CptS575 Hw2

Mengxiao

Part 1

a. Read the data into R

```
data = read.csv('https://scads.eecs.wsu.edu/wp-content/uploads/2017/09/College.csv')
```

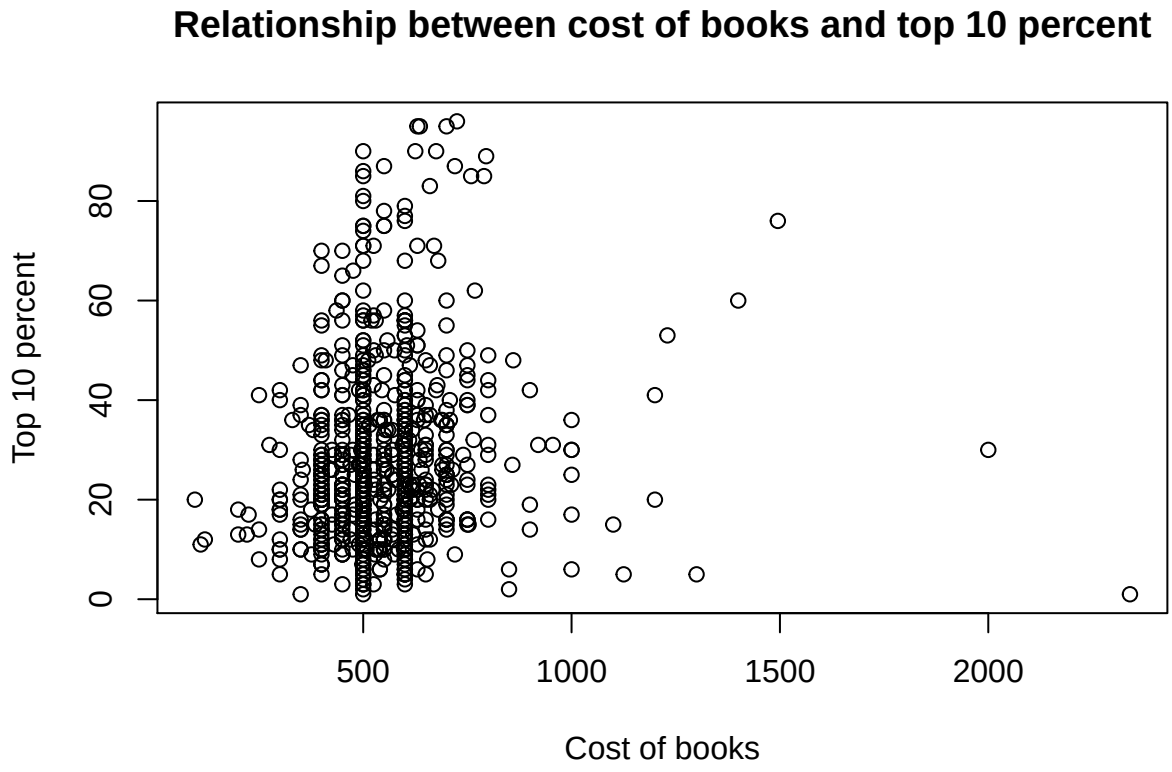
b. Find the median cost of books

```
books_median = median(data['Books'][, 1])
```

```
## [1] 500
```

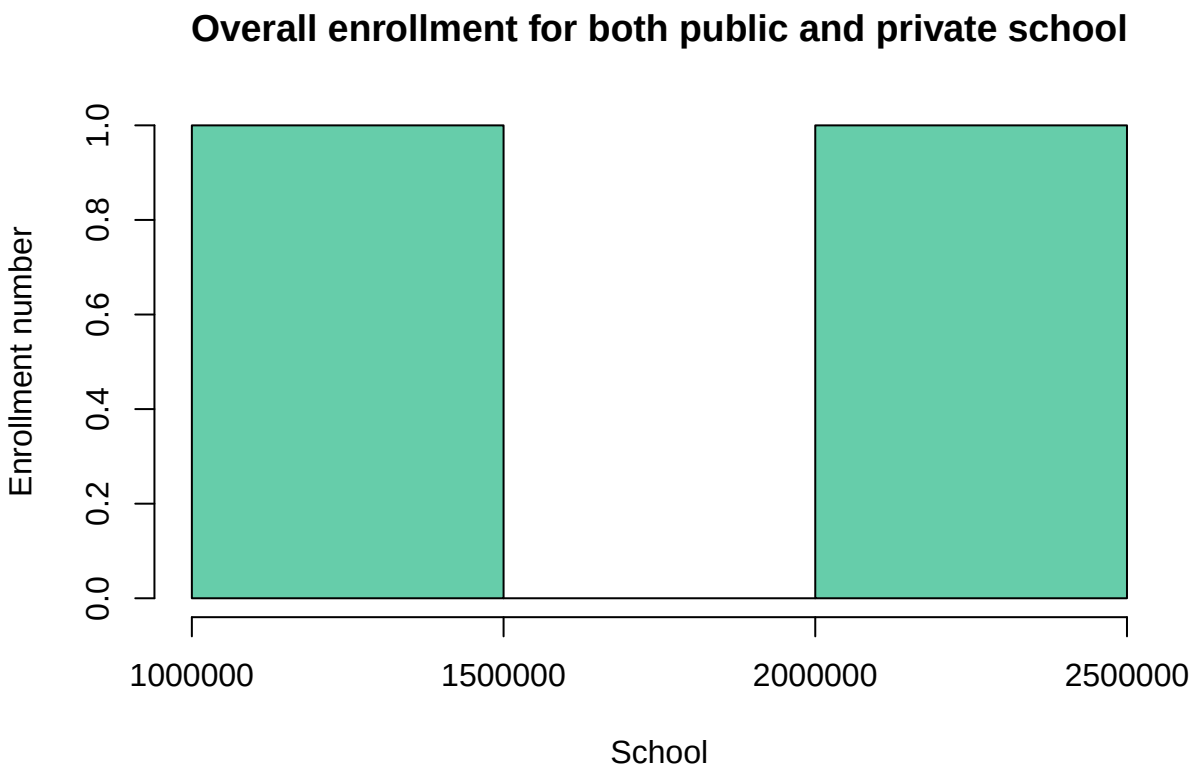
c. Produce a scatterplot to show the relationship between the cost of books and Top 10 percent students.

```
plot(x = data$Books,
     y = data$Top10perc,
     xlab = "Cost of books",
     ylab = "Top 10 percent",
     main = "Relationship between cost of books and top 10 percent"
     )
```



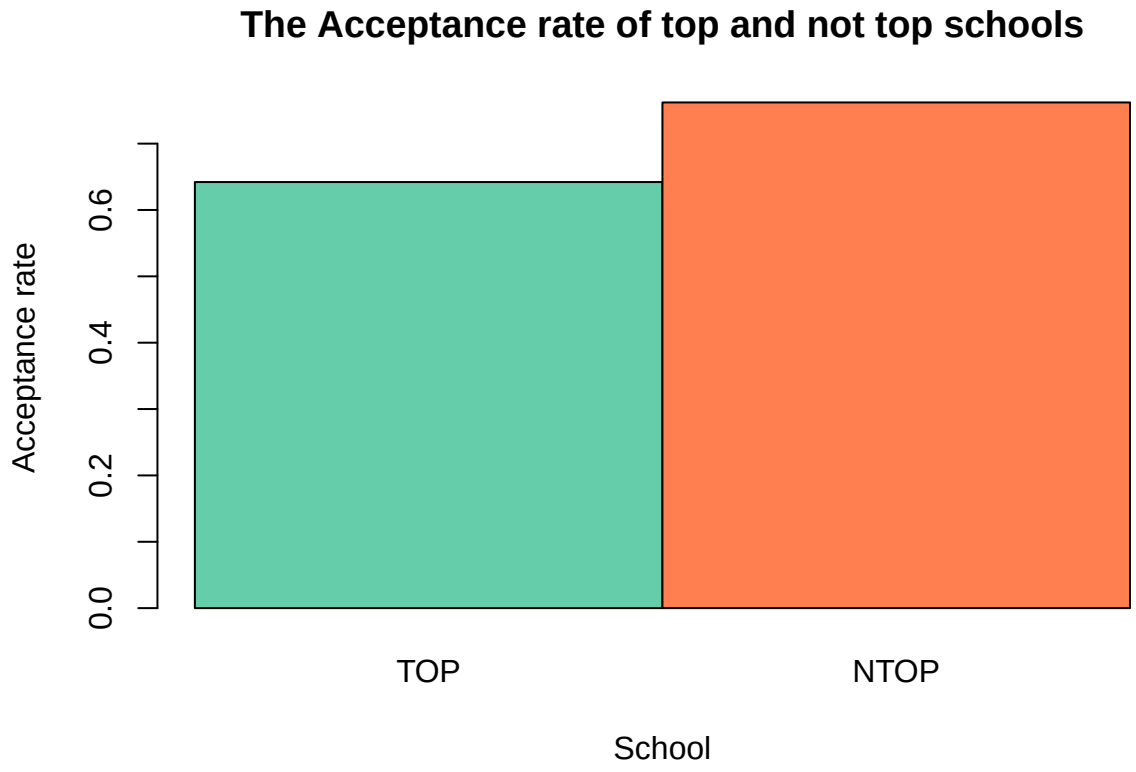
d. Produce a histogram showing the overall enrollment numbers for both public and private schools.

```
OverallEnroll <- c(sum(data[data$Private == 'No',]$P.Undergrad)
                   +sum(data[data$Private == 'No',]$F.Undergrad),
                   sum(data[data$Private == 'Yes',]$P.Undergrad)
                   +sum(data[data$Private == 'Yes',]$F.Undergrad))
hist(OverallEnroll, main = "Overall enrollment for both public and private school",
     xlab = "School", ylab = "Enrollment number", col=c("aquamarine3", "coral"),
    )
```



e. Seperate the schools to two parts, top and ntop(not top)

```
top <- sum(data[data$Top25perc>50,]$Accept)/sum(data[data$Top25perc>50,]$Apps)
ntop<- sum(data[data$Top25perc<=50,]$Accept)/sum(data[data$Top25perc<=50,]$Apps)
barplot(matrix(c(top, ntop)), ylab="Acceptance rate", xlab="School", col=c("aquamarine3", "coral"),
    main="The Acceptance rate of top and not top schools", beside=TRUE, width=0.2,
    names.arg = c("TOP", "NTOP"))
```

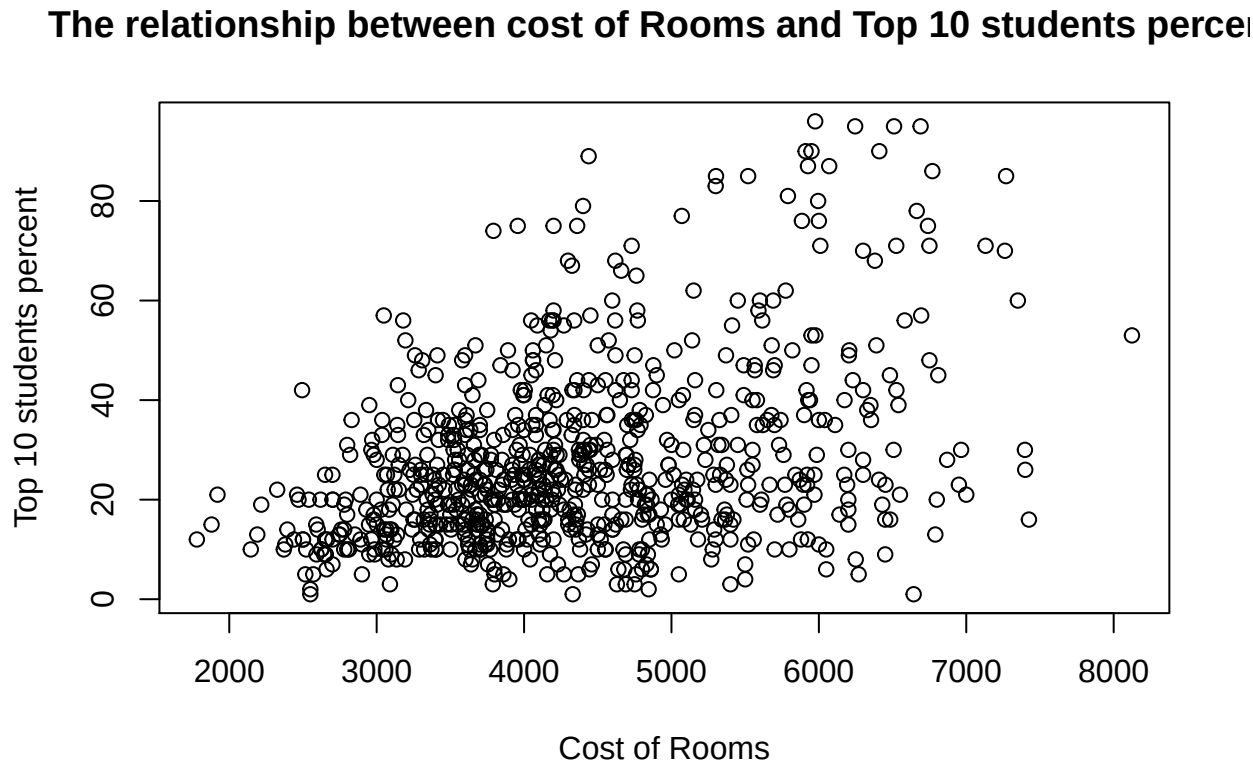



```
#legend("topleft", legend=c("TOP","NTOP"), fill=c("aquamarine3", "coral"))
```

f. Produce two new plots.

1. I want to explore the relationship between cost of Rooms and Top 10 students percentage.

```
plot(x = data$Room.Board,
     y = data$Top10perc,
     xlab = "Cost of Rooms",
     ylab = "Top 10 students percent",
     main = "The relationship between cost of Rooms and Top 10 students percent")
```



2. I want to build a pie chart to show the proportion of private and public schools.

```
school_public <- length(data[data$Private == 'Yes',]$Private)
school_private<- length(data[data$Private == 'No', ]$Private)
pie(c(school_public, school_private),
    labels = c("Private", "Public"))
```