second_iteration

2024-06-23

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(gridExtra) # This is use to create multiple grid when working with ggplot else we would have us
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

1. Loading the data

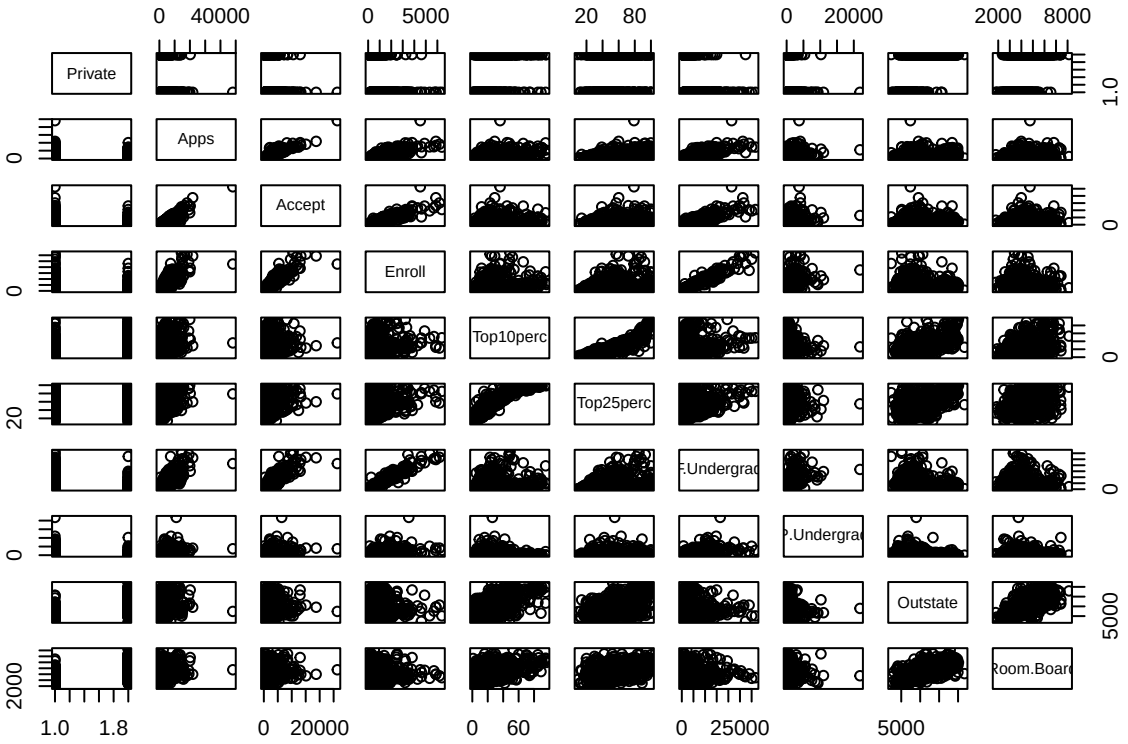
```
college <- read.csv("D:\\R\\ISLR\\College.csv")
View(college)
```

2. Replacing the row names with the name of colleges
3. Calculating acceptance rate by dividing the total accepted students by the total number of studenets who applied.

```
rownames(college) <- college[,1]
college <- college[,-1]
```

4. Converting the variable private to factor

```
college$Private <- as.factor(college$Private)
pairs(college[,1:10])
```

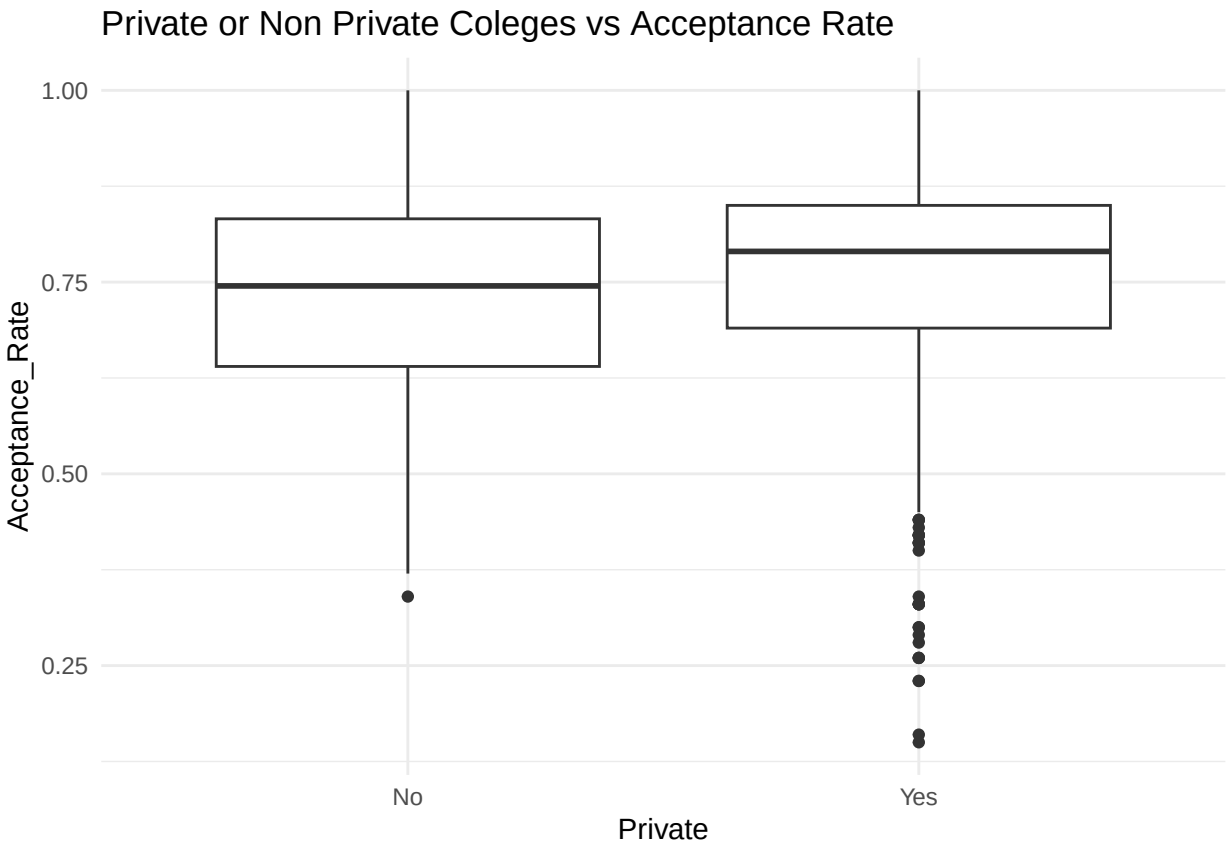


5. Finding Elite universities with Top10perc > 50

```
college$Elite <- college$Top10perc > 50
college$Acceptance_Rate<- round(college$Accept/college$Apps,2)
```

6. Acceptance rate for Private and Non-Private Colleges.

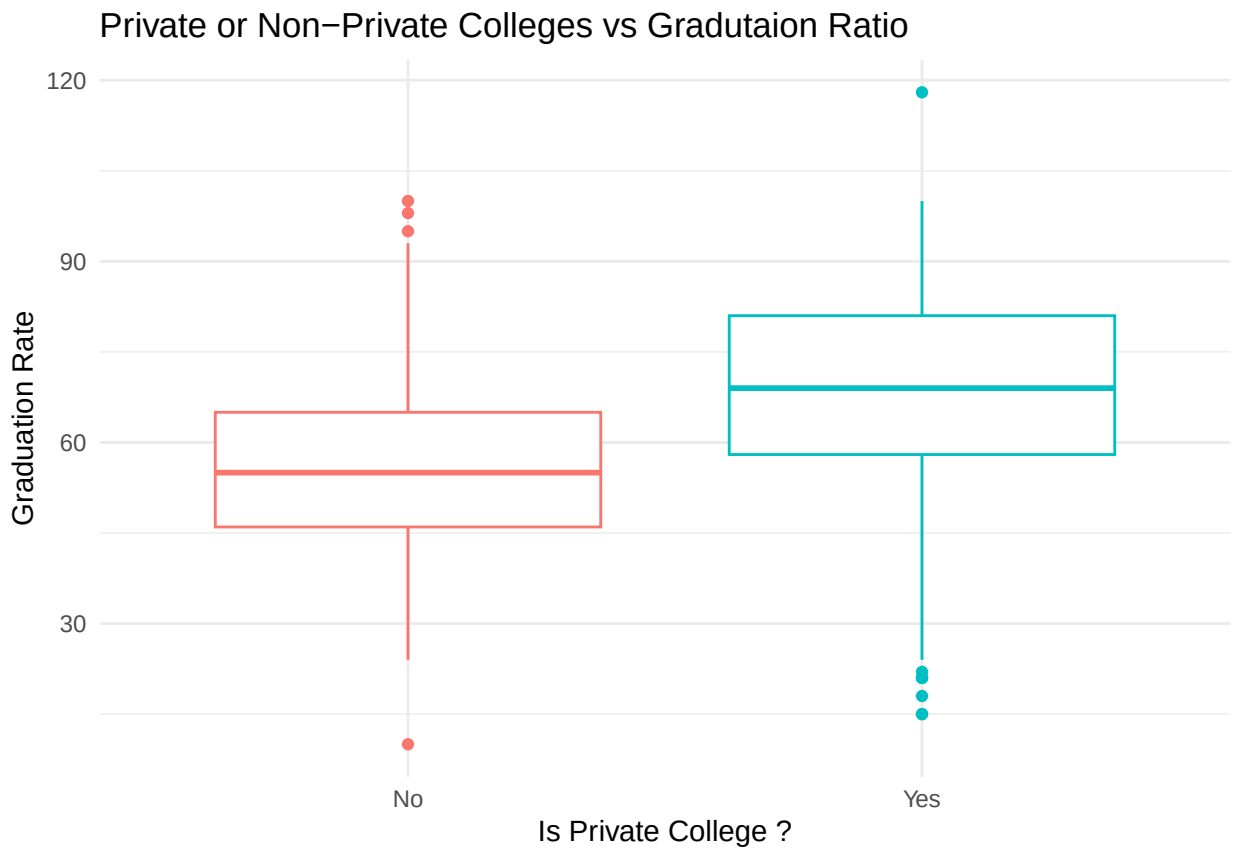
```
college %>%
  ggplot(aes(x=Private,y=Acceptance_Rate))+
  geom_boxplot()+
  theme_minimal()+
  labs(title="Private or Non Private Coleges vs Acceptance Rate")
```



For Non Private colleges the acceptance rate is slightly higher than non private colleges.

7. What is the situation of graduation rate for private and non private colleges.

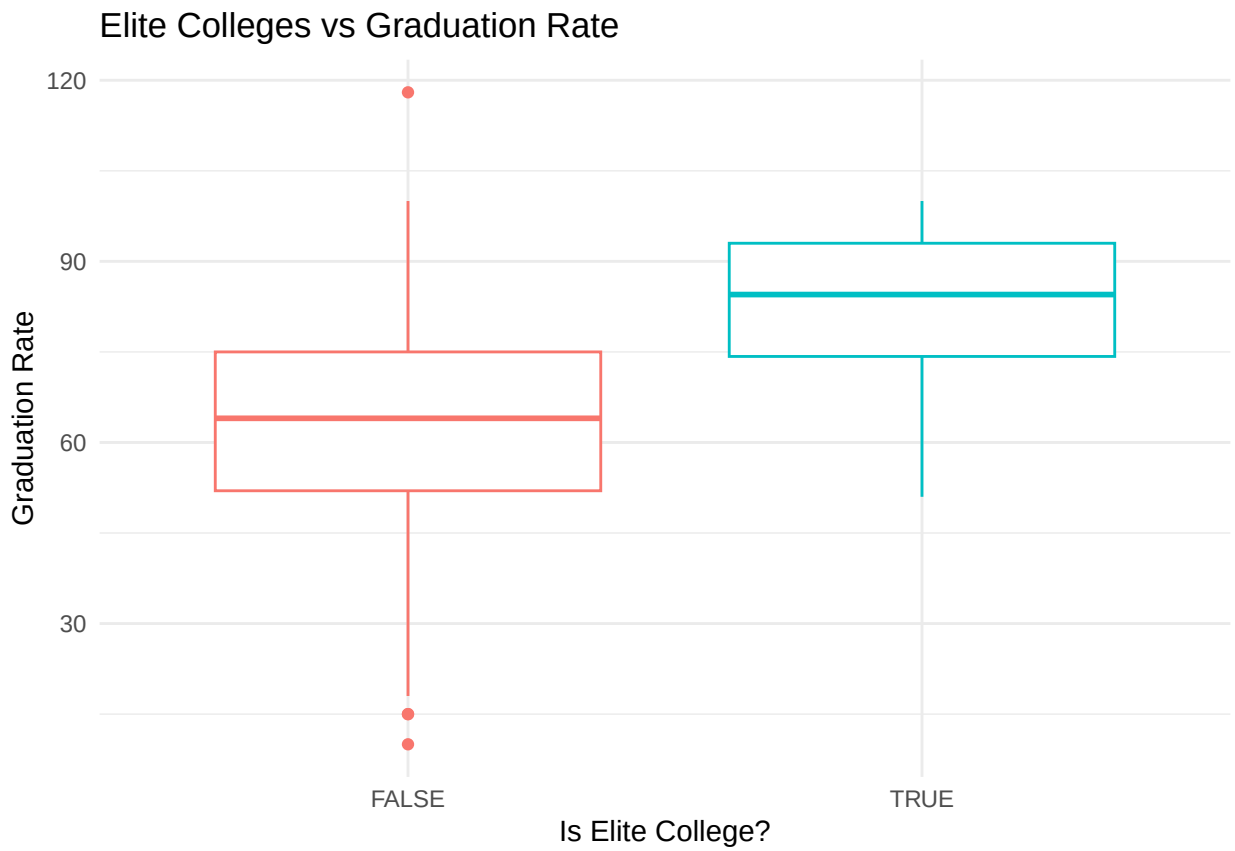
```
college %>%
  ggplot(aes(x=Private,y=Grad.Rate))+
  geom_boxplot(aes(color=Private),show.legend=FALSE)+
  theme_minimal()+
  labs(x="Is Private College ?",
       y="Graduation Rate",
       title="Private or Non-Private Colleges vs Gradutaion Ratio")
```



The Graduation Rate for Private Colleges are higher than the graduation rate for Non-Private Colleges.

8. Graduation Rate for Elite or Non Elite colleges

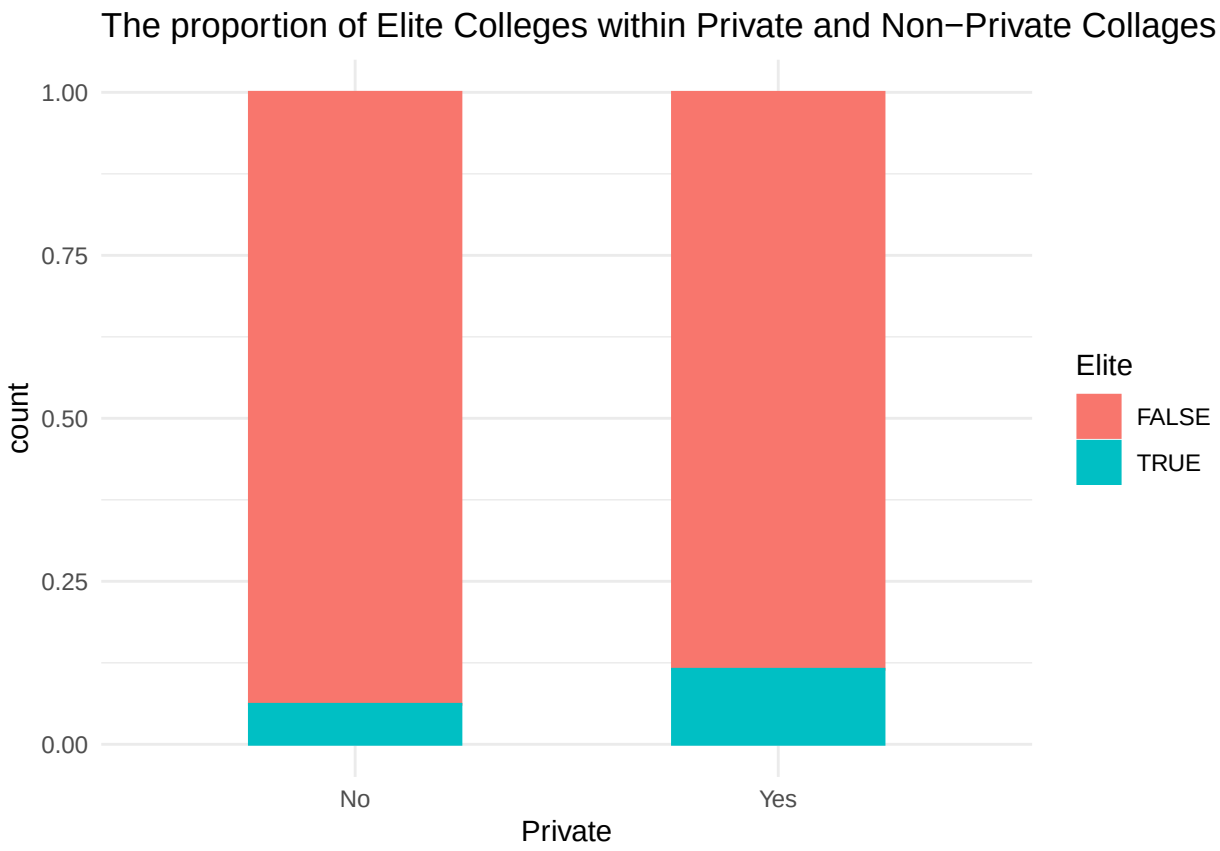
```
college %>%
  ggplot(aes(x=Elite,y=Grad.Rate))+
  geom_boxplot(aes(color=Elite),show.legend=FALSE)+
  theme_minimal()+
  labs(x="Is Elite College?",
       y="Graduation Rate",
       title="Elite Colleges vs Graduation Rate")
```



Clearly The Elite Colleges have higher Graduation Rate than Non Elite Colleges

8. Is Elite college belog to Private or non Private

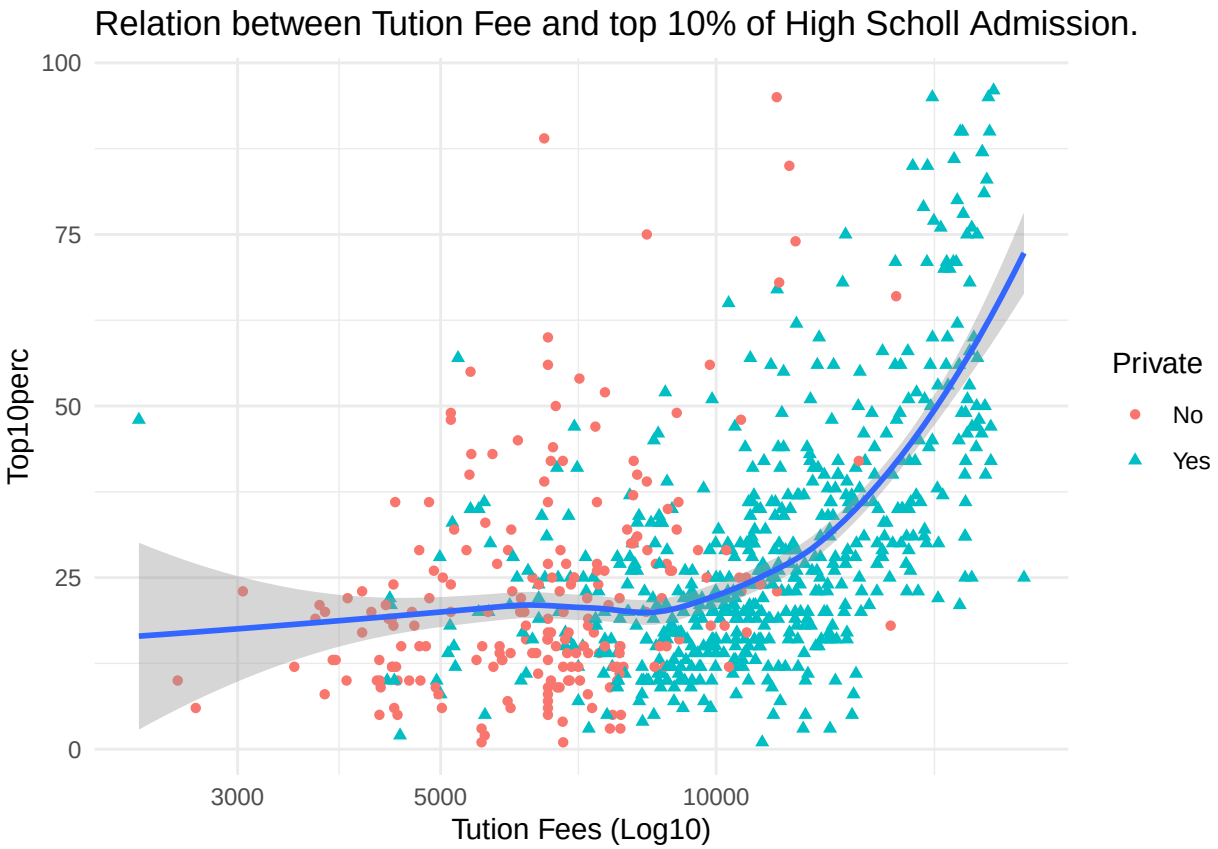
```
college %>%
  ggplot(aes(x=Private))+
  geom_bar(aes(color=Elite,fill=Elite),
           position="fill",
           stat="count",
           width=0.5)+
  theme_minimal()+
  labs(title="The proportion of Elite Colleges within Private and Non-Private Collages")
```



9. Is there a relationship between top 10 high school students admitted and the tution fee

```
college %>%
  ggplot(aes(x=Outstate,y=Top10perc))+
  geom_point(aes(color=Private,shape=Private))+
  scale_x_log10()+
  geom_smooth()+
  theme_minimal()+
  labs(x="Tution Fees (Log10)",
       title="Relation between Tution Fee and top 10% of High Scholl Admission.")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

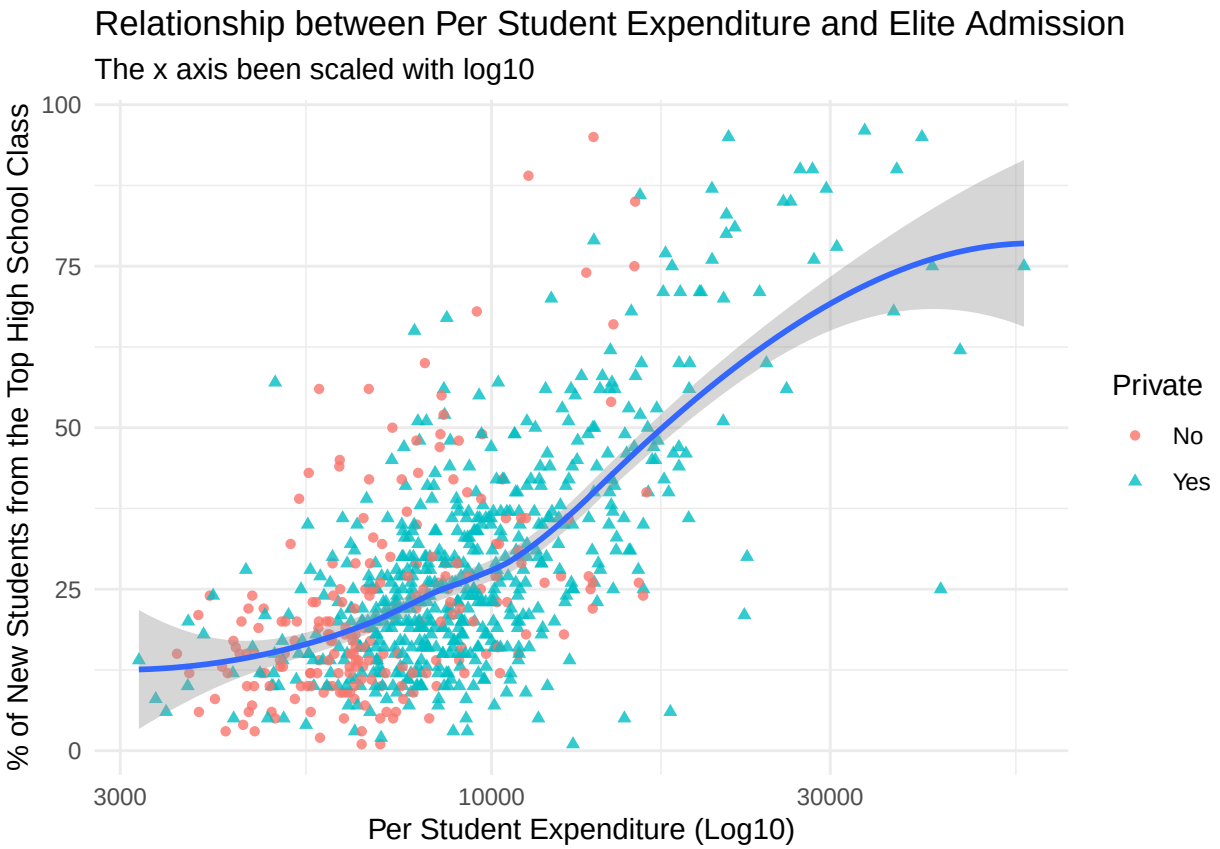


The higher fees correlate to Private Colleges and high number of admission of new students from top 10% of high school class in those colleges as well.

10. Is the Expenditure per student related with the admission of new students from top 10% of high school

```
college %>%
  ggplot(aes(x=Expend,y=Top10perc))+
  geom_point(aes(color=Private,shape=Private),alpha=0.8)+
  geom_smooth()+
  scale_x_log10()+
  theme_minimal()+
  labs(x="Per Student Expenditure (Log10)",
       y="% of New Students from the Top High School Class",
       title="Relationship between Per Student Expenditure and Elite Admission",
       subtitle="The x axis been scaled with log10")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

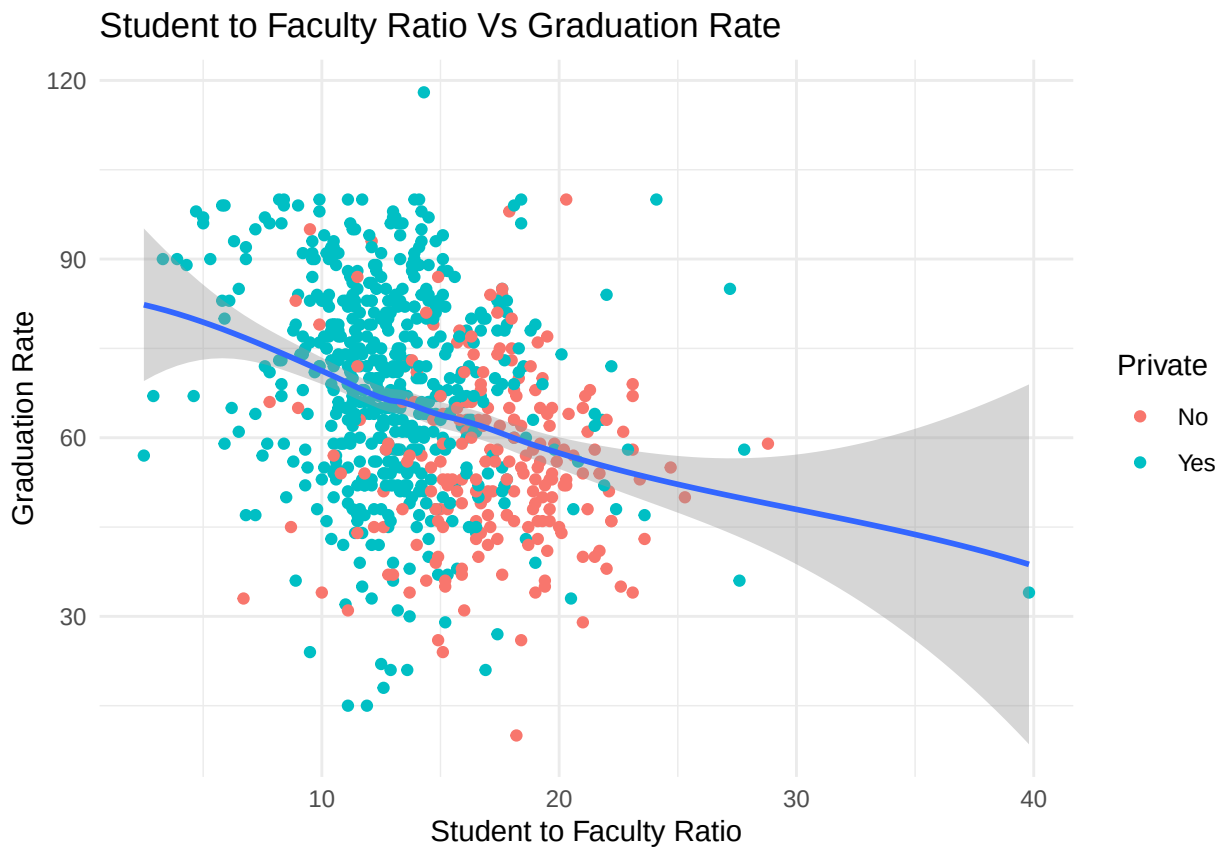


The colleges which have high student expenditure, there seems to be higher admission from top10% of high school classes.

11. Does Student to Faculty Ratio is Related to Graduation Rate

```
college %>%
  ggplot(aes(x=S.F.Ratio,y=Grad.Rate))+
  geom_point(aes(color=Private))+
  geom_smooth()+
  theme_minimal()+
  labs(x="Student to Faculty Ratio",
       y="Graduation Rate",
       title="Student to Faculty Ratio Vs Graduation Rate")
```

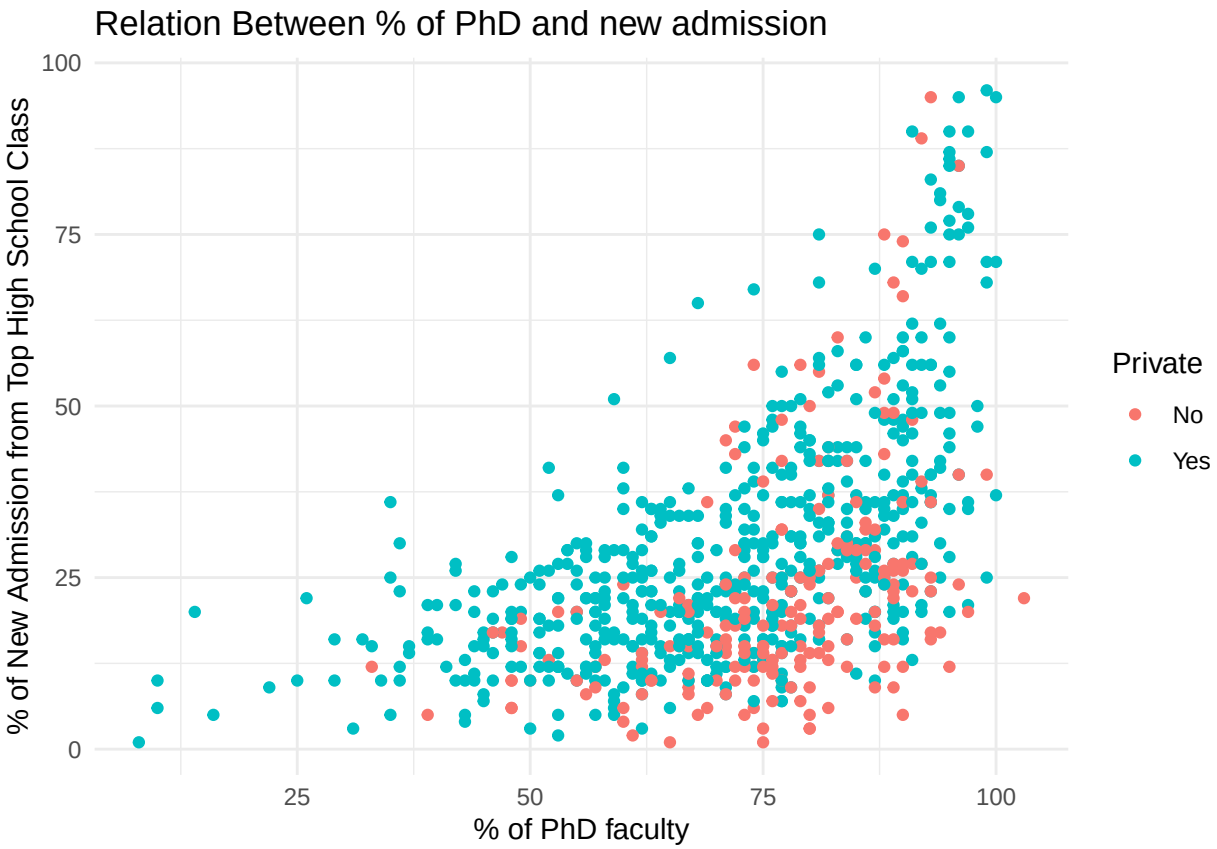
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



There is Downward Trend Seen. When Student to Faculty Ratio increases The Graduation Rate decreases. And Student to Faculty Ratio In Private college are lesser than that of Non-Private Colleges.

12. Is the Admission of new Students from top 10% of high school related to the percentage of Phd Faculty present in the university.

```
college %>%
  ggplot(aes(x=PhD,y=Top10perc))+
  geom_point(aes(color=Private))+
  theme_minimal()+
  labs(x="% of PhD faculty",
       y="% of New Admission from Top High School Class",
       title="Relation Between % of PhD and new admission")
```


Relation Between % of PhD and new admission
% of New Admission from Top High School Class
100
75
50
25
0
25
50
75
100
% of PhD faculty
Private
No
Yes