Exploratory Data Analysis Project

Aubrey Teckam

2024-03-04

Introduction

The High School and Beyond data set shows students' background information, test scores, and school programs in 200 observations of 11 variables. I want to study the correlation between public or private school and the type of program a student enters after high school. I attended an early college high school, more similar to a private school than public, and after graduating was able to finish my bachelor's degree in less than four years. I want to see if other students that attended public schools were more likely to attend academic programs after high school than public school students.

```
hsb2 <- read.table("/Users/aubreyteckam/Downloads/math130/data/hsb2.txt", header=TRUE, sep="\t")
library(ggplot2)
library(RColorBrewer)
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
```
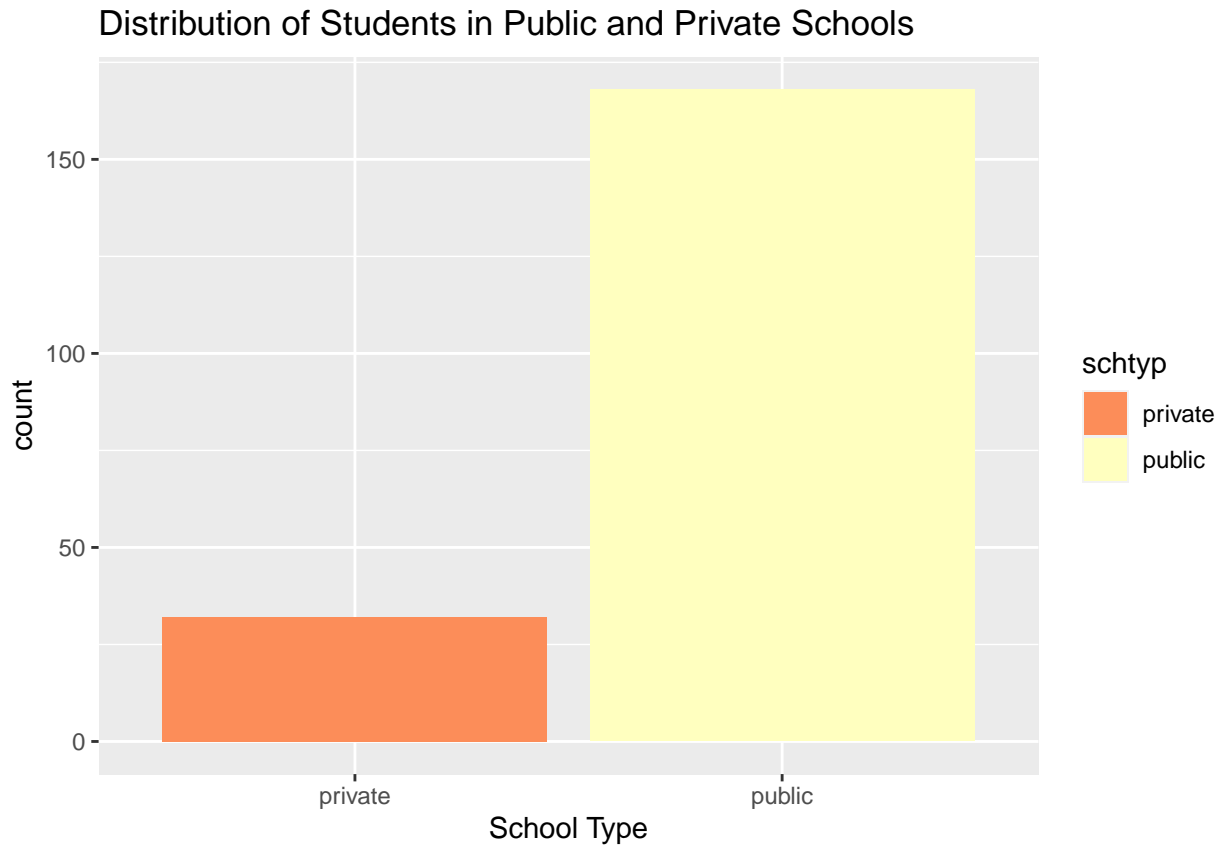
Univariate Data

School Type

```
table(hsb2$schtyp)
```

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

Below is a bar chart displaying the distribution of students in public and private schools.

```
ggplot(hsb2, aes(x=schtyp, fill=schtyp)) + geom_bar() + ggtitle("Distribution of Students in Public and
```

Distribution of Students in Public and Private Schools
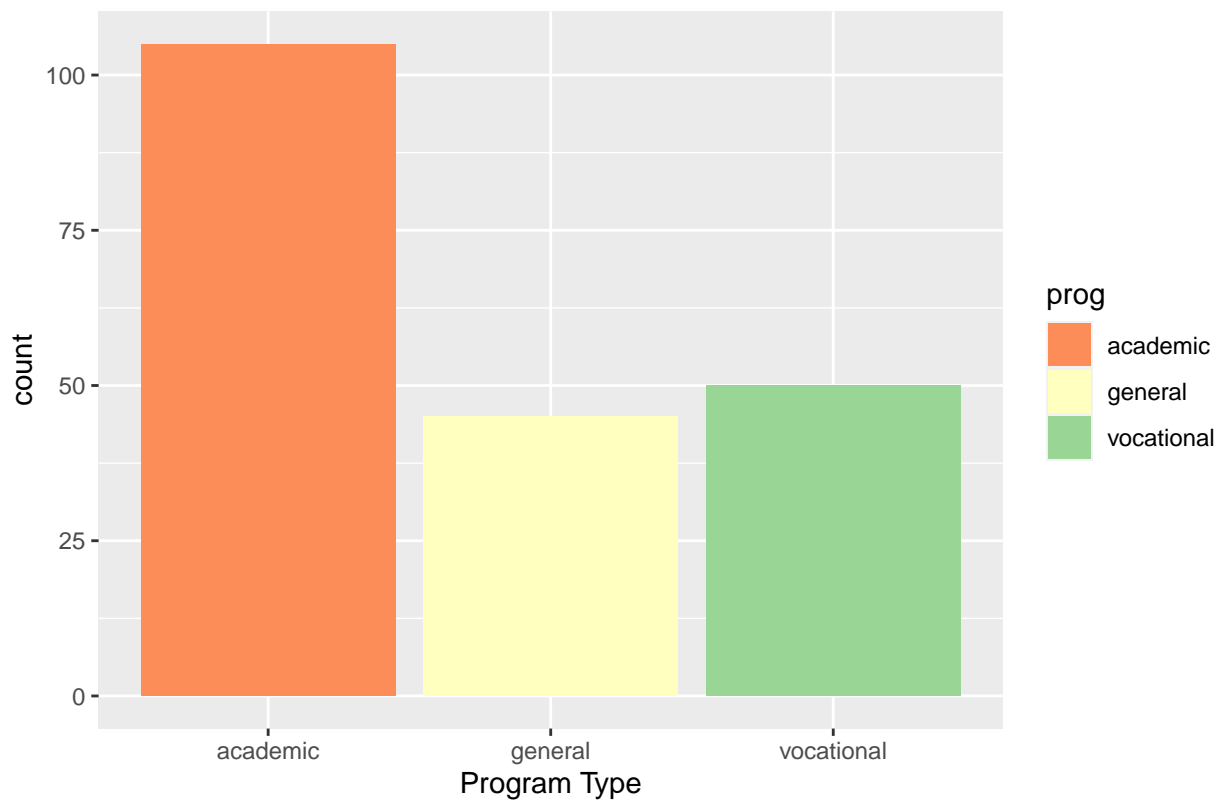

Program

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

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

Below is a bar chart that displays the distribution of students between program types.

```
ggplot(hsb2, aes(x=prog, fill=prog)) + geom_bar() + ggtitle("Distribution of Students in Various Educati
```

Distribution of Students in Various Educational Programs
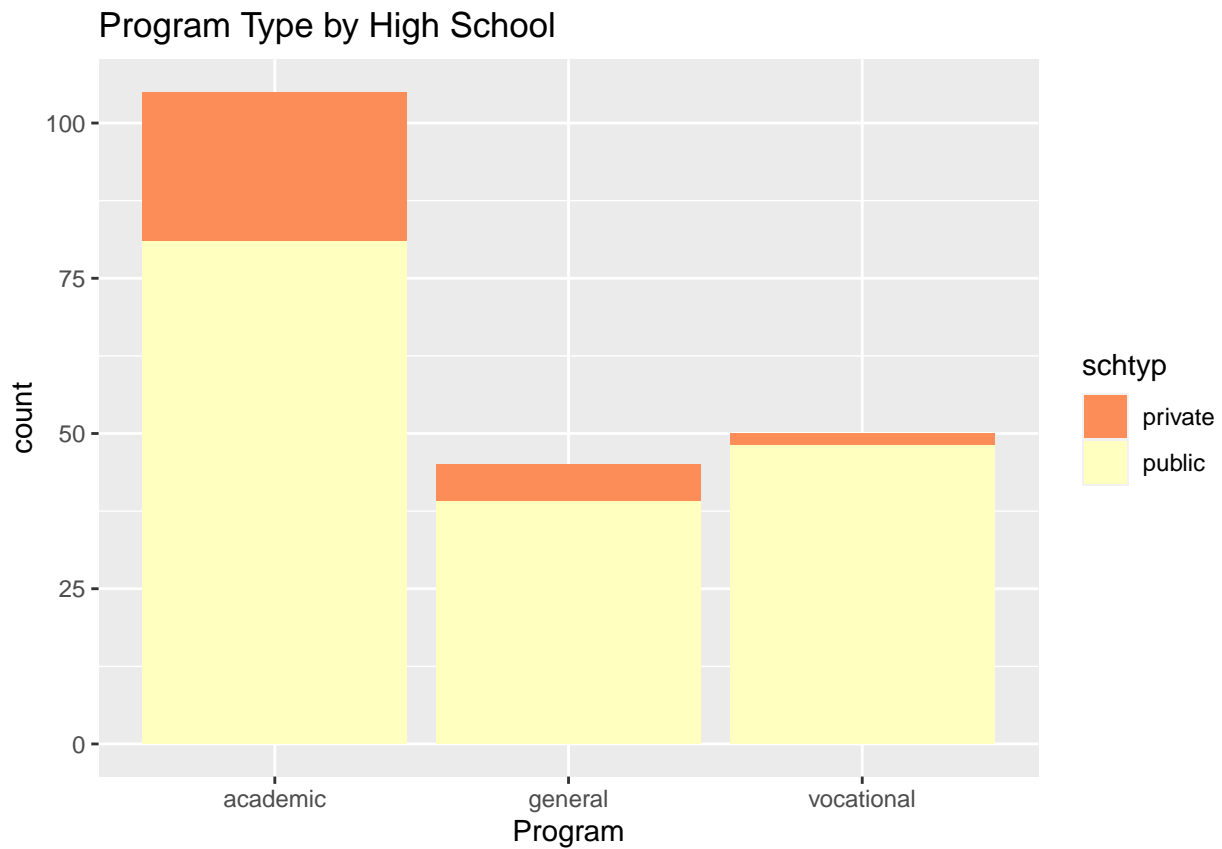

Bivariate Data

```
table(hsb2$schtyp, hsb2$prog) %>% prop.table(margin=1)
```

```
##
##           academic   general vocational
##   private 0.7500000 0.1875000  0.0625000
##   public  0.4821429 0.2321429  0.2857143
```

Below is a boxplot showing the relationship between

```
ggplot(hsb2, aes(x=prog, fill=schtyp)) + geom_bar() + ggtitle("Program Type by High School") + xlab("Pro
```

Program Type by High School

Conclusion

In conclusion, private school students are more likely to go to an academic program after high school than a general or vocational program. In general students are more likely to go to an academic program, however public school students are more likely to go into general or vocational programs than private school students. These findings prove my hypothesis to be correct.