homework2

Na Yun Cho

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
library(splines)
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-33. For overview type 'help("mgcv-package")'.
```

```
library(pdp)
library(earth)
```

```
## Loading required package: Formula
```

```
## Loading required package: plotmo
```

```
## Loading required package: plotrix
```

```
## Loading required package: TeachingDemos
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x purrr::lift()     masks caret::lift()
## x purrr::partial()  masks pdp::partial()
```

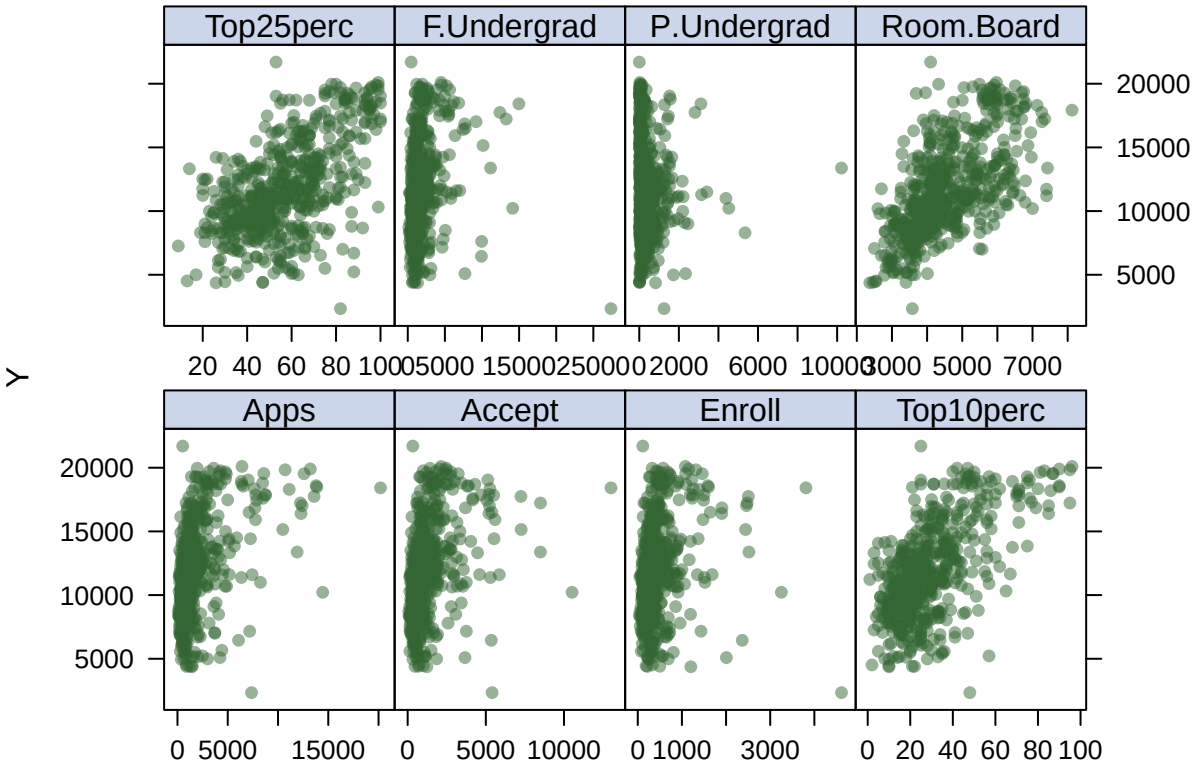
```
library(ggplot2)
```

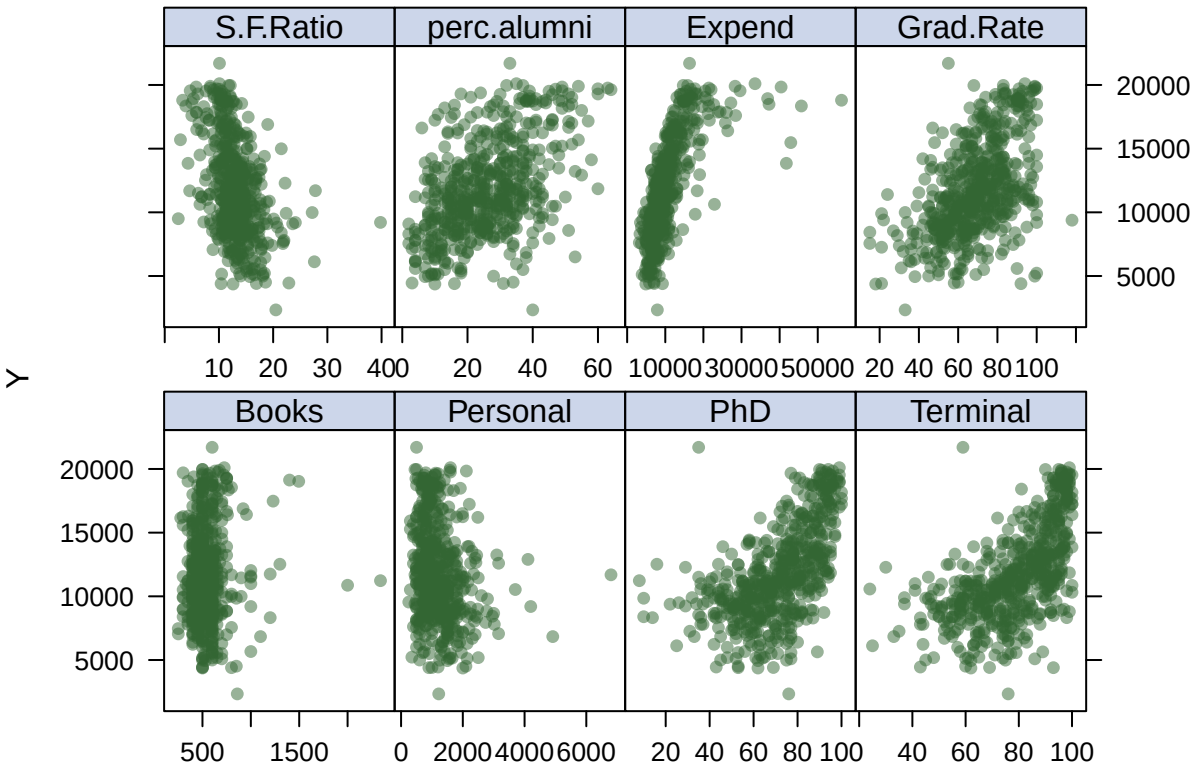
(a) Exploratory data analysis

```
college = read.csv("./data/data.csv")
college1 <- college[-125,]

college2 = data.matrix(college1, rownames.force = NA)
x <- college2 [ ,-c(1,9)]
y <- college2 [ , 9]


theme1 <- trellis.par.get()
theme1$plot.symbol$col <- rgb(.2, .4, .2, .5)
theme1$plot.symbol$pch <- 16
theme1$plot.line$col <- rgb(.8, .1, .1, 1)
theme1$plot.line$lwd <- 2
theme1$strip.background$col <- rgb(.0, .2, .6, .2)
trellis.par.set(theme1)
featurePlot(x, y, plot = "scatter", labels = c("","Y"),
            type = c("p"), layout = c(4, 2))
```




S.F.Ratio
perc.alumni
Expend
Grad.Rate
Books
Personal
PhD
Terminal
Y
20000
15000
10000
5000
10 20 30 40
20 40 60
10000 30000 50000
20 40 60 80 100
500 1500
0 2000 4000 6000
20 40 60 80 100
40 60 80 100