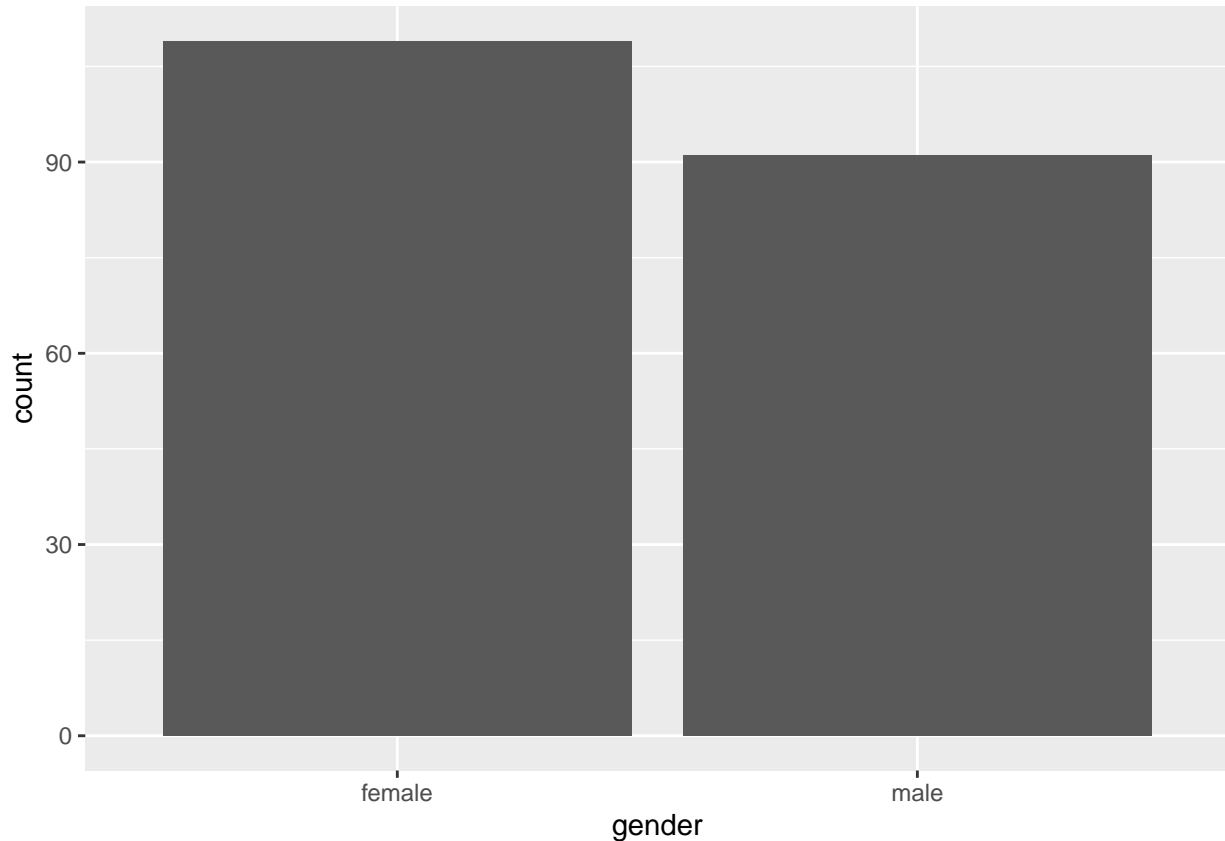
Gender Table

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

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

In this data illustrated in the table above we can see that from our overall data that 109 people are female and 91 are male. This gives us a visual layout of how many people from the observation are female vs. male.

Gender graph

```
ggplot(pol, aes(x=gender)) + geom_bar()
```


Above we can see the difference in genders through a bar plot. We can see from a scale of 100% that 54.5% of the observed are female, and 45.5% of the observed are male.
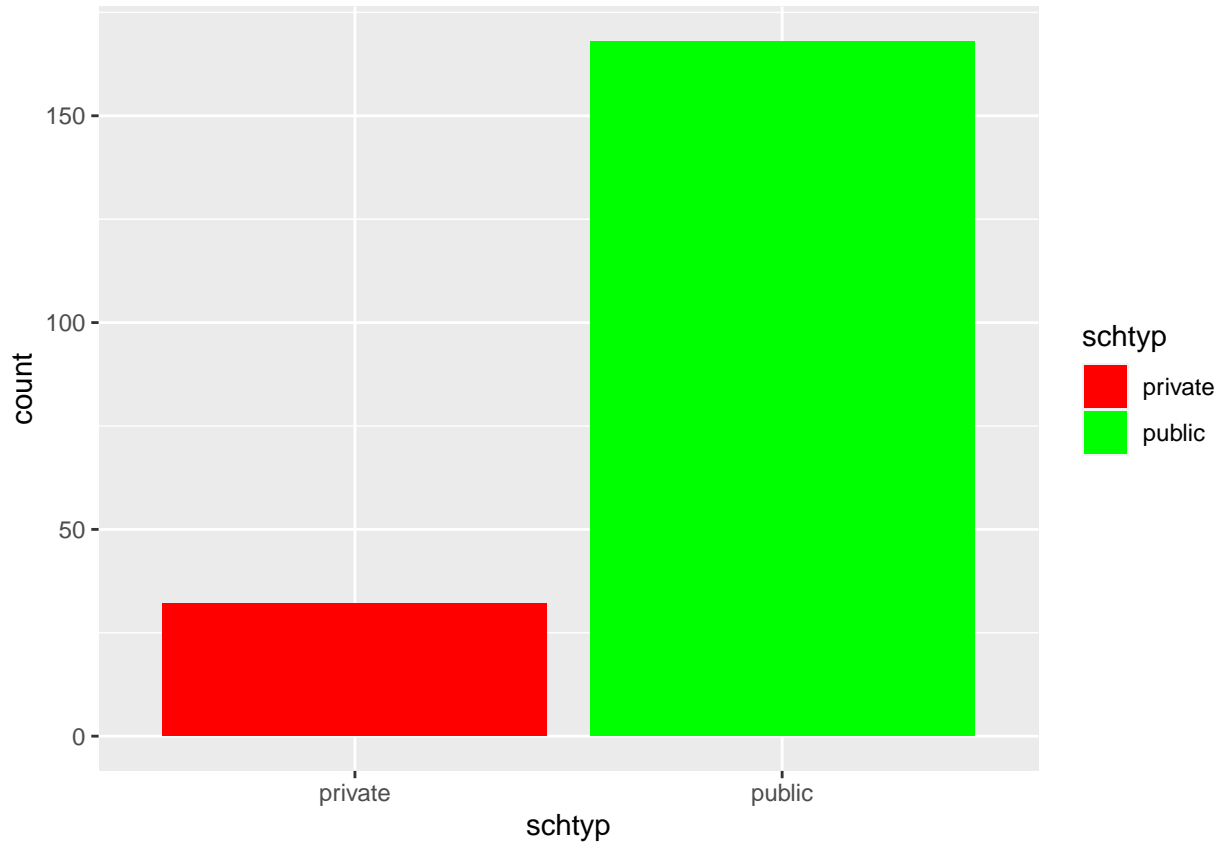
School type table

```
pol$schtyp %>% table()
```

```
## .
## private  public
##      32     168
```

This table shows us our data difference in public vs private schools. We can tell from the data given that 32 are private and 168 are public. This means that from all the observed this is how many attended private vs public schools.

School Type Bar Graph

```
ggplot(pol, aes(x=schtyp, fill=schtyp)) + geom_bar() +
    scale_fill_manual(values=c("red", "green"))
```

This shows the amount of public schools versus the amount of private schools taken from our HS and Beyond data set. A better visual representation.

school type gender graph

```
ggplot(pol, aes(y=race, x=gender, col=schtyp)) + geom_boxplot() +
  scale_color_discrete(name="Size of number")
```
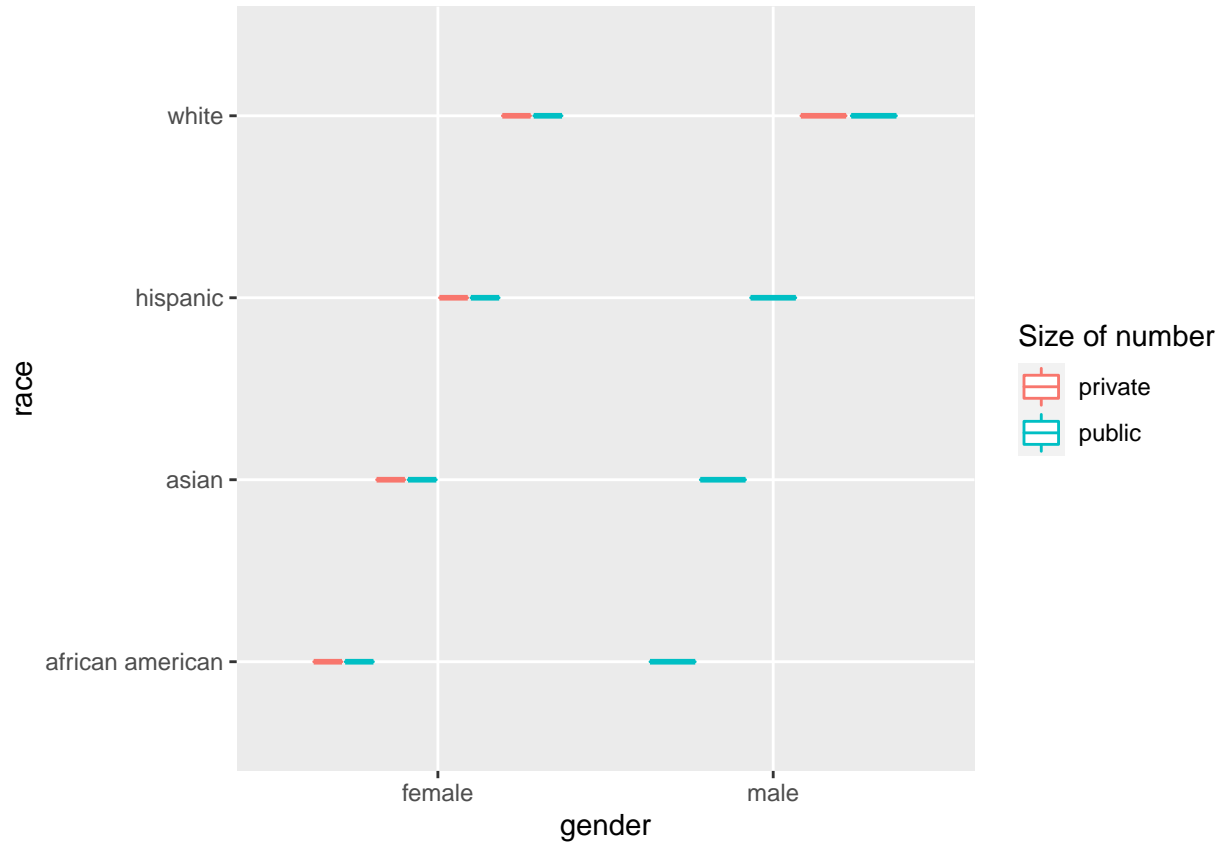
In the box plot above we can see the gender to race ratio from our data set. This shows the race of females and the race in males from our data set in general. Along with if the females or males in each race attended private and or public schools.

3 Bivariate Comparison of variables:

Grouped Table

```
table(pol$gender, pol$race, pol$schtyp)
```

```
## , ,  = private
##
##
##          african american asian hispanic white
##   female                2     1        2    13
##   male                  0     0        0    14
##
## , ,  = public
##
##
##          african american asian hispanic white
##   female               11     7        9    64
##   male                  7     3       13    54
```

Above we can see the statistics of the three main variables we decided to use. This shows us a side by side

of all the outputs from the data. It shows us the difference in demographics and gender between public and private schools.

School Type Vs. Race Chart

```
ggplot(pol, aes(x=schtyp, fill=race)) + geom_bar(position = "dodge")
```
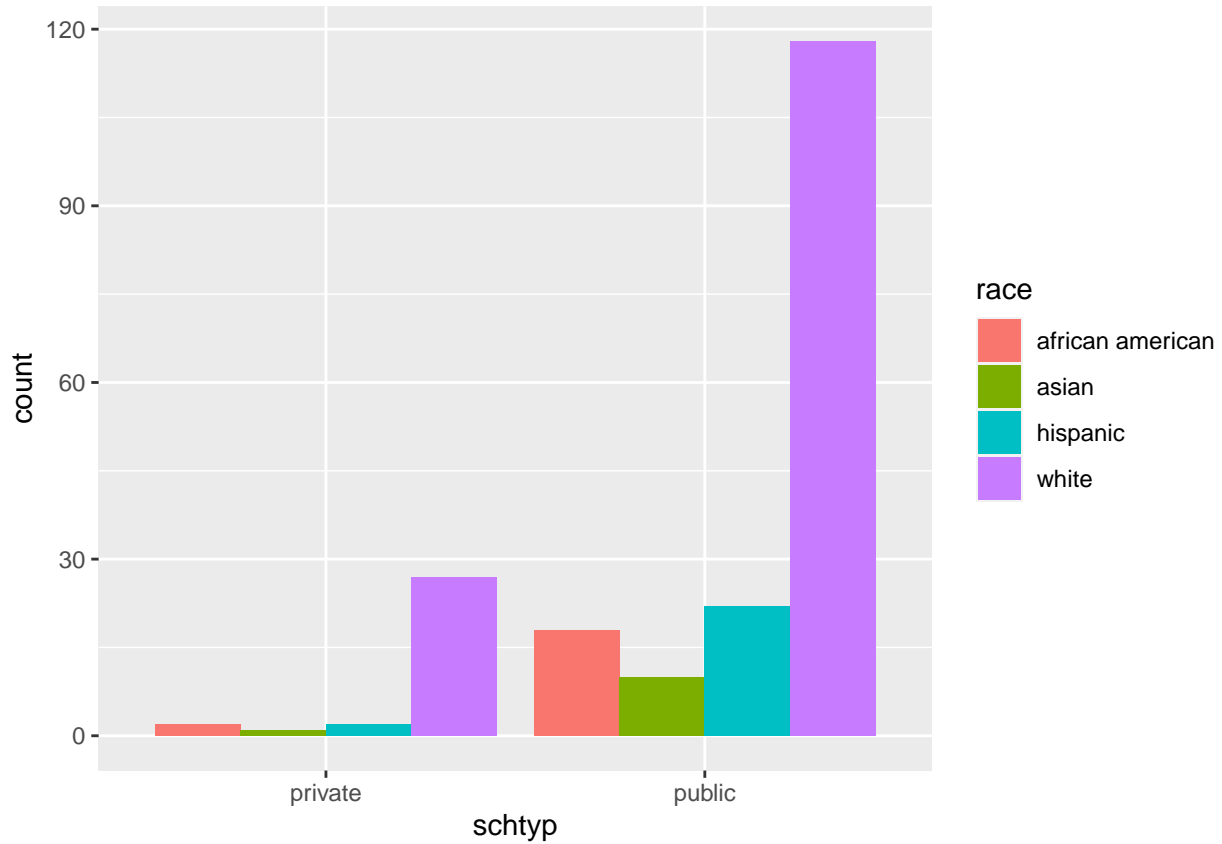
Here is a combined bar chart in which we can see the difference of Race(African American, Asian, White, and Hispanic) vs. School Type(private, Public).

4 Conclusion

Although each school had an approximately similar race ratio of whites, we can see that public and private schools consist of mainly both white demographics. We can also tell from both charts that public schools have a bigger demographic since there are more students to be in public rather than in a private school. With majority of them females (109) rather than males (91).