output: html_document

Data Exploration/Exploration

Setup

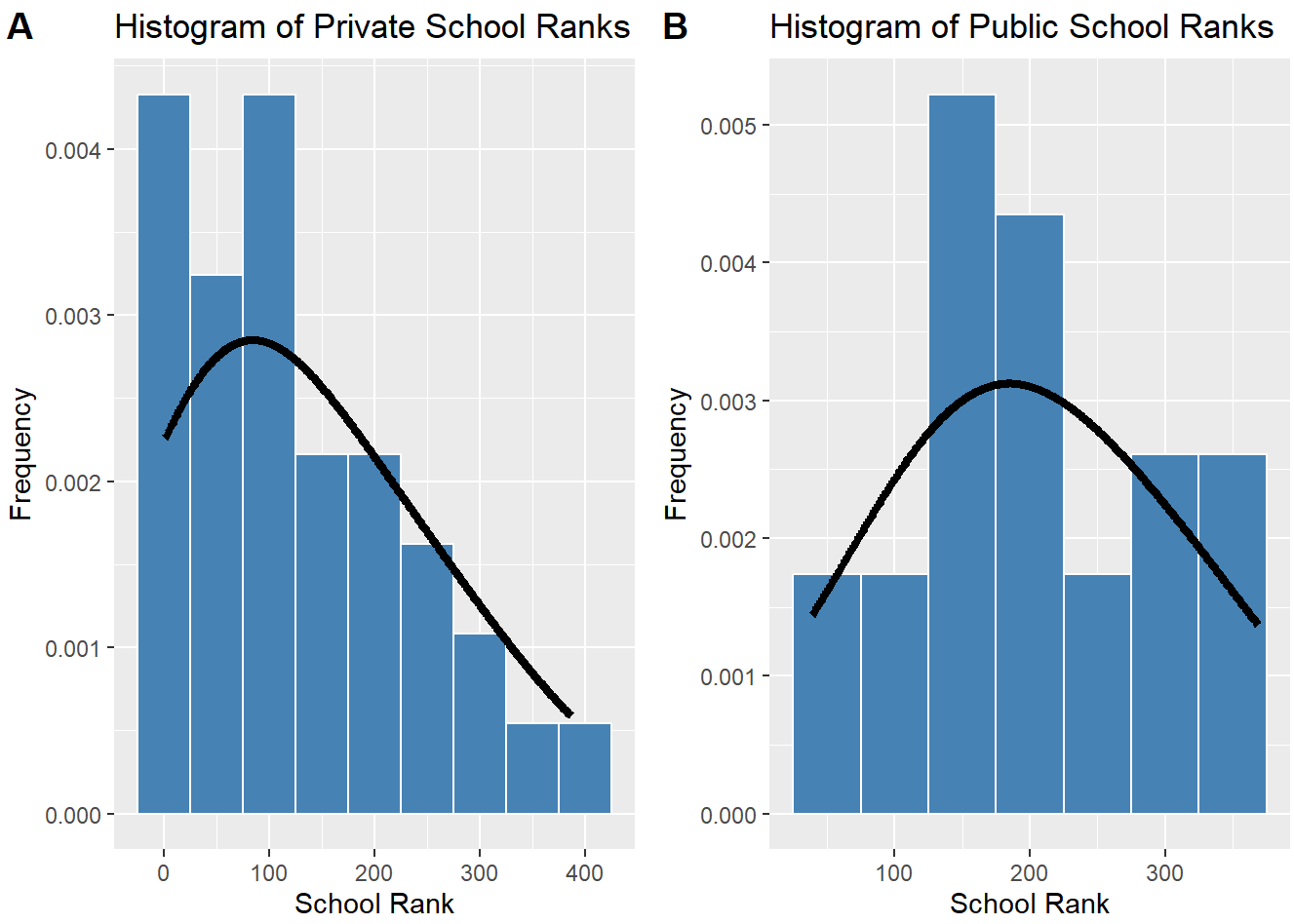
```
library(viridis)
library(dplyr)
library(corrplot)
library(ggplot2)
library(ggpubr)
library(kableExtra)
library(janitor)

# Load Data from previous section
obs_60_final<- read.csv('C:\\Users\\roryq\\Downloads\\Stat 1223\\obs_60_final.csv')

# Filter by private or public schools
Private_60 = obs_60_final[which(obs_60_final$institutionalControl == "private"),]
Private_60<- Private_60 %>% select(Tuition,Expend,Median_Income, number_Undergrads,Rank)
Public_60 = obs_60_final[which(obs_60_final$institutionalControl == "public"),]
Public_60<- Public_60 %>% select(Tuition,Expend,Median_Income, number_Undergrads,Rank)
```

Explanatory Variable Relationships/Distributions

```
A<-ggplot(Private_60, aes(x = Rank)) +
    geom_histogram(aes(y=..density..),color = 'white', fill='steelblue', binwidth = 50,) +
    labs(x = 'School Rank',  y = 'Frequency')+ geom_density(adjust=2, size= 1.5)+
  ggtitle("Histogram of Private School Ranks")
```

```
B<- ggplot(Public_60, aes(x = Rank)) +
    geom_histogram(aes(y=..density..),color = 'white', fill='steelblue', binwidth = 50,) +
    labs(x = 'School Rank',  y = 'Frequency')+ geom_density(adjust=2, size= 1.5)+
  ggtitle("Histogram of Public School Ranks")
```

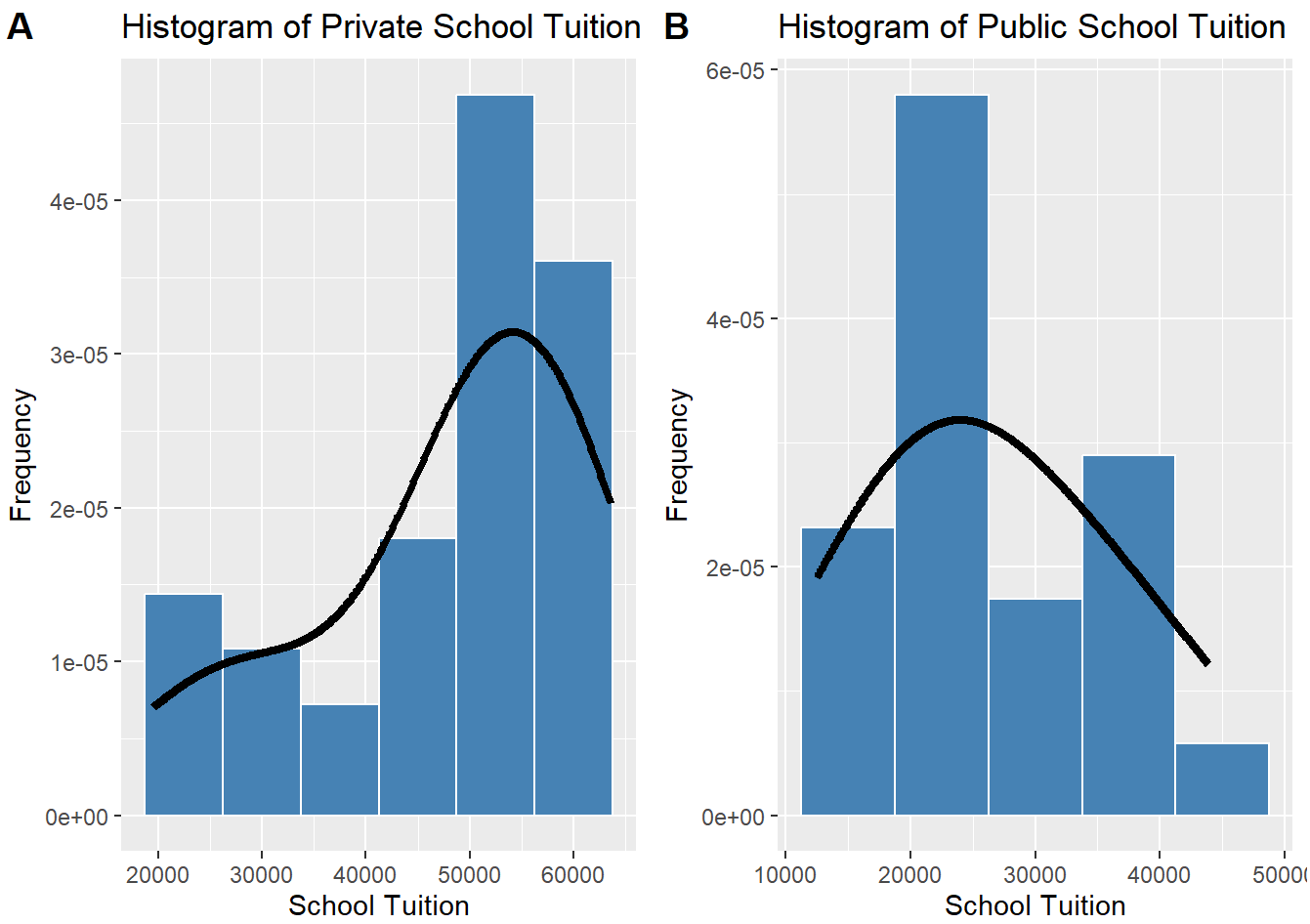
```
# Make plot side by side by control
ggarrange(A, B, labels = c("A", "B"))
```




```
A<-ggplot(Private_60, aes(x = Tuition)) +
    geom_histogram(aes(y=..density..),color = 'white', fill='steelblue', binwidth=7500) +
    labs(x = 'School Tuition',  y = 'Frequency')+ geom_density(adjust=2, size= 1.5)+
  ggtitle("Histogram of Private School Tuition")
```

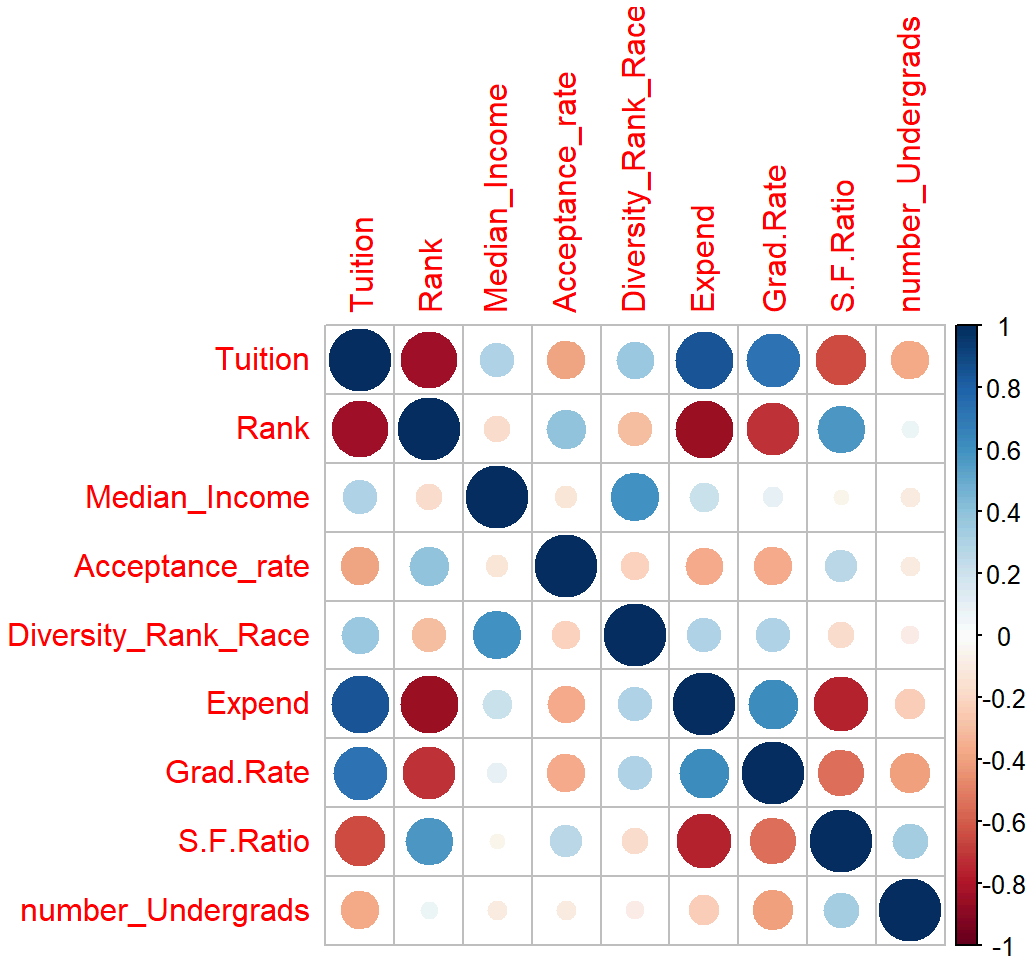
```
B<- ggplot(Public_60, aes(x = Tuition)) +
    geom_histogram(aes(y=..density..),color = 'white', fill='steelblue', binwidth=7500) +
    labs(x = 'School Tuition',  y = 'Frequency')+ geom_density(adjust=2, size= 1.5)+
  ggtitle("Histogram of Public School Tuition")
```

```
# Make plot side by side by control
ggarrange(A, B, labels = c("A", "B"))
```

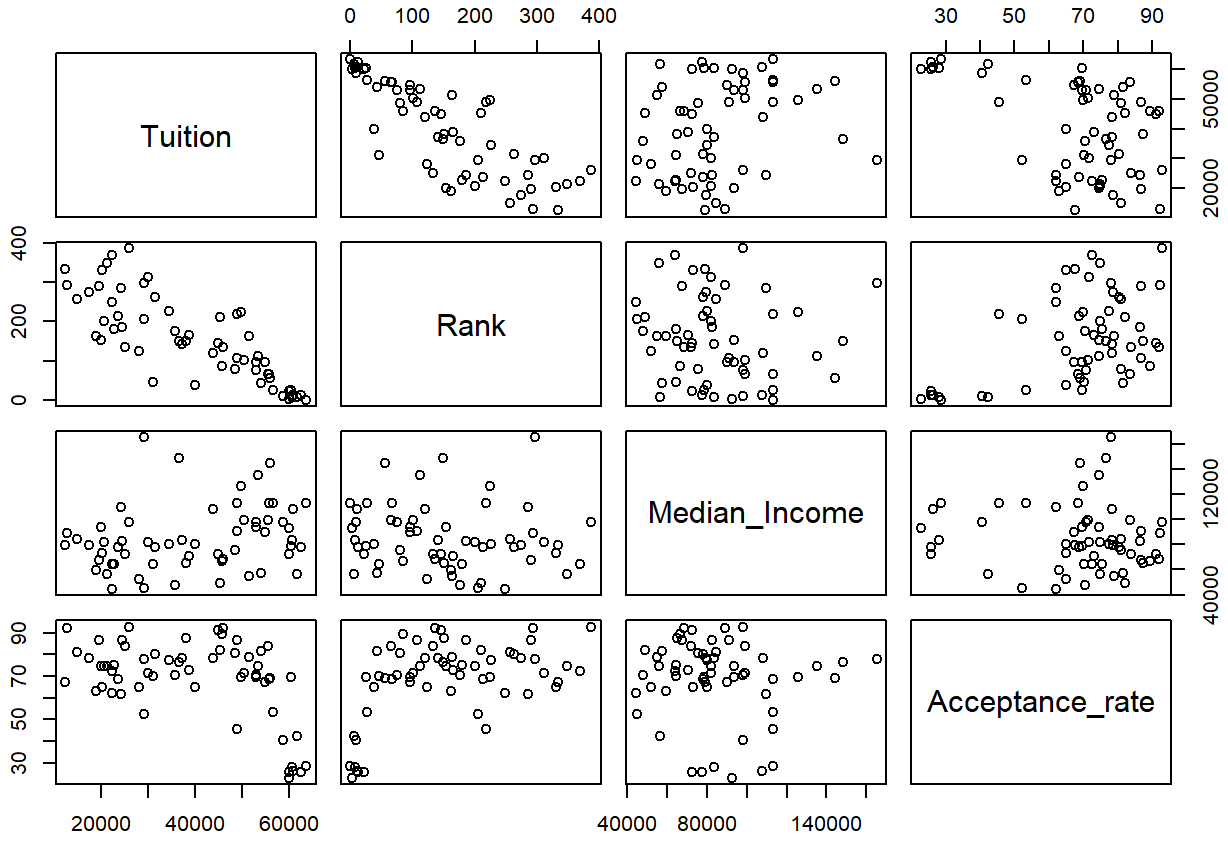



```
# Select factors that we are most interested in

# Create correlation matrix
key_predictors<- obs_60_final %>% select("Tuition","Rank", "Median_Income","Acceptance_rate","Di
versity_Rank_Race","Expend", "Grad.Rate", "S.F.Ratio", "number_Undergrads")
cor<- cor(key_predictors,method = c("spearman"))
```

```
# Create correlogram to visualize connections
corrplot(cor)
```




```
# Narrow down some particularly interesting factos and look closer
key_predictors_2<- obs_60_final %>% select("Tuition","Rank", "Median_Income","Acceptance_rate")
plot(key_predictors_2)
```



Descriptive Statistics

`

```
# Create empty vectors to fill
Mean <- numeric(ncol(Private_60))
Standard_Deviation<-numeric(ncol(Private_60))

# Loop through each column to compute the mean
for (i in 1:ncol(Private_60)) {
  column_data <- Private_60[[i]]        # Extract column data
  Mean[i] <- round(mean(column_data), digits = 0)  # Calculate mean and store in vector
  Standard_Deviation[i]<- round(sd(column_data),digits = 0)
}

# Set names of the vector to column names for clarity
names(Mean) <- names(Private_60)
names(Standard_Deviation)<-names(Private_60)

# Print the mean values vector
ds<-data.frame(Mean,Standard_Deviation)
ds %>%
  kbl(caption= "Descriptive Statistics for Private Schools") %>%
  kable_styling()
```

Descriptive Statistics for Private Schools

	Mean	Standard_Deviation
Tuition	47880	12794
Expend	15013	10726
Median_Income	85944	24297
number_Undergrads	5227	3265
Rank	126	106

```
# Same as above except with public schools
Mean <- numeric(ncol(Public_60))
Standard_Deviation<-numeric(ncol(Public_60))
# Loop through each column to compute the mean
for (i in 1:ncol(Public_60)) {
  column_data <- Public_60[[i]]        # Extract column data
  Mean[i] <- round(mean(column_data), digits = 0)  # Calculate mean and store in vector
  Standard_Deviation[i]<- round(sd(column_data),digits = 0)
}

# Set names of the vector to column names for clarity
names(Mean) <- names(Public_60)
names(Standard_Deviation)<-names(Public_60)

# Print the mean values vector
ds<-data.frame(Mean,Standard_Deviation)
ds %>%
  kbl(caption= "Descriptive Statistics for Public Schools") %>%
  kable_styling()
```

Descriptive Statistics for Public Schools

	Mean	Standard_Deviation
Tuition	26163	8922
Expend	7653	2151
Median_Income	83064	28239
number_Undergrads	15494	6470
Rank	202	91

```
# Number of observations of private schools
nrow(Private_60)
```

```
## [1] 37
```

```
# Number of observations of public schools
nrow(Public_60)
```

```
## [1] 23
```