

B3

Jeremiah Theisen

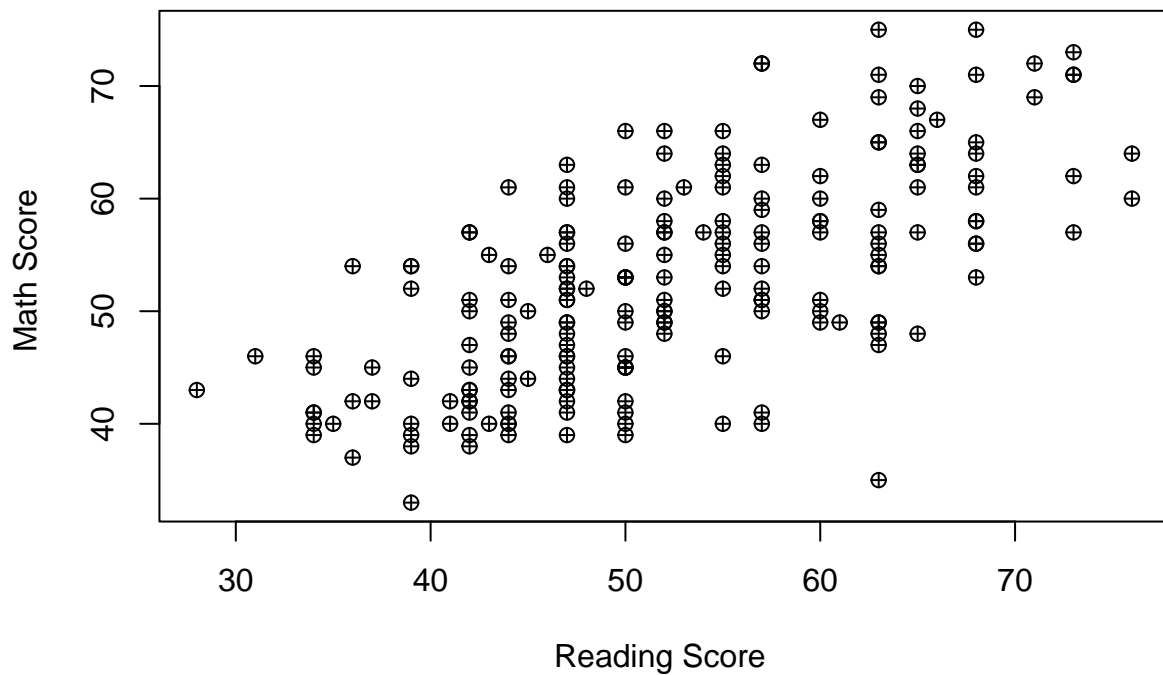
2024-08-27

```
hsb2 = read.csv("https://github.com/TienChih/tbil-stats/raw/main/data/hsb2.csv")
names(hsb2)
```

```
## [1] "id"      "gender"  "race"    "ses"     "schtyp"  "prog"    "read"
## [8] "write"   "math"    "science" "socst"
```

Activity 1.3.1

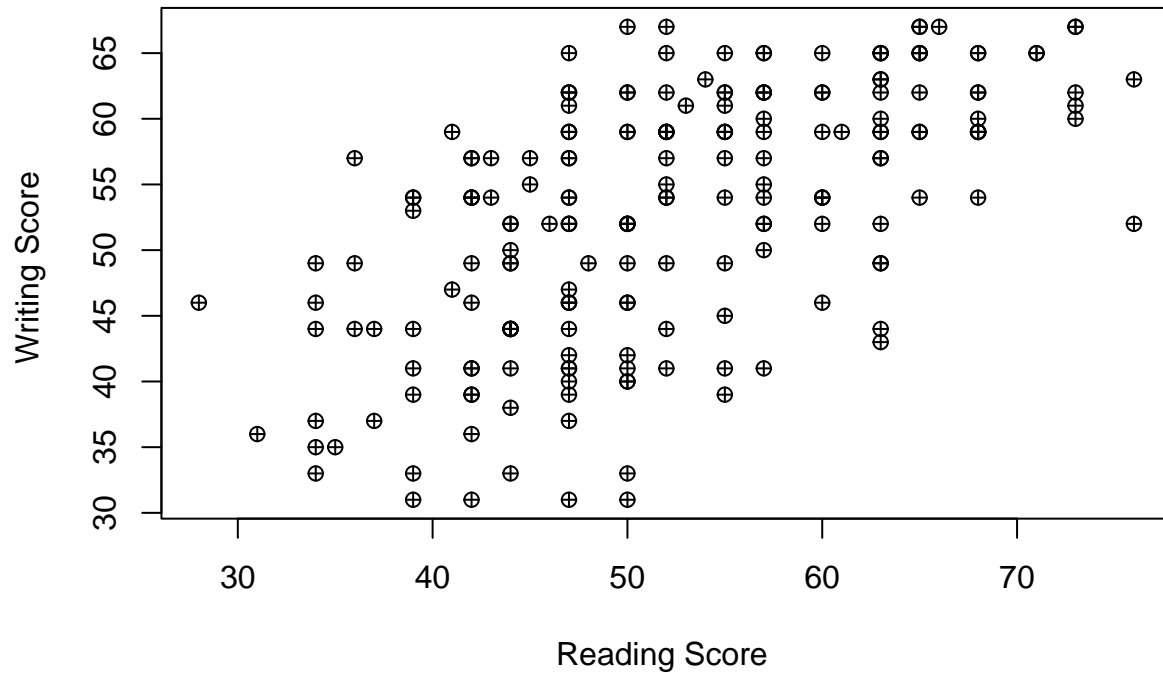
```
plot(hsb2$read, hsb2$math, xlab="Reading Score ", ylab="Math Score ", pch=10)
```



a. Each dot represents one student. Its coordinates represent the student's reading and math score.

- b. There is a strong positive relationship.
- c. done
- d.

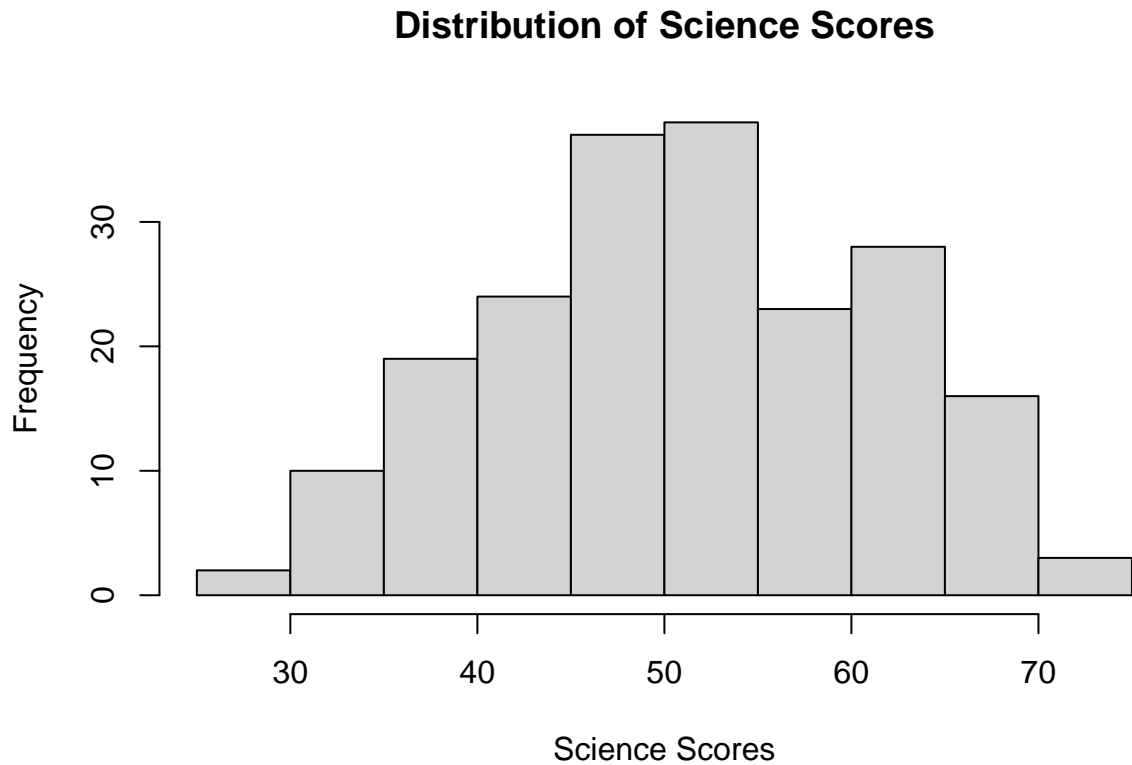
```
plot(hsb2$read, hsb2$write, xlab="Reading Score ", ylab="Writing Score ", pch=10)
```



- e. Each dot is a student. Its coordinates represent the student's Reading and Writing scores.
- f. There is a positive relationship, although not as strong as I was expecting

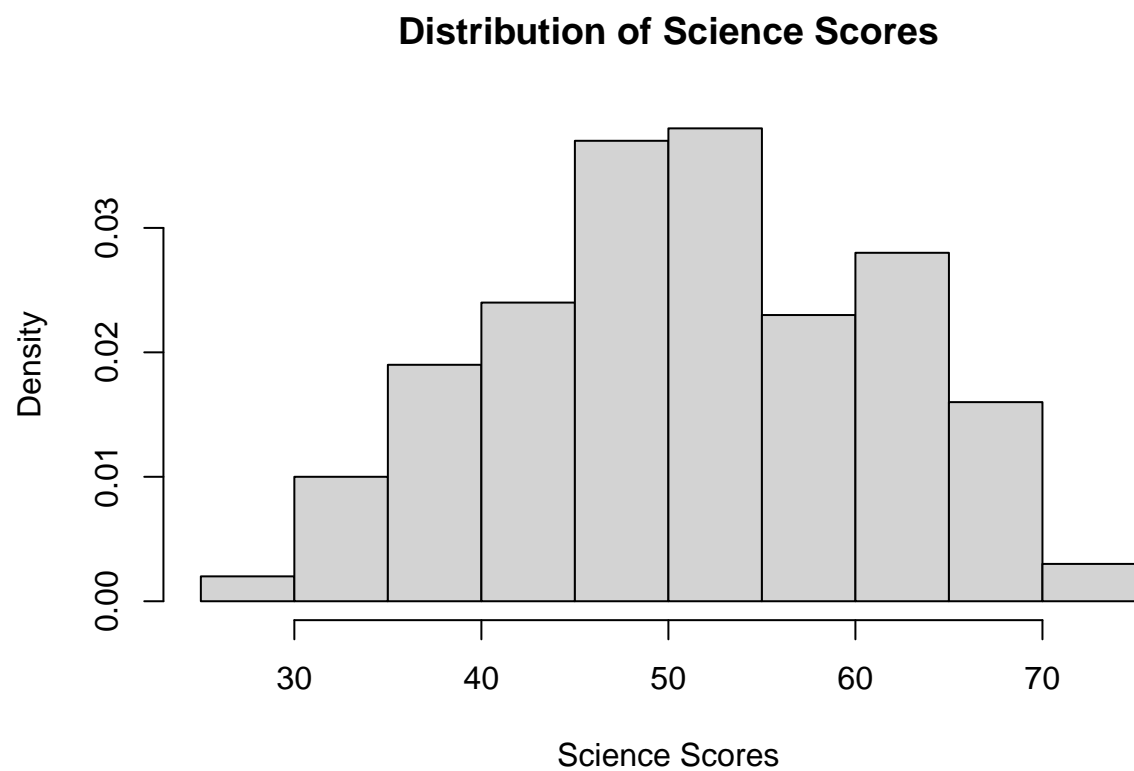
Activity 1.3.2

```
hist(hsb2$science, xlab="Science Scores", main="Distribution of Science Scores" )
```



- a. The x value represents different science score, the height represents the proportion of students which got that score.
- b. The graph looks like a triangle, but with more on the right side, after 50, than before 50. 45 and 50 are spikes, which are more frequent than any of the ones around them.
- c.

```
hist(hsb2$science, xlab="Science Scores", main="Distribution of Science Scores",freq=FALSE )
```



- d.
- e.

```
hist(hsb2$math, xlab="Math Scores", main="Distribution of Math Scores" )
```



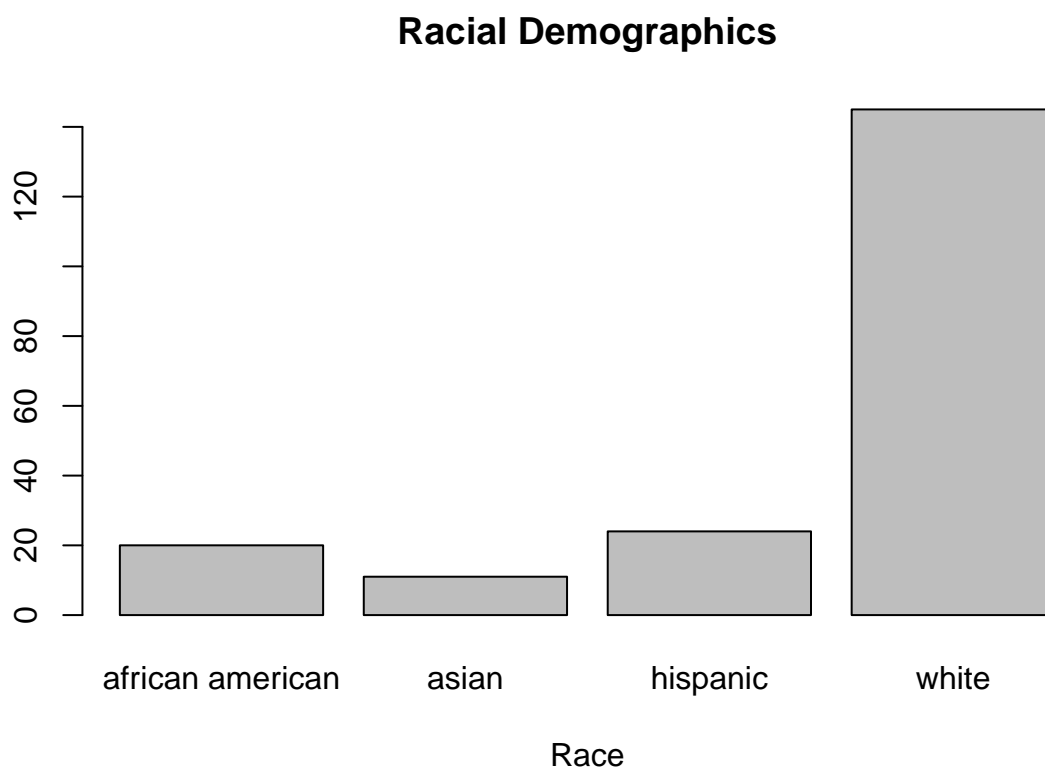
- f. The x value represents a score in math, the height represents the number of students which got that score.
- g. Although this plot still retains the triangle from the Science scores, there is a more even spread of scores between 40-60, rather than the two spikes from the Science score.
- h.

```
hist(hsb2$math, xlab="Math Scores", main="Distribution of Math Scores", freq=FALSE )
```



Activity 1.3.3

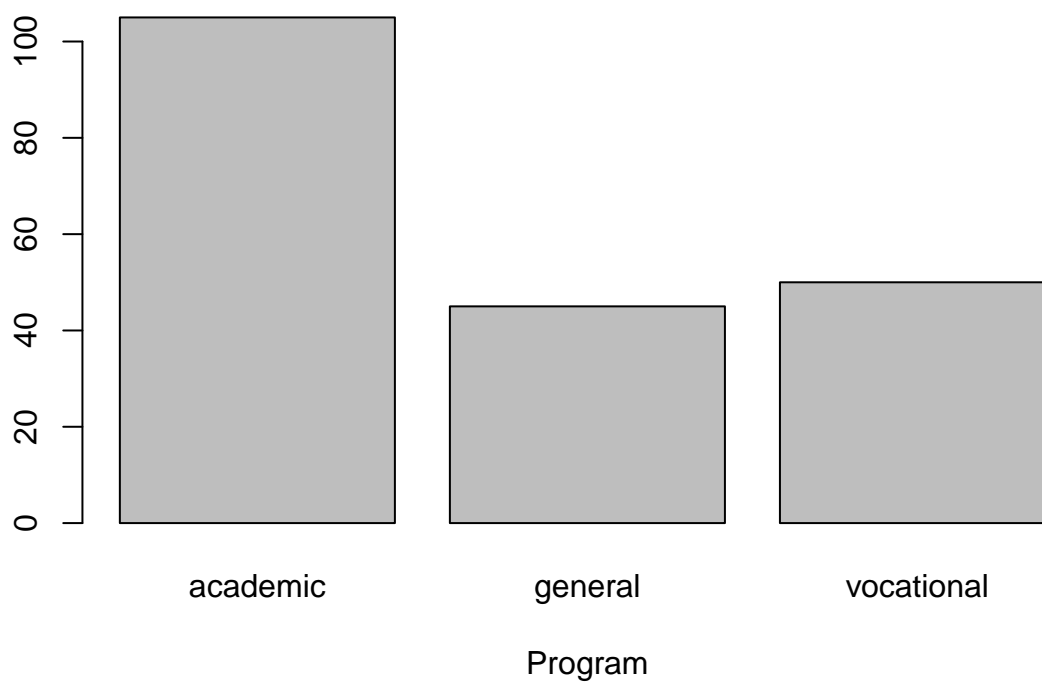
```
counts = table(hsb2$race)
barplot(counts, main="Racial Demographics", xlab="Race")
```



- a. Each bar represents a racial group, the height represents the number of students of said group.
- b.
- c.

```
counts = table(hsb2$prog)
barplot(counts, main="School Programs", xlab="Program")
```

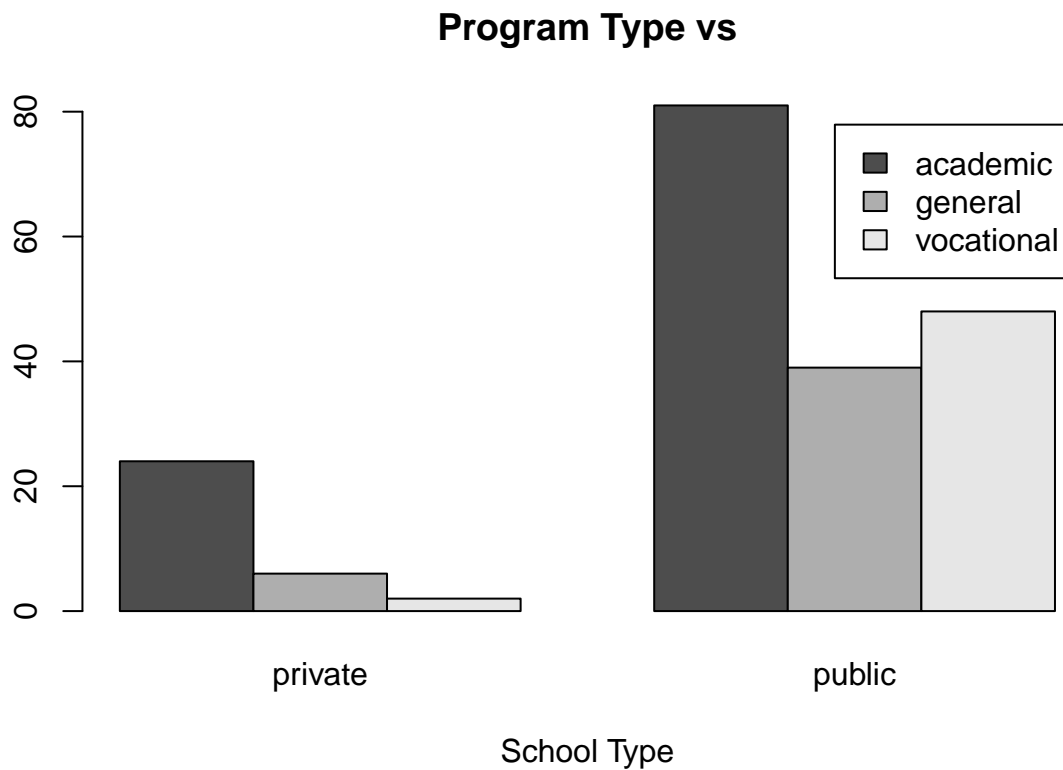
School Programs



- d. Most students are in academic programs.
- e. Each bar represents a program, the heights represent number of students are in each type of program.

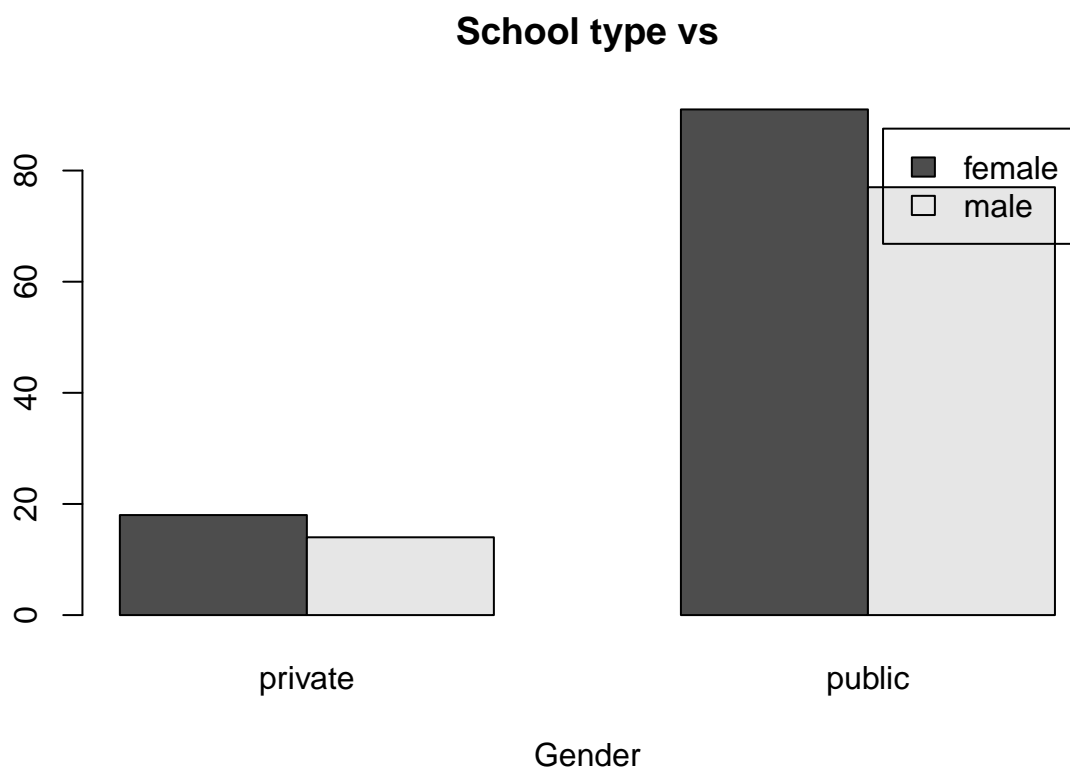
Activity 1.3.4

```
counts = table(hsb2$prog, hsb2$schtyp)
barplot(counts, main="Program Type vs", xlab="School Type",
        legend = rownames(counts), beside=TRUE)
```

- Each bar represents a program, the heights represent the number of students in each program at a specific type of school
- The private schools are smaller, and have proportionally less students in vocational programs
-
-

```
counts = table(hsb2$gender, hsb2$schtyp)
barplot(counts, main="School type vs", xlab="Gender",
        legend = rownames(counts), beside=TRUE)
```

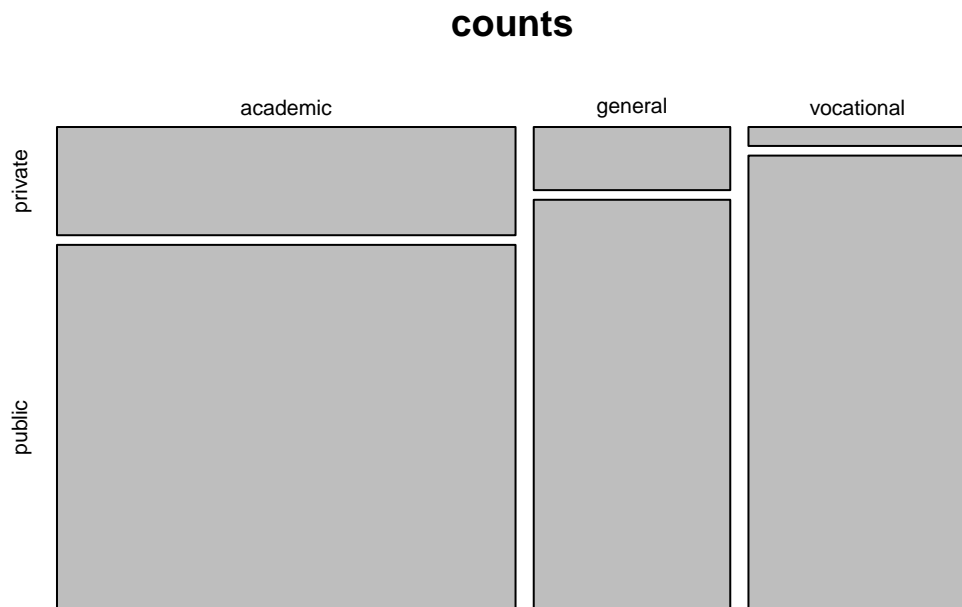


e. There are more female students overall

f. Each bar is a gender at a different type of school, the heights are the number of students.

Activity 1.3.5

```
counts = table(hsb2$prog, hsb2$schtyp)
mosaicplot(counts)
```



- The area of each rectangle is the relative number of students in a program in one type of school compared to another type of school. The position is what type of program it is, and what type of school it is at.
- This graph visualizes the relative difference between program sizes at different school types.
-
-

```
counts = table(hsb2$gender, hsb2$prog)
mosaicplot(counts)
```

counts



- e. The gender ratio across different programs is almost the same.
- f. The rectangles represent the proportion of each gender that goes to each type of school.