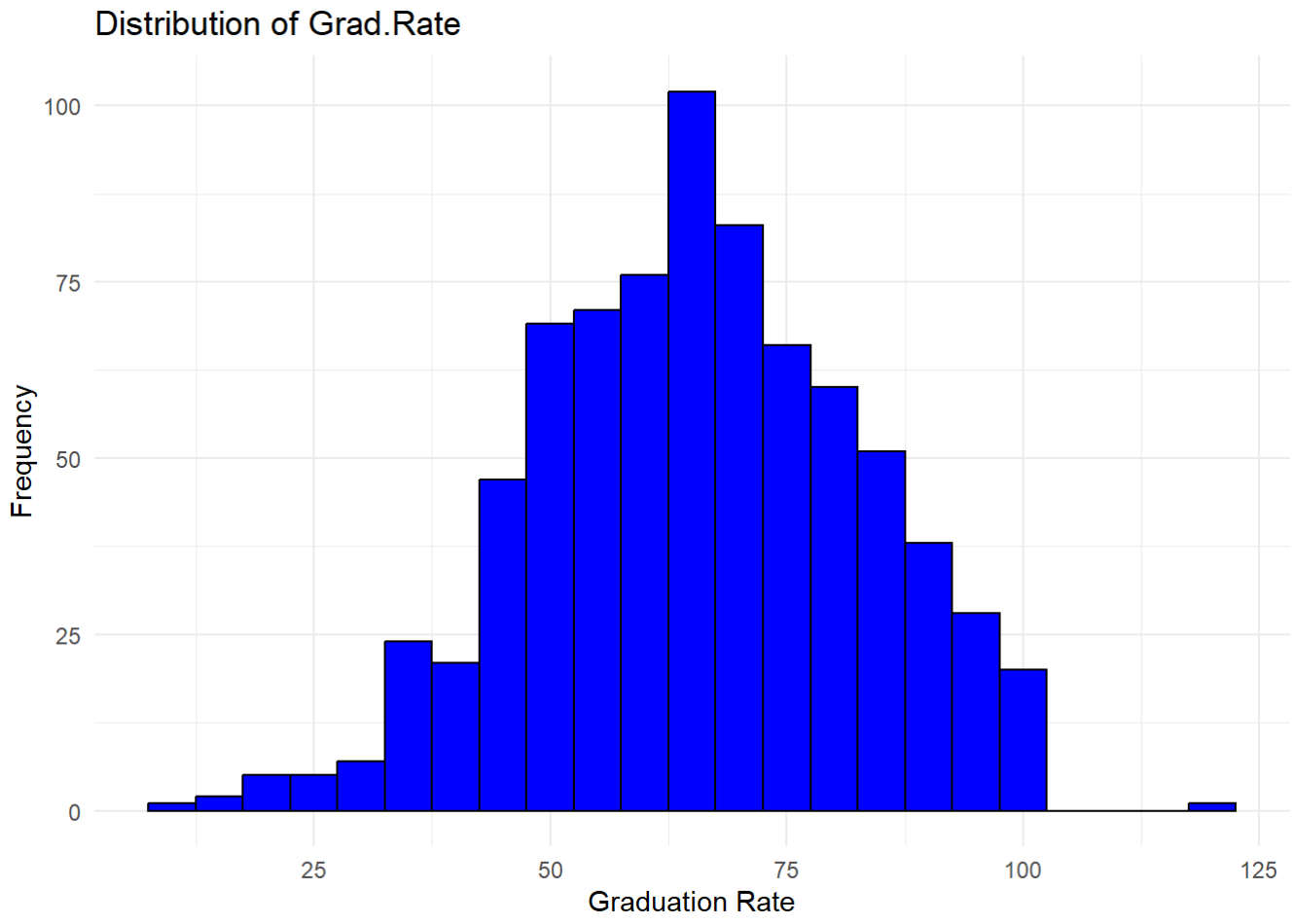
Assignment 3 problem 1

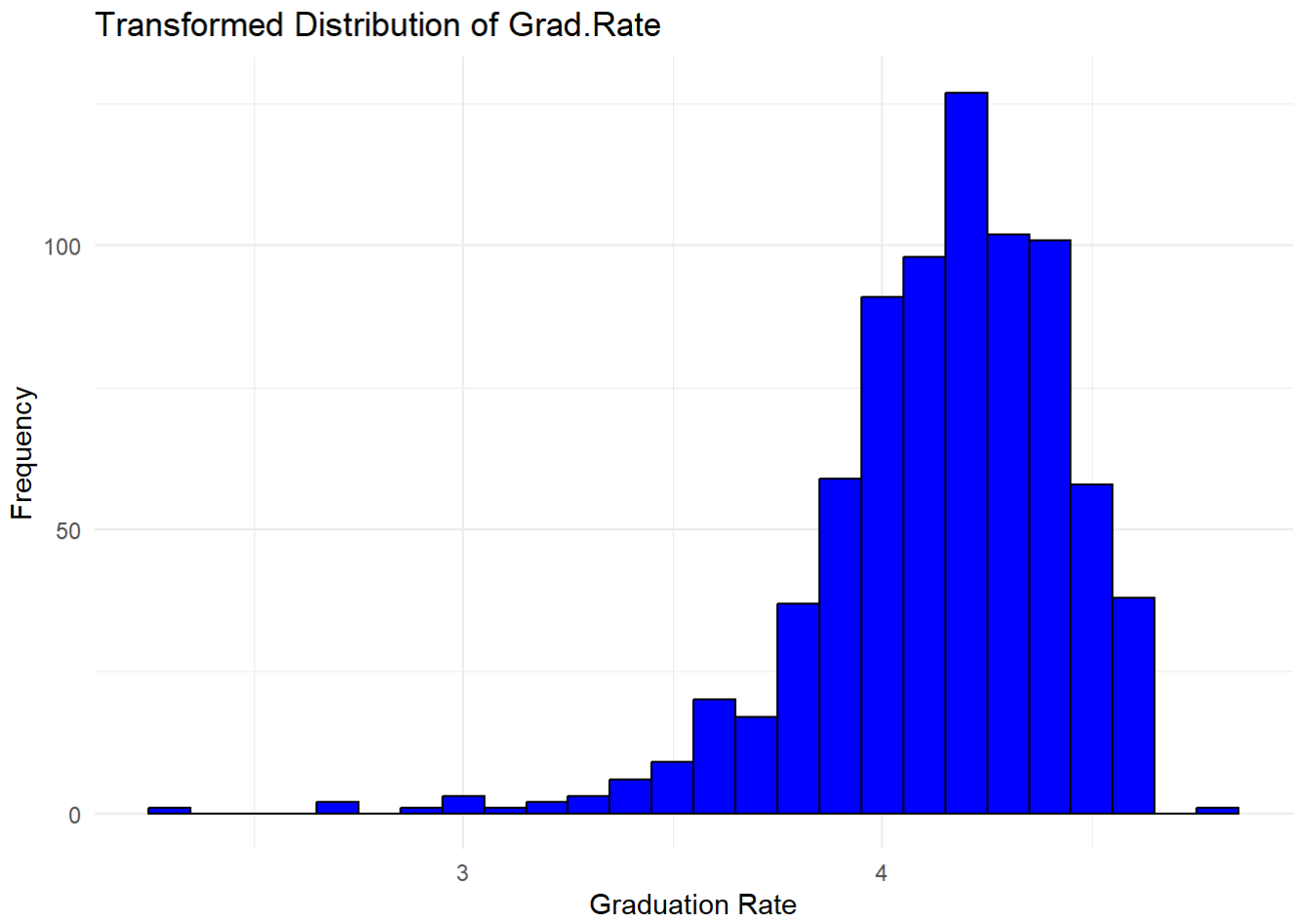
SRK Yarra

2023-10-25

```
#Problem1(a)
# Read the data
data <- read.csv("college.csv")

#a) Analyze distribution of Grad.Rate
# Load necessary libraries
library(ggplot2)
# Plot the distribution of Grad.Rate
ggplot(data, aes(x=Grad.Rate)) + geom_histogram(binwidth=5, fill="blue", color="black") +
  labs(title="Distribution of Grad.Rate", x="Graduation Rate", y="Frequency") +
  theme_minimal()
```



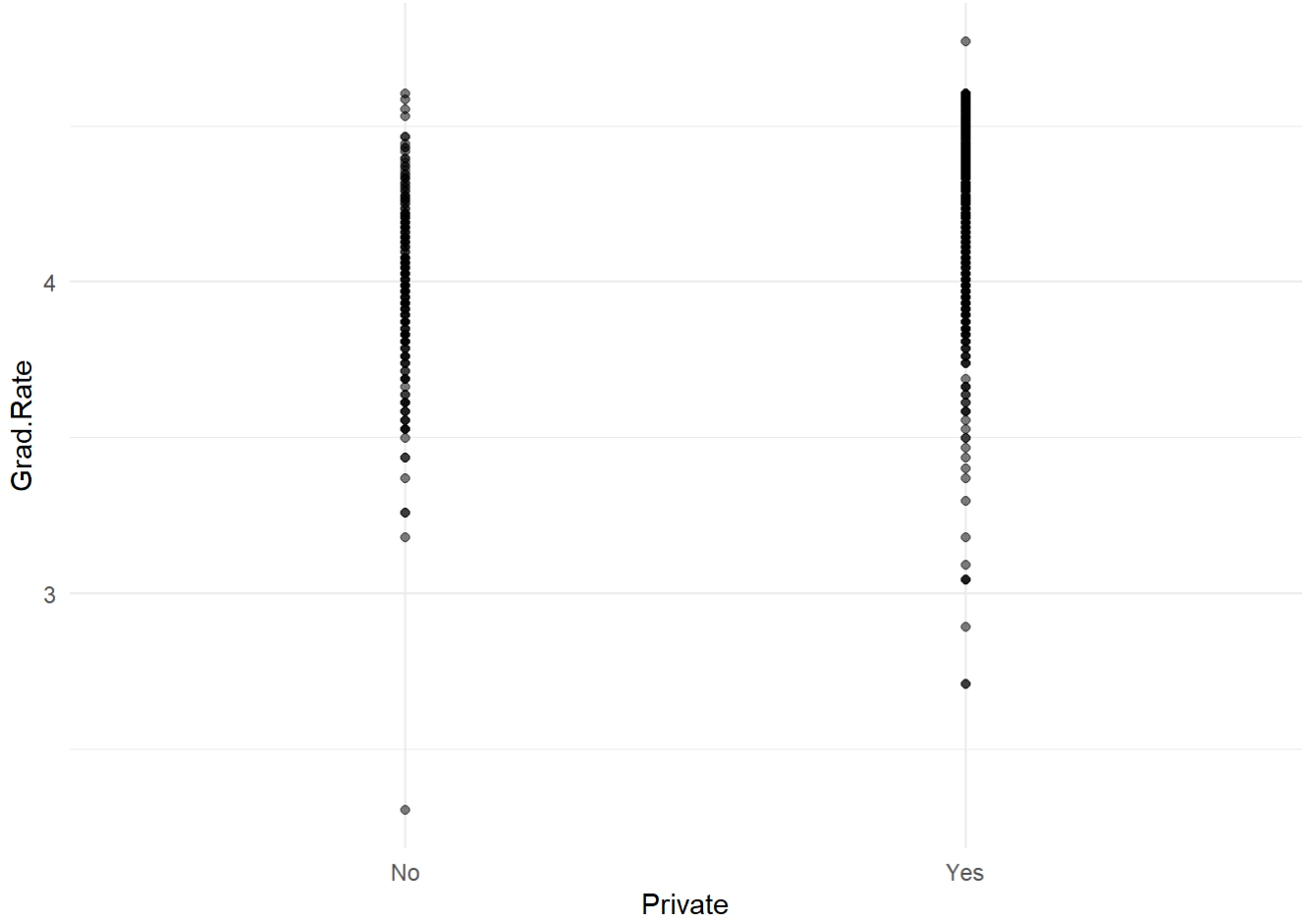

```
data$Grad.Rate<-log(data$Grad.Rate)

ggplot(data, aes(x=Grad.Rate)) + geom_histogram(binwidth=0.1, fill="blue", color="black") +
  labs(title="Transformed Distribution of Grad.Rate", x="Graduation Rate", y="Frequency") +
  theme_minimal()
```

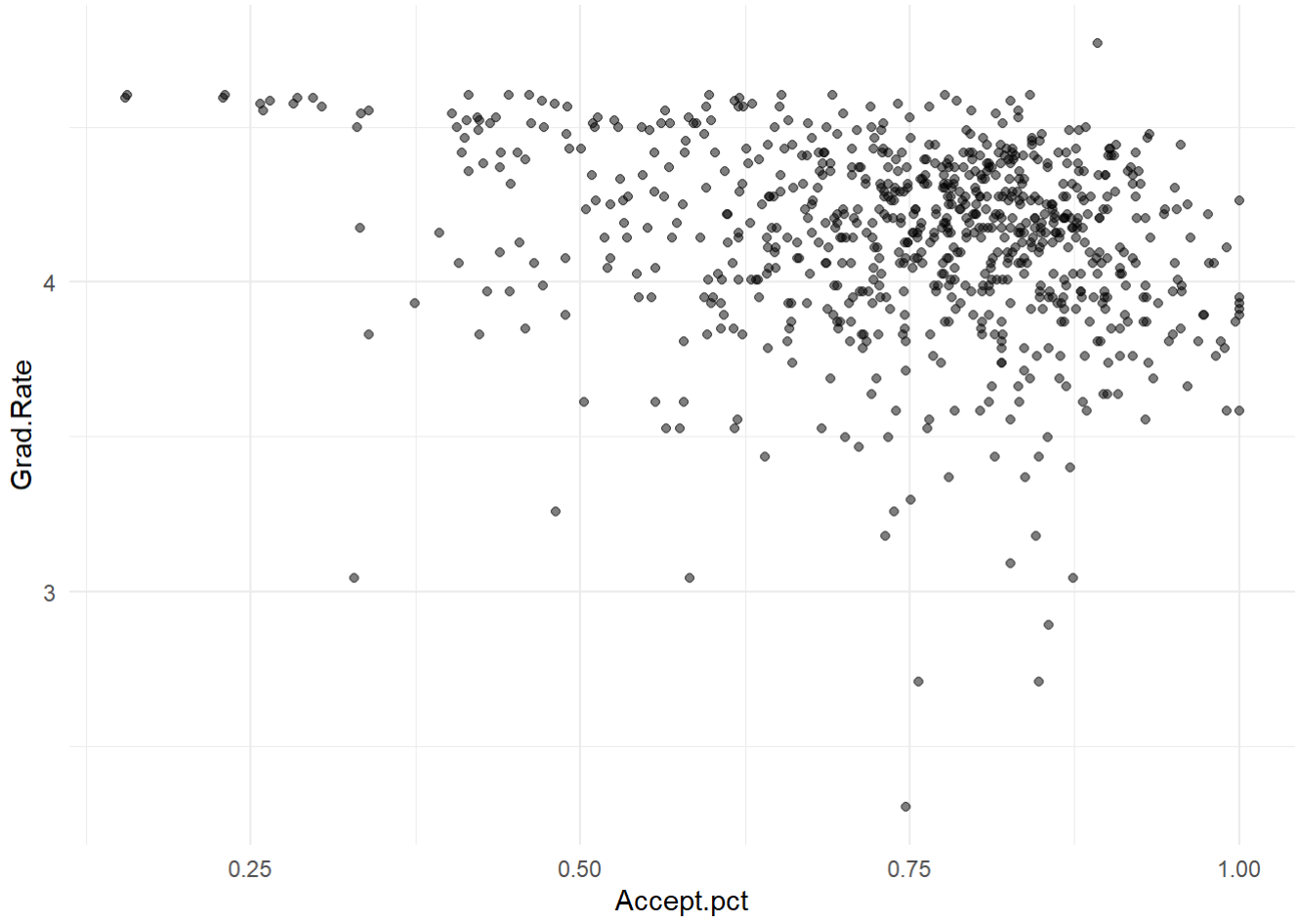


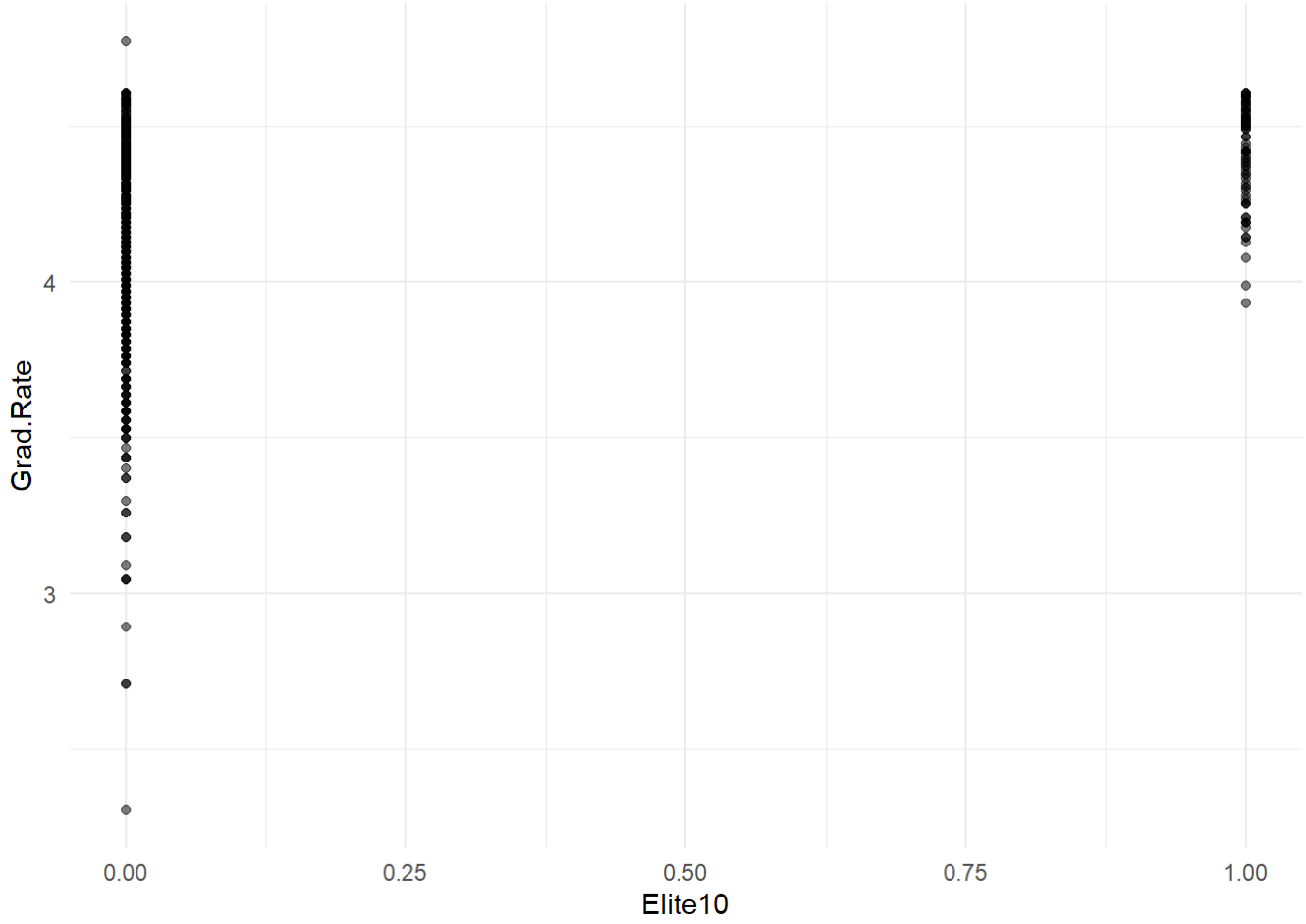
#Problem1(b)

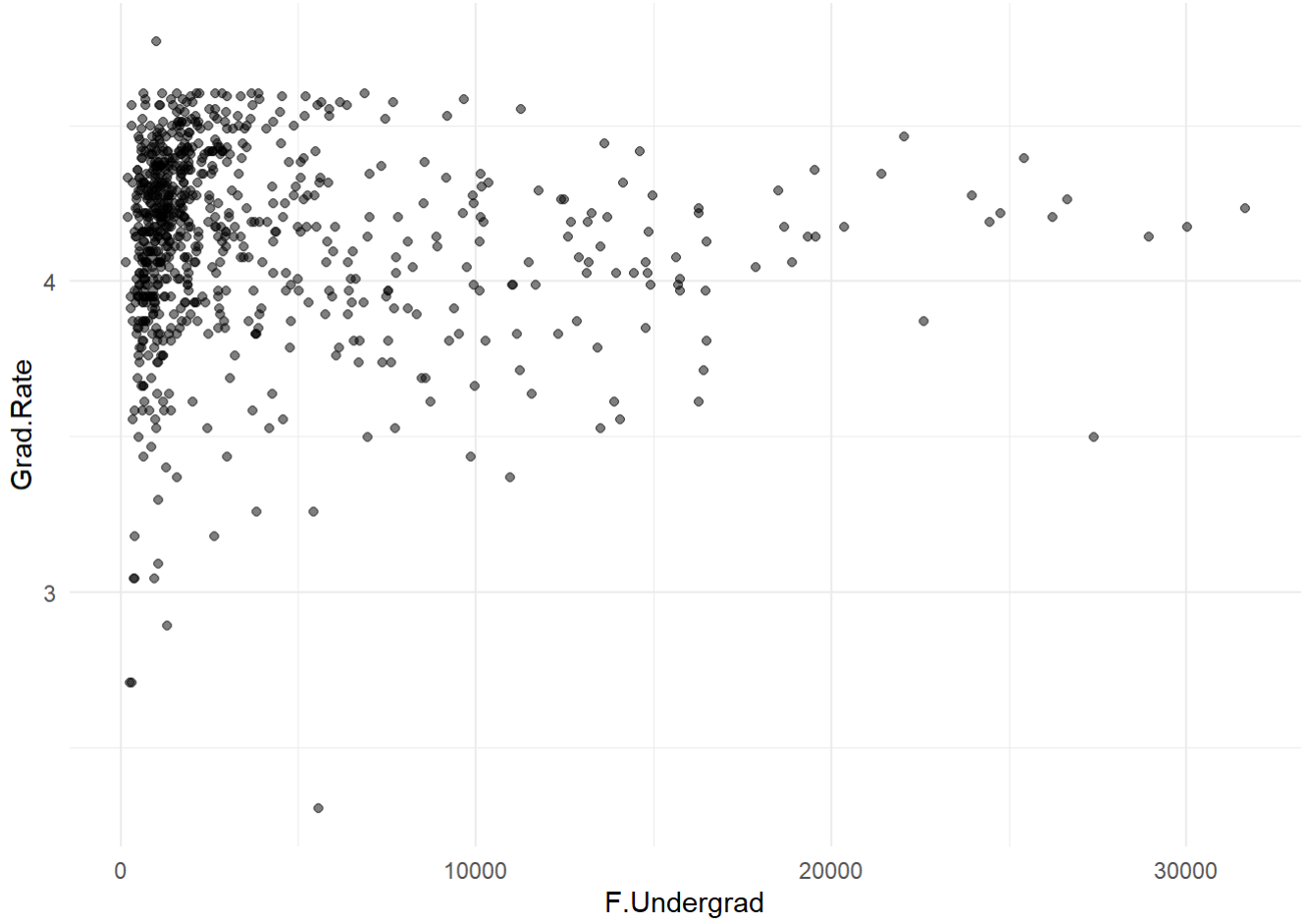
```
#b) Scatterplots for Grad.Rate vs each independent variable
# Creating scatterplots for each variable
independent_vars <- names(data)[2:(ncol(data)-1)] # Exclude 'school' and 'Grad.Rate'
for (var in independent_vars) {
  print(ggplot(data, aes_string(x=var, y="Grad.Rate")) + geom_point(alpha=0.5) + theme_minima
l())
}
```

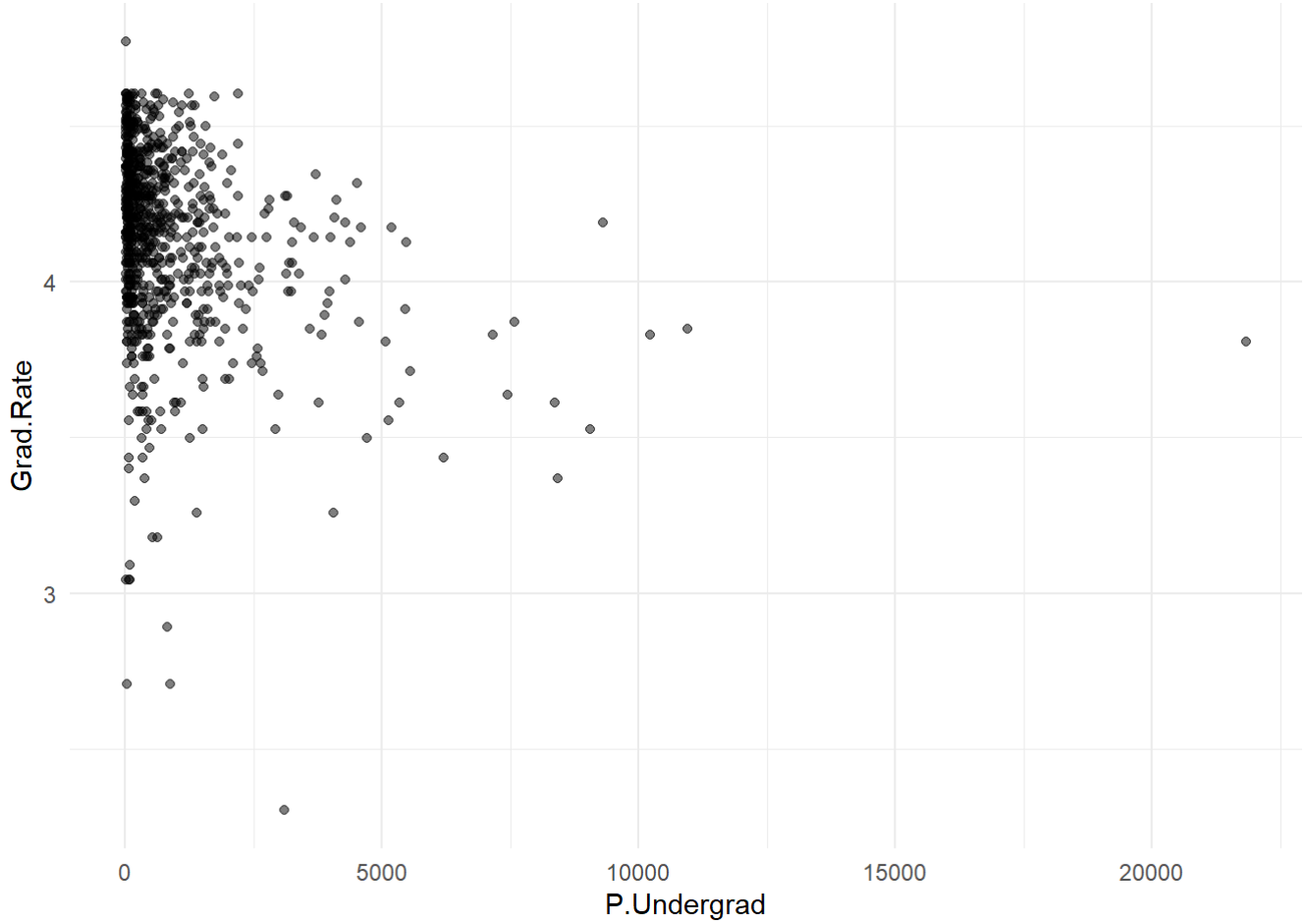
```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## ℹ Please use tidy evaluation idioms with `aes()`.
## ℹ See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

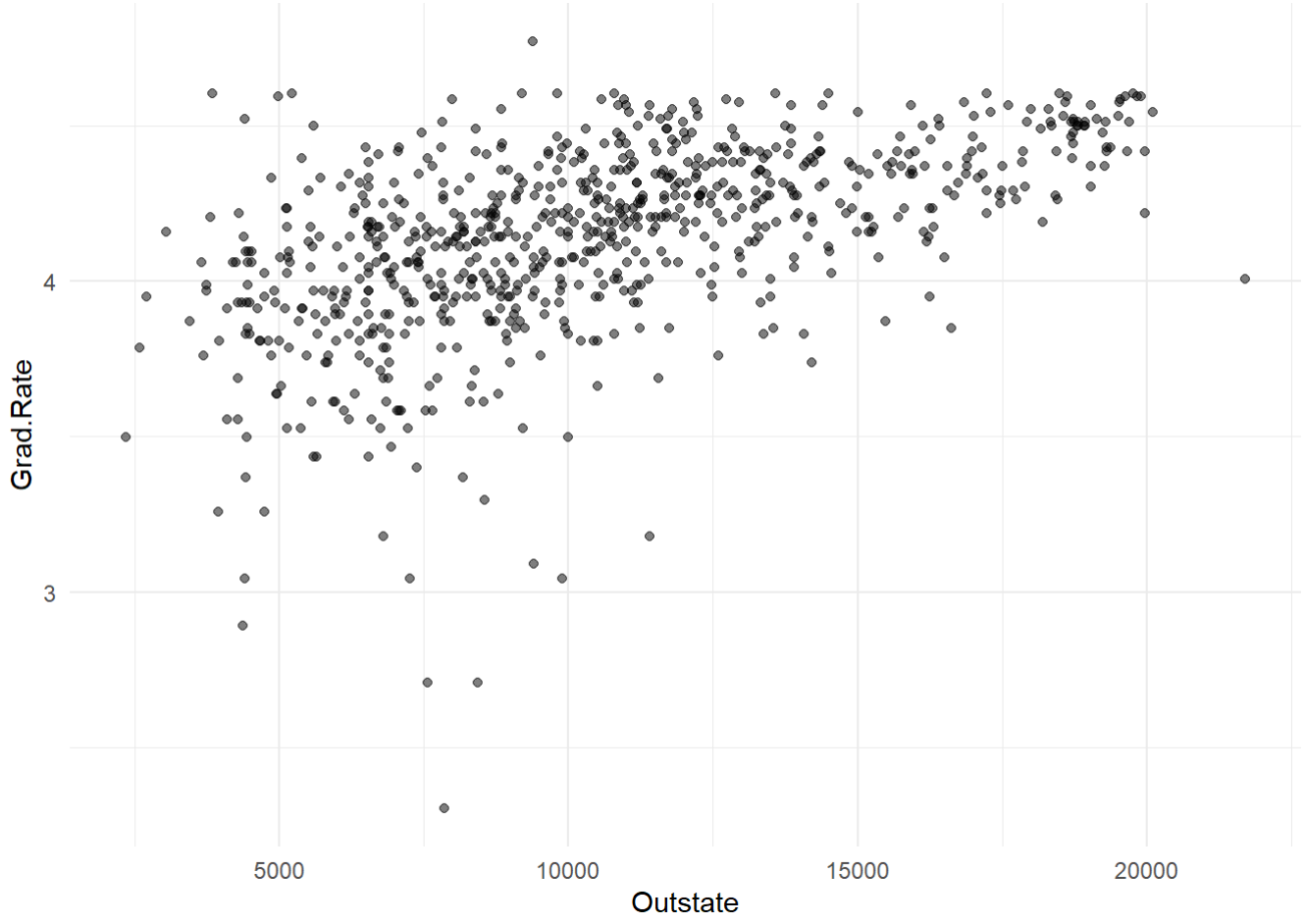

Grad.Rate
4
3
No
Yes
Private

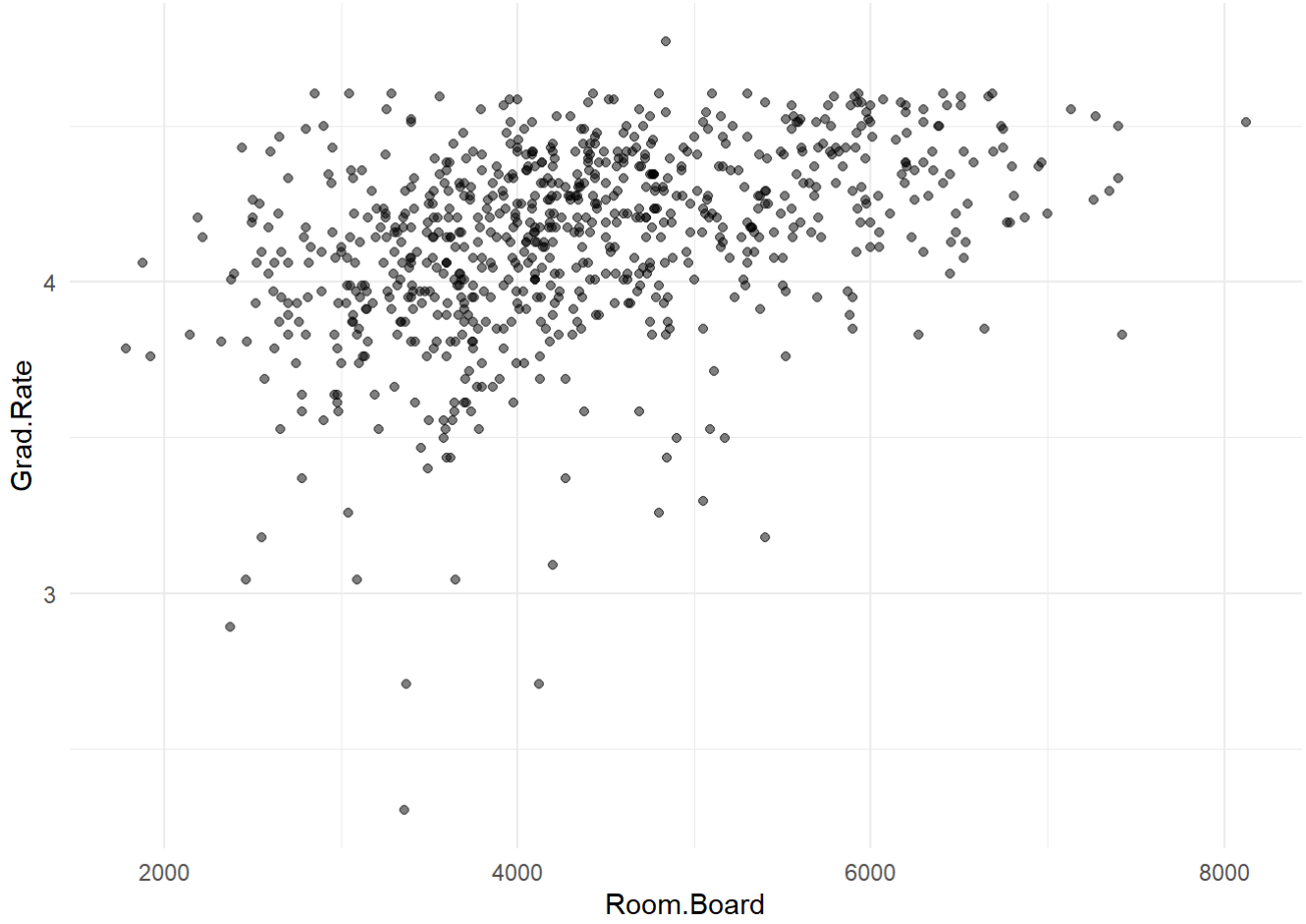

Grad.Rate
4
3
0.25
0.50
0.75
1.00
Accept.pct


Grad.Rate
4
3
0.00
0.25
0.50
0.75
1.00
Elite10

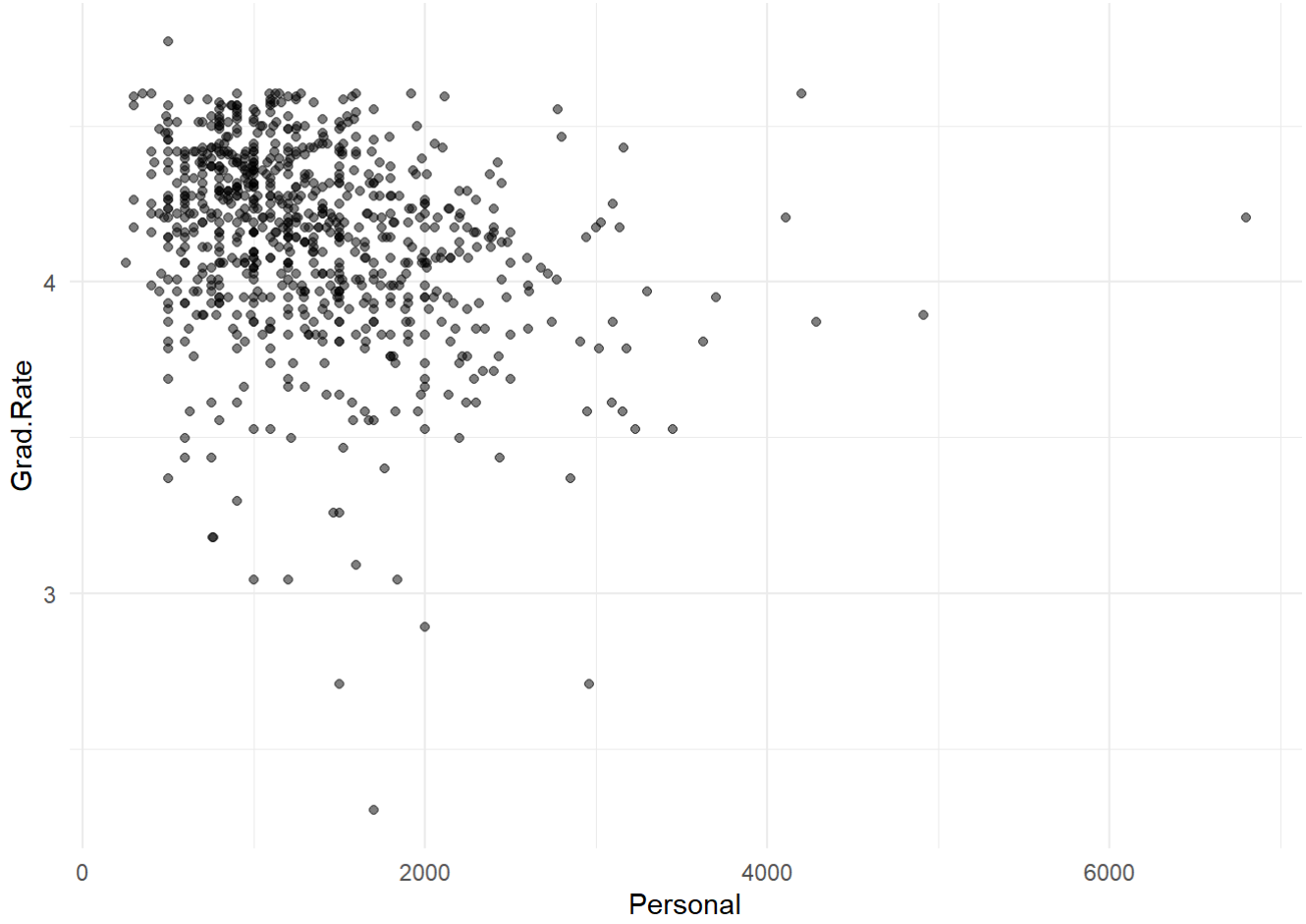

Grad.Rate
4
3
0
10000
20000
30000
F.Undergrad

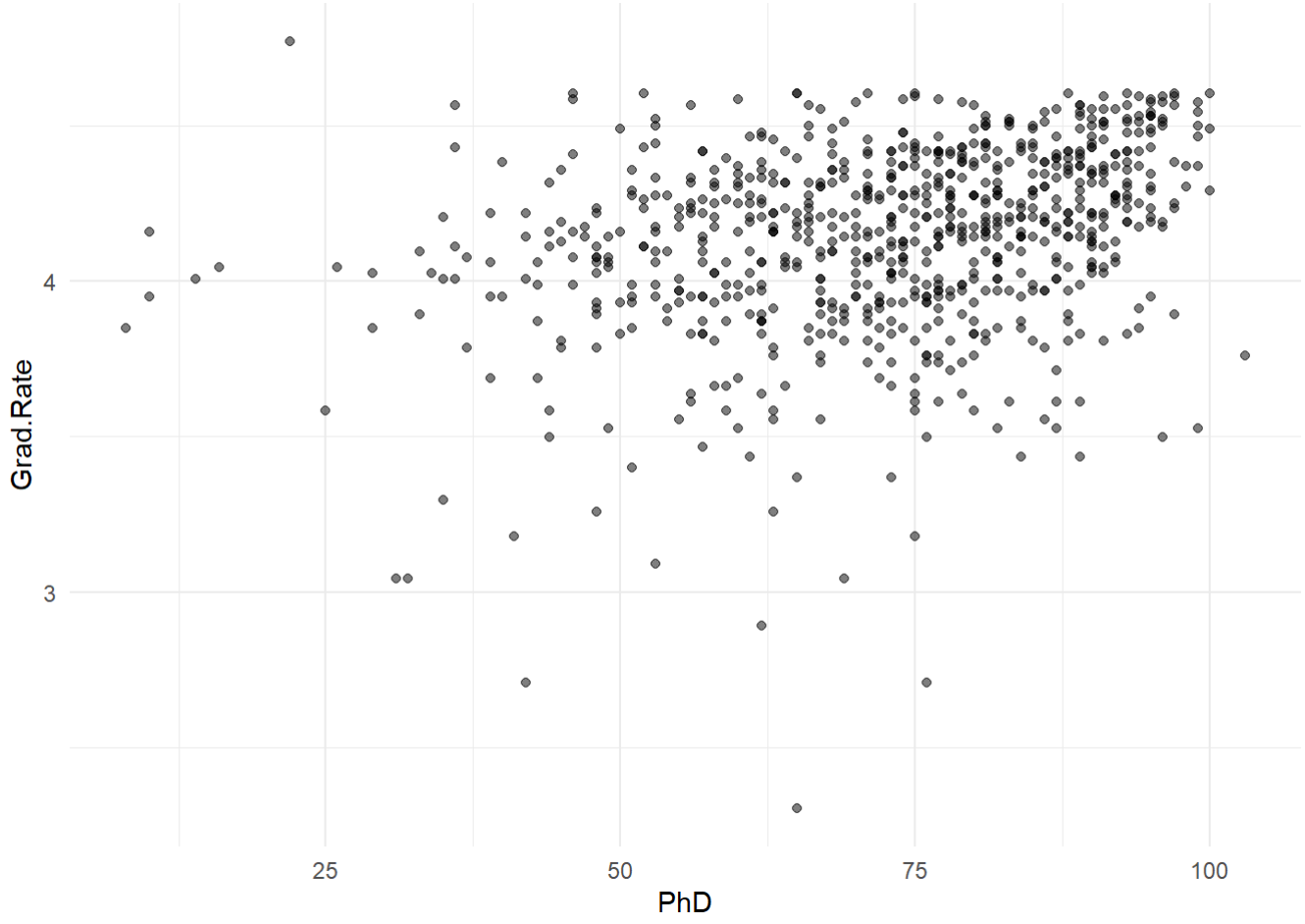

Grad.Rate
4
3
0
5000
10000
15000
20000
P.Undergrad

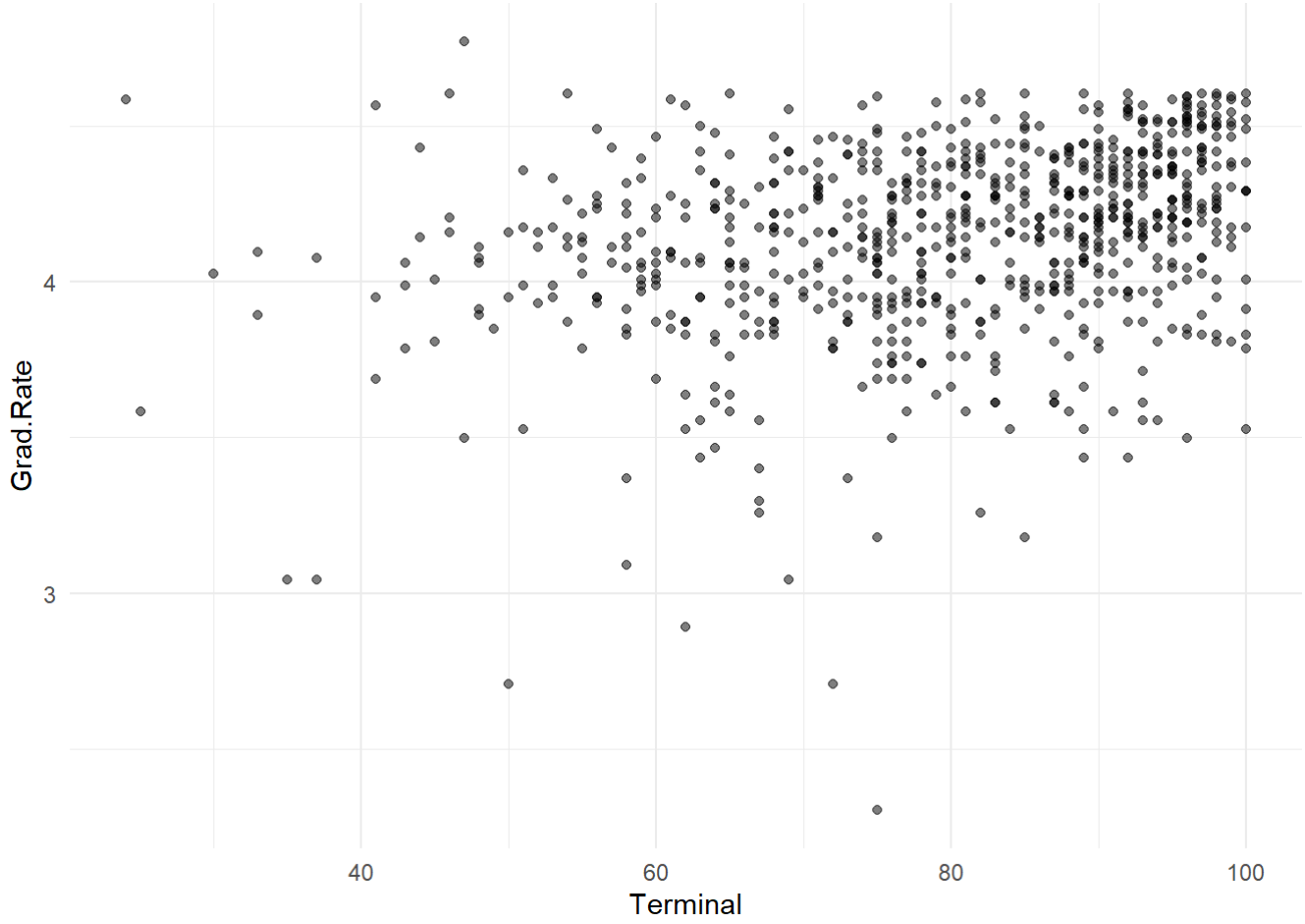

Grad.Rate
4
3
5000
10000
15000
20000
Outstate

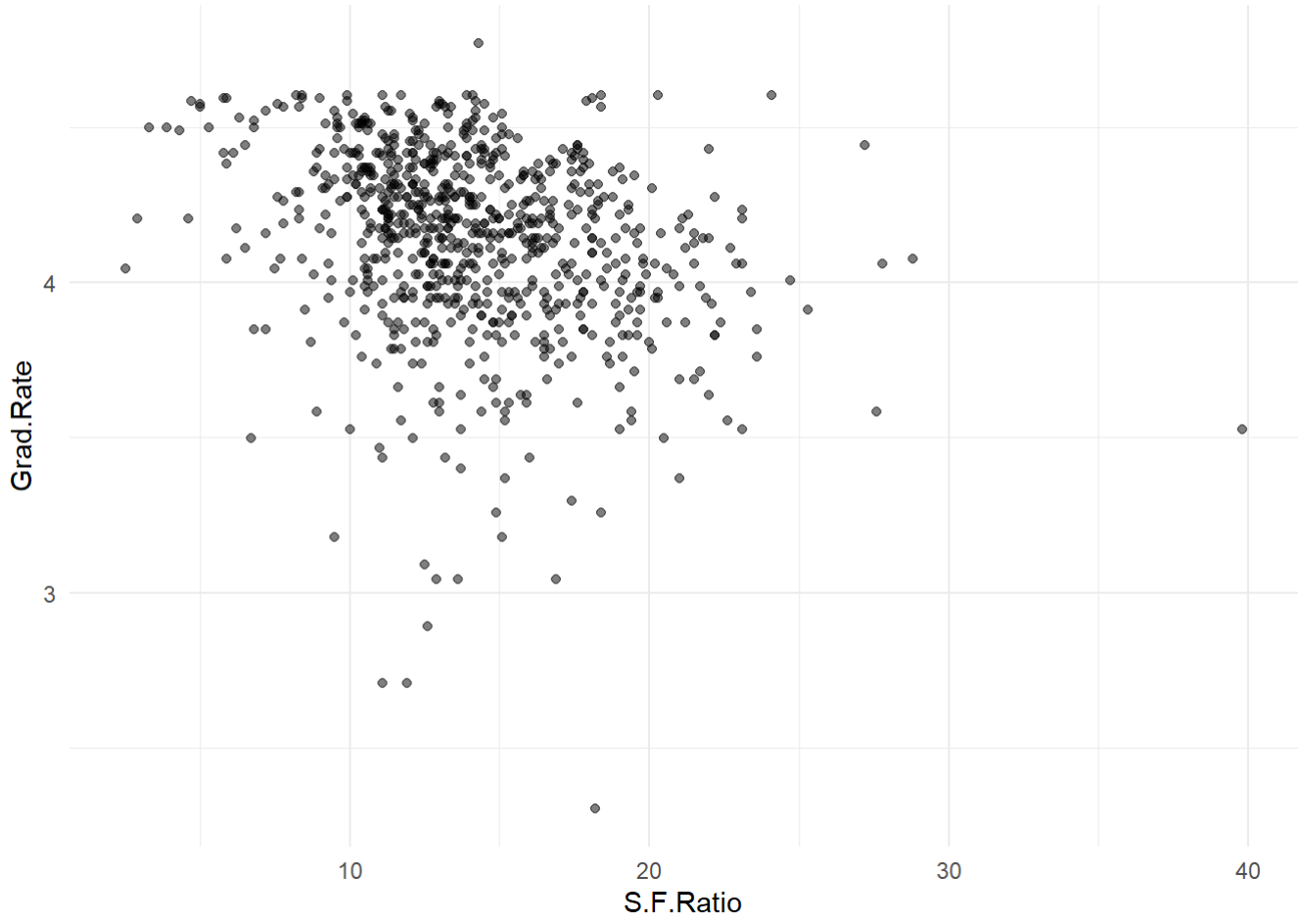

Grad.Rate
4
3
2000
4000
6000
8000
Room.Board

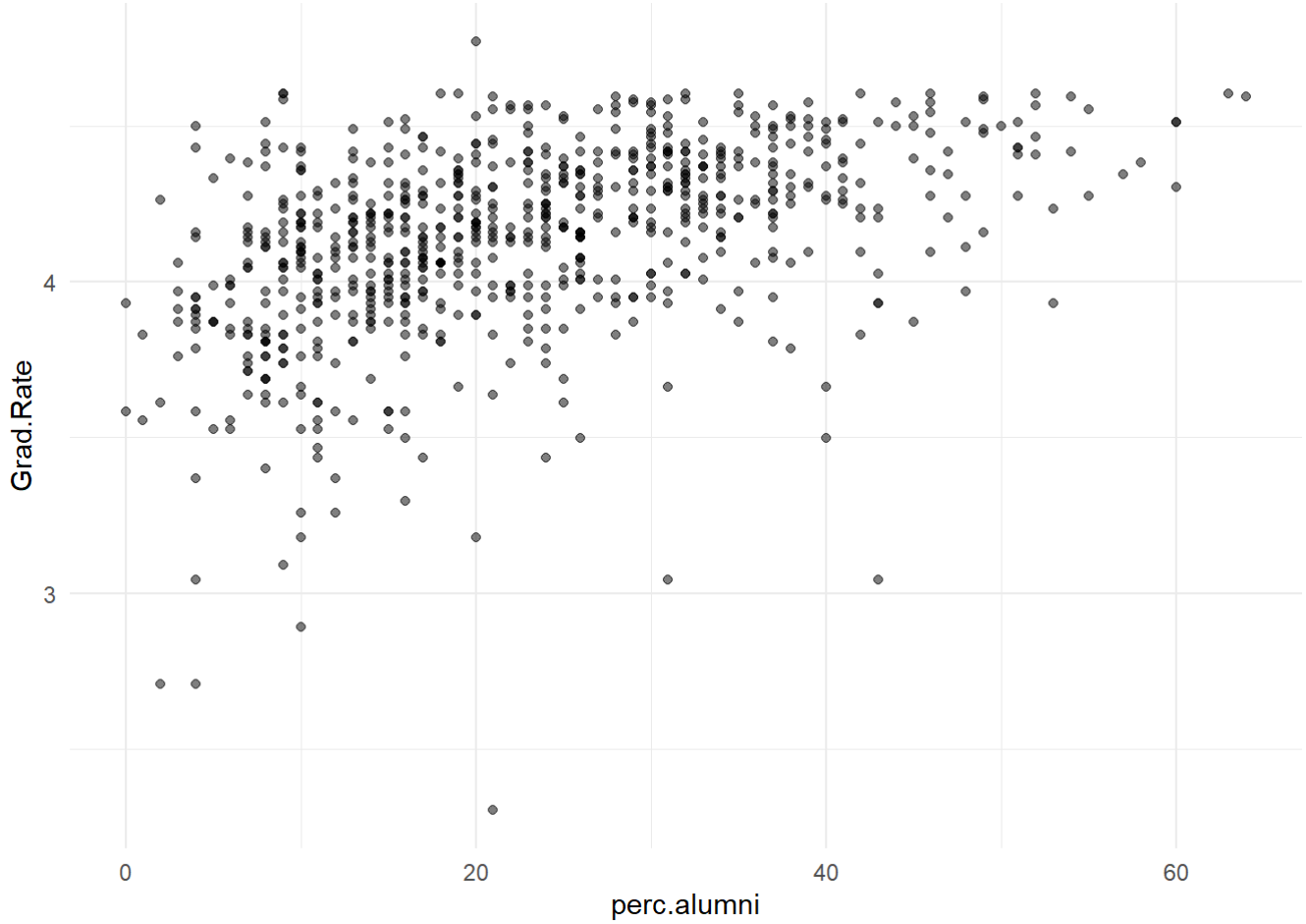

Grad.Rate
4
3
0
500
1000
1500
2000
Books

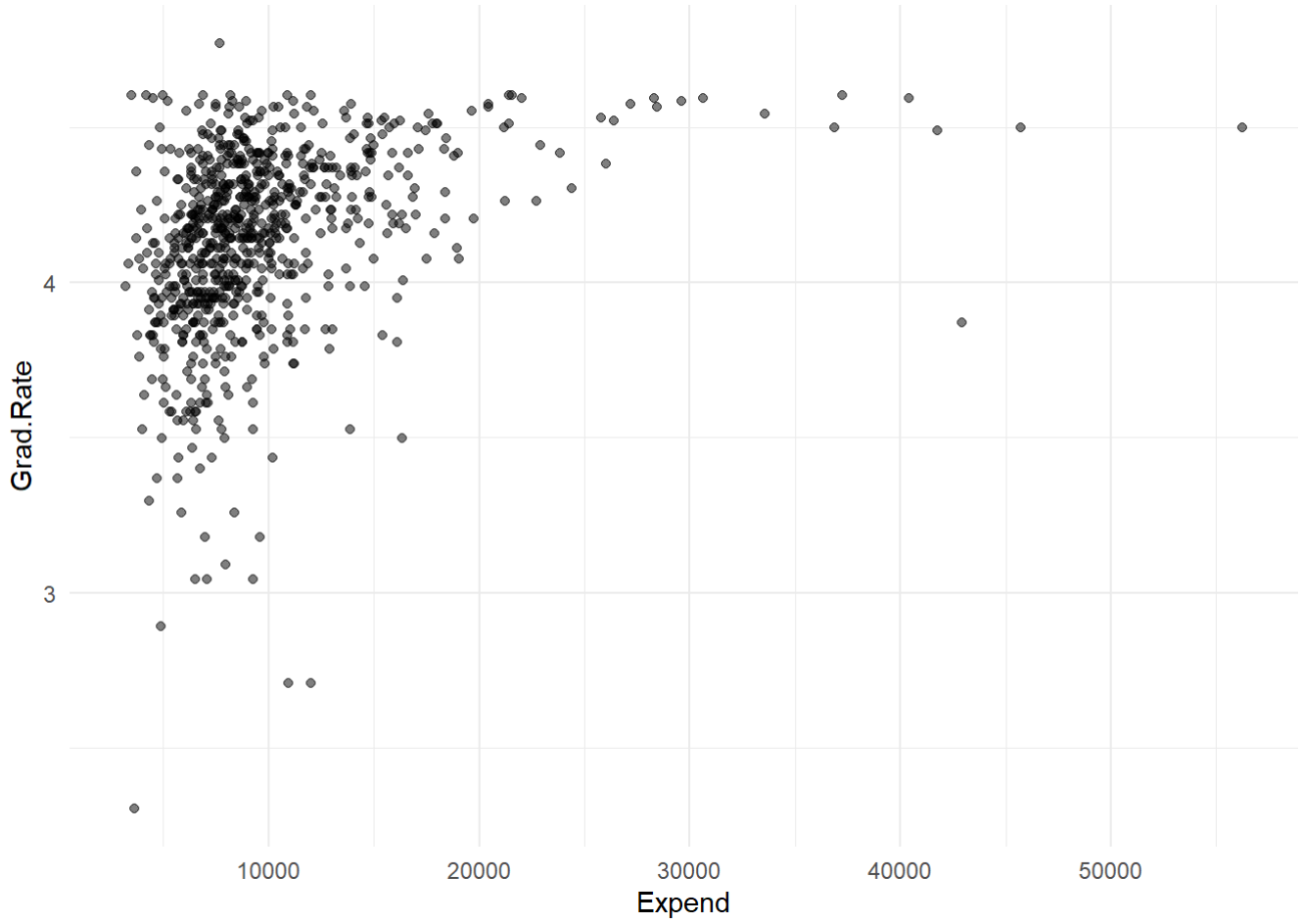

Grad.Rate
4
3
0
2000
4000
6000
Personal


Grad.Rate
4
3
25
50
75
100
PhD


Grad.Rate
4
3
40
60
80
100
Terminal


Grad.Rate
4
3
10
20
30
40
S.F.Ratio


Grad.Rate
4
3
0
20
40
60
perc.alumni

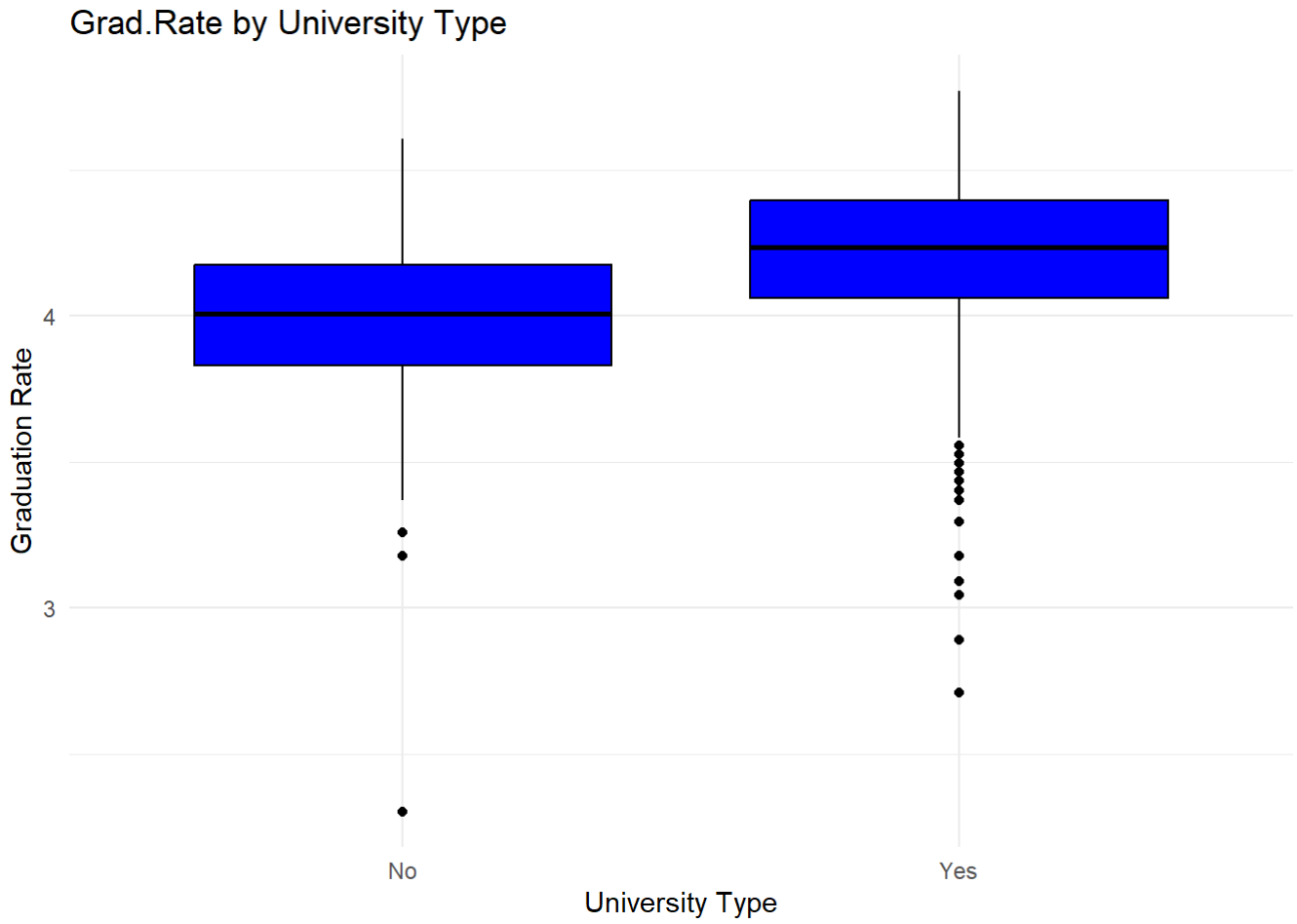

Grad.Rate
4
3
10000
20000
30000
40000
50000
Expend

```
# Correlation analysis

correlation_matrix<-cor(data[,-c(1,2)])
correlation_matrix
```

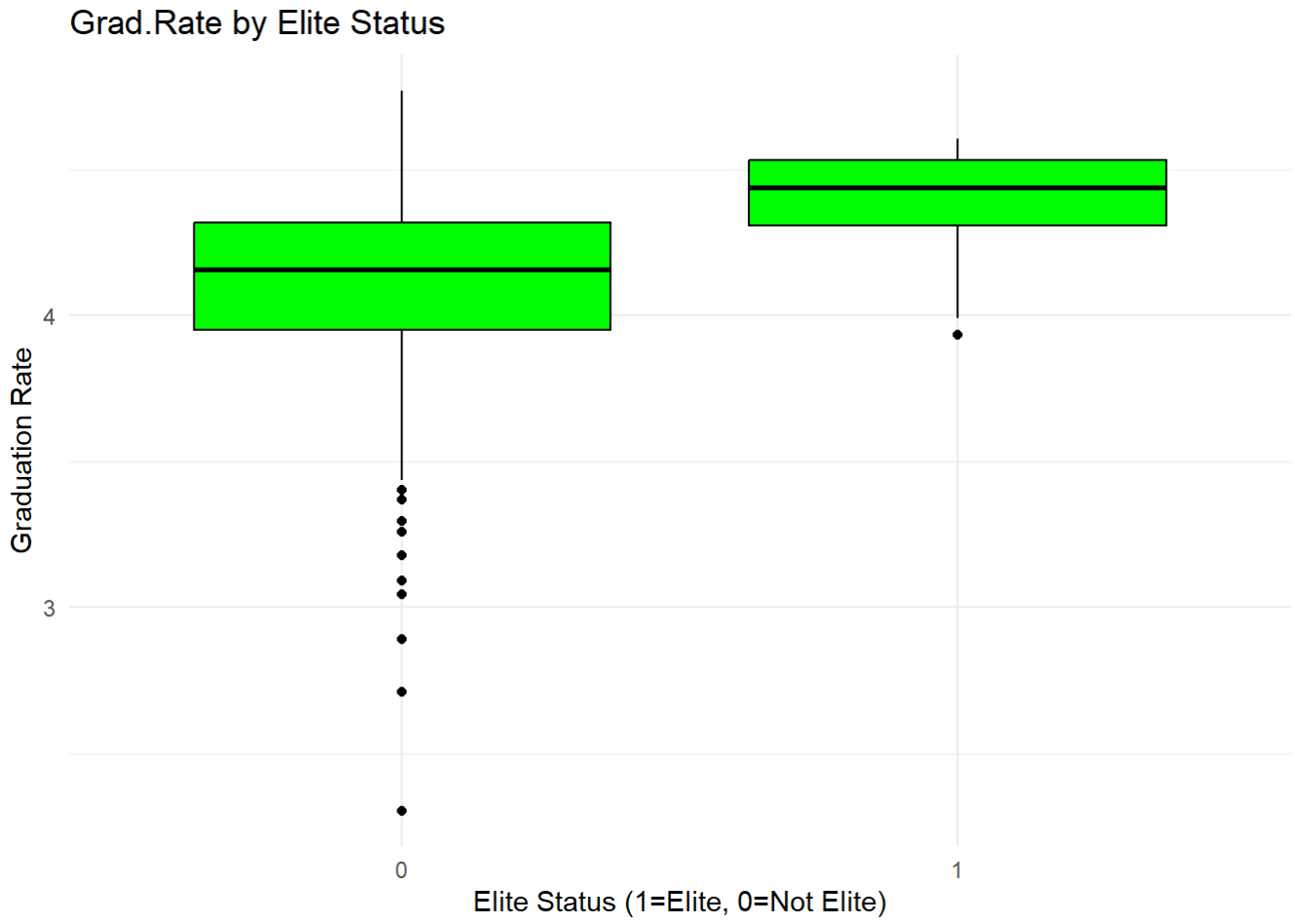
```
##              Accept.pct     Elite10 F.Undergrad P.Undergrad    Outstate
## Accept.pct    1.00000000 -0.46245330 -0.15565379 -0.09228664 -0.24095073
## Elite10      -0.46245330  1.00000000  0.06083999 -0.11644570  0.39947675
## F.Undergrad  -0.15565379  0.06083999  1.00000000  0.57051219 -0.21574200
## P.Undergrad  -0.09228664 -0.11644570  0.57051219  1.00000000 -0.25351232
## Outstate     -0.24095073  0.39947675 -0.21574200 -0.25351232  1.00000000
## Room.Board   -0.31030204  0.29847208 -0.06889039 -0.06132551  0.65425640
## Books        -0.17407288  0.09217607  0.11554976  0.08119952  0.03885487
## Personal      0.01997851 -0.07526924  0.31719954  0.31988162 -0.29908690
## PhD          -0.31833394  0.34106219  0.31833697  0.14911422  0.38298241
## Terminal     -0.30379999  0.32664984  0.30001894  0.14190357  0.40798320
## S.F.Ratio     0.10998188 -0.29349738  0.27970335  0.23253051 -0.55482128
## perc.alumni  -0.13210402  0.30259090 -0.22946222 -0.28079236  0.56626242
## Expend       -0.40862232  0.55977784  0.01865162 -0.08356842  0.67277862
## Grad.Rate    -0.22948475  0.30150616 -0.06248495 -0.25004967  0.53324387
##               Room.Board        Books    Personal         PhD    Terminal
## Accept.pct   -0.31030204 -0.174072883  0.01997851 -0.31833394 -0.30379999
## Elite10       0.29847208  0.092176073 -0.07526924  0.34106219  0.32664984
## F.Undergrad -0.06889039  0.115549761  0.31719954  0.31833697  0.30001894
## P.Undergrad -0.06132551  0.081199521  0.31988162  0.14911422  0.14190357
## Outstate      0.65425640  0.038854868 -0.29908690  0.38298241  0.40798320
## Room.Board    1.00000000  0.127962970 -0.19942818  0.32920228  0.37453955
## Books         0.12796297  1.000000000  0.17929476  0.02690573  0.09995470
## Personal     -0.19942818  0.179294764  1.00000000 -0.01093579 -0.03061311
## PhD           0.32920228  0.026905731 -0.01093579  1.00000000  0.84958703
## Terminal      0.37453955  0.099954700 -0.03061311  0.84958703  1.00000000
## S.F.Ratio    -0.36262774 -0.031929274  0.13634483 -0.13053011 -0.16010395
## perc.alumni   0.27236345 -0.040207736 -0.28596808  0.24900866  0.26713029
## Expend        0.50173942  0.112409075 -0.09789189  0.43276168  0.43879922
## Grad.Rate     0.39867414 -0.006404662 -0.25855499  0.29453029  0.28611348
##                S.F.Ratio perc.alumni      Expend    Grad.Rate
## Accept.pct    0.10998188 -0.13210402 -0.40862232 -0.229484755
## Elite10      -0.29349738  0.30259090  0.55977784  0.301506162
## F.Undergrad   0.27970335 -0.22946222  0.01865162 -0.062484953
## P.Undergrad   0.23253051 -0.28079236 -0.08356842 -0.250049668
## Outstate     -0.55482128  0.56626242  0.67277862  0.533243871
## Room.Board   -0.36262774  0.27236345  0.50173942  0.398674142
## Books        -0.03192927 -0.04020774  0.11240908 -0.006404662
## Personal      0.13634483 -0.28596808 -0.09789189 -0.258554989
## PhD          -0.13053011  0.24900866  0.43276168  0.294530290
## Terminal     -0.16010395  0.26713029  0.43879922  0.286113484
## S.F.Ratio     1.00000000 -0.40292917 -0.58383204 -0.274912522
## perc.alumni  -0.40292917  1.00000000  0.41771172  0.457131186
## Expend       -0.58383204  0.41771172  1.00000000  0.345480592
## Grad.Rate    -0.27491252  0.45713119  0.34548059  1.000000000
```

#Problem1(c)

```
#c) Boxplots for graduation rates by university type and elite status

# Boxplot for Grad.Rate by Private/Public University
ggplot(data, aes(x=Private, y=Grad.Rate)) +
  geom_boxplot(fill="blue", color="black") +
  labs(title="Grad.Rate by University Type", x="University Type", y="Graduation Rate") +
  theme_minimal()
```




```
# Boxplot for Grad.Rate by Elite/Not Elite Status
ggplot(data, aes(x=factor(Elite10), y=Grad.Rate)) +
  geom_boxplot(fill="green", color="black") +
  labs(title="Grad.Rate by Elite Status", x="Elite Status (1=Elite, 0=Not Elite)", y="Graduat
ion Rate") +
  theme_minimal()
```



#Problem1(d)

```
#d) Fit a full model
full_model <- lm(Grad.Rate ~ . - school, data=data)
summary(full_model)
```

```
##
## Call:
## lm(formula = Grad.Rate ~ . - school, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65101 -0.09527  0.00743  0.12974  0.78402
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.793e+00  1.120e-01  33.883  < 2e-16 ***
## PrivateYes   7.732e-02  3.148e-02   2.456  0.01426 *
## Accept.pct  -2.152e-01  7.026e-02  -3.063  0.00227 **
## Elite10      5.042e-02  3.662e-02   1.377  0.16901
## F.Undergrad  1.292e-05  2.612e-06   4.949 9.20e-07 ***
## P.Undergrad -3.697e-05  7.138e-06  -5.179 2.86e-07 ***
## Outstate     2.066e-05  4.180e-06   4.944 9.43e-07 ***
## Room.Board   3.066e-05  1.087e-05   2.821  0.00491 **
## Books       -4.489e-05  5.423e-05  -0.828  0.40804
## Personal    -3.050e-05  1.423e-05  -2.144  0.03236 *
## PhD          2.085e-03  1.028e-03   2.029  0.04281 *
## Terminal    -5.456e-04  1.144e-03  -0.477  0.63348
## S.F.Ratio    3.465e-04  2.959e-03   0.117  0.90681
## perc.alumni  5.023e-03  8.846e-04   5.678 1.93e-08 ***
## Expend      -8.255e-06  2.775e-06  -2.975  0.00303 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2361 on 762 degrees of freedom
## Multiple R-squared:  0.3898, Adjusted R-squared:  0.3786
## F-statistic: 34.77 on 14 and 762 DF,  p-value: < 2.2e-16
```

#Problem1(e)

```
#e) Multi-collinearity and VIF statistics
library(car)
```

```
## Loading required package: carData
```

```
vif_values <- vif(full_model)
print(vif_values)
```

```
##     Private  Accept.pct     Elite10 F.Undergrad P.Undergrad    Outstate
##    2.739521    1.486633    1.687903    2.233117    1.643393    3.935059
##  Room.Board       Books    Personal         PhD    Terminal   S.F.Ratio
##    1.976762    1.115823    1.290983    3.917716    3.946581    1.909722
## perc.alumni      Expend
##    1.672367    2.922643
```

#Problem1(f)

```r
#f) Variable selection procedures
# Backward Selection
backward_model <- step(full_model, direction="backward")
```

```
## Start:  AIC=-2228.14
## Grad.Rate ~ (school + Private + Accept.pct + Elite10 + F.Undergrad +
##     P.Undergrad + Outstate + Room.Board + Books + Personal +
##     PhD + Terminal + S.F.Ratio + perc.alumni + Expend) - school
##
##               Df Sum of Sq    RSS     AIC
## - S.F.Ratio    1   0.00076 42.488 -2230.1
## - Terminal     1   0.01269 42.500 -2229.9
## - Books        1   0.03821 42.526 -2229.4
## - Elite10      1   0.10568 42.593 -2228.2
## <none>                     42.488 -2228.1
## - PhD          1   0.22953 42.717 -2225.9
## - Personal     1   0.25627 42.744 -2225.5
## - Private      1   0.33641 42.824 -2224.0
## - Room.Board   1   0.44380 42.931 -2222.1
## - Expend       1   0.49334 42.981 -2221.2
## - Accept.pct   1   0.52308 43.011 -2220.6
## - Outstate     1   1.36271 43.850 -2205.6
## - F.Undergrad  1   1.36555 43.853 -2205.6
## - P.Undergrad  1   1.49556 43.983 -2203.3
## - perc.alumni  1   1.79786 44.285 -2197.9
##
## Step:  AIC=-2230.12
## Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad + P.Undergrad +
##     Outstate + Room.Board + Books + Personal + PhD + Terminal +
##     perc.alumni + Expend
##
##               Df Sum of Sq    RSS     AIC
## - Terminal     1   0.01283 42.501 -2231.9
## - Books        1   0.03813 42.527 -2231.4
## - Elite10      1   0.10560 42.594 -2230.2
## <none>                     42.488 -2230.1
## - PhD          1   0.23150 42.720 -2227.9
## - Personal     1   0.25962 42.748 -2227.4
## - Private      1   0.33928 42.828 -2225.9
## - Room.Board   1   0.44359 42.932 -2224.1
## - Accept.pct   1   0.53070 43.019 -2222.5
## - Expend       1   0.60488 43.093 -2221.1
## - Outstate     1   1.36521 43.854 -2207.6
## - F.Undergrad  1   1.39627 43.885 -2207.0
## - P.Undergrad  1   1.49651 43.985 -2205.2
## - perc.alumni  1   1.80214 44.291 -2199.8
##
## Step:  AIC=-2231.89
## Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad + P.Undergrad +
##     Outstate + Room.Board + Books + Personal + PhD + perc.alumni +
##     Expend
##
##               Df Sum of Sq    RSS     AIC
## - Books        1   0.04461 42.546 -2233.1
## - Elite10      1   0.10636 42.608 -2231.9
## <none>                     42.501 -2231.9
## - Personal     1   0.25582 42.757 -2229.2
## - PhD          1   0.35827 42.859 -2227.4
## - Private      1   0.35898 42.860 -2227.3
```

```
## - Room.Board   1   0.43261 42.934 -2226.0
## - Accept.pct   1   0.53614 43.037 -2224.2
## - Expend       1   0.61044 43.112 -2222.8
## - Outstate     1   1.35246 43.854 -2209.6
## - F.Undergrad  1   1.38679 43.888 -2208.9
## - P.Undergrad  1   1.50352 44.005 -2206.9
## - perc.alumni  1   1.78933 44.291 -2201.8
##
## Step:  AIC=-2233.07
## Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + perc.alumni + Expend
##
##               Df Sum of Sq    RSS     AIC
## - Elite10      1   0.10504 42.651 -2233.2
## <none>                     42.546 -2233.1
## - Personal     1   0.30274 42.849 -2229.6
## - Private      1   0.34894 42.895 -2228.7
## - PhD          1   0.37773 42.924 -2228.2
## - Room.Board   1   0.40928 42.955 -2227.6
## - Accept.pct   1   0.50692 43.053 -2225.9
## - Expend       1   0.62302 43.169 -2223.8
## - F.Undergrad  1   1.35834 43.904 -2210.7
## - Outstate     1   1.36108 43.907 -2210.6
## - P.Undergrad  1   1.49937 44.045 -2208.2
## - perc.alumni  1   1.80726 44.353 -2202.8
##
## Step:  AIC=-2233.16
## Grad.Rate ~ Private + Accept.pct + F.Undergrad + P.Undergrad +
##     Outstate + Room.Board + Personal + PhD + perc.alumni + Expend
##
##               Df Sum of Sq    RSS     AIC
## <none>                     42.651 -2233.2
## - Personal     1   0.29979 42.951 -2229.7
## - Private      1   0.33284 42.984 -2229.1
## - Room.Board   1   0.39919 43.050 -2227.9
## - PhD          1   0.40176 43.053 -2227.9
## - Expend       1   0.52456 43.175 -2225.7
## - Accept.pct   1   0.71923 43.370 -2222.2
## - Outstate     1   1.38174 44.033 -2210.4
## - F.Undergrad  1   1.41008 44.061 -2209.9
## - P.Undergrad  1   1.66949 44.320 -2205.3
## - perc.alumni  1   1.88618 44.537 -2201.5
```

```
# Forward Selection
null_model <- lm(Grad.Rate ~ 1, data=data)
forward_model <- step(null_model, scope=list(lower=null_model, upper=full_model), direction
="forward")
```

```
## Start:  AIC=-1872.31
## Grad.Rate ~ 1
##
##               Df Sum of Sq    RSS     AIC
## + Outstate     1   19.7992 49.831 -2130.3
## + perc.alumni  1   14.5505 55.079 -2052.5
## + Room.Board   1   11.0671 58.563 -2004.8
## + Expend       1    8.3108 61.319 -1969.1
## + Private      1    6.7672 62.863 -1949.8
## + Elite10      1    6.3298 63.300 -1944.4
## + PhD          1    6.0403 63.590 -1940.8
## + Terminal     1    5.7000 63.930 -1936.7
## + S.F.Ratio    1    5.2624 64.368 -1931.4
## + Personal     1    4.6548 64.975 -1924.1
## + P.Undergrad  1    4.3536 65.276 -1920.5
## + Accept.pct   1    3.6669 65.963 -1912.3
## + F.Undergrad  1    0.2719 69.358 -1873.3
## <none>                     69.630 -1872.3
## + Books        1    0.0029 69.627 -1870.3
##
## Step:  AIC=-2130.27
## Grad.Rate ~ Outstate
##
##               Df Sum of Sq    RSS     AIC
## + perc.alumni  1   2.46802 47.363 -2167.7
## + P.Undergrad  1   0.98181 48.849 -2143.7
## + Accept.pct   1   0.75406 49.077 -2140.1
## + Personal     1   0.75053 49.080 -2140.1
## + PhD          1   0.66547 49.165 -2138.7
## + Elite10      1   0.64873 49.182 -2138.4
## + Terminal     1   0.39264 49.438 -2134.4
## + Room.Board   1   0.30187 49.529 -2133.0
## + F.Undergrad  1   0.20173 49.629 -2131.4
## <none>                     49.831 -2130.3
## + Books        1   0.05130 49.779 -2129.1
## + S.F.Ratio    1   0.04412 49.787 -2129.0
## + Private      1   0.02916 49.802 -2128.7
## + Expend       1   0.02242 49.808 -2128.6
##
## Step:  AIC=-2167.74
## Grad.Rate ~ Outstate + perc.alumni
##
##               Df Sum of Sq    RSS     AIC
## + Accept.pct   1   0.76895 46.594 -2178.5
## + Room.Board   1   0.65088 46.712 -2176.5
## + PhD          1   0.56267 46.800 -2175.0
## + P.Undergrad  1   0.53489 46.828 -2174.6
## + F.Undergrad  1   0.44159 46.921 -2173.0
## + Elite10      1   0.42245 46.940 -2172.7
## + Personal     1   0.41023 46.952 -2172.5
## + Terminal     1   0.30456 47.058 -2170.8
## + S.F.Ratio    1   0.17380 47.189 -2168.6
## <none>                     47.363 -2167.7
## + Expend       1   0.05994 47.303 -2166.7
## + Books        1   0.01170 47.351 -2165.9
```

```
## + Private      1   0.00395 47.359 -2165.8
##
## Step:  AIC=-2178.46
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct
##
##               Df Sum of Sq    RSS     AIC
## + P.Undergrad  1   0.78875 45.805 -2189.7
## + Personal     1   0.47712 46.117 -2184.4
## + Room.Board   1   0.40579 46.188 -2183.2
## + Expend       1   0.33908 46.255 -2182.1
## + PhD          1   0.29847 46.295 -2181.4
## + F.Undergrad  1   0.23346 46.360 -2180.4
## + S.F.Ratio    1   0.15344 46.440 -2179.0
## + Terminal     1   0.12805 46.466 -2178.6
## <none>                     46.594 -2178.5
## + Elite10      1   0.09922 46.495 -2178.1
## + Books        1   0.06809 46.526 -2177.6
## + Private      1   0.03330 46.560 -2177.0
##
## Step:  AIC=-2189.72
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad
##
##               Df Sum of Sq    RSS     AIC
## + F.Undergrad  1   1.22083 44.584 -2208.7
## + PhD          1   0.64320 45.162 -2198.7
## + Room.Board   1   0.51487 45.290 -2196.5
## + Terminal     1   0.37589 45.429 -2194.1
## + Expend       1   0.25996 45.545 -2192.2
## + Personal     1   0.24233 45.563 -2191.8
## + S.F.Ratio    1   0.22478 45.580 -2191.6
## <none>                     45.805 -2189.7
## + Elite10      1   0.06171 45.743 -2188.8
## + Books        1   0.04467 45.760 -2188.5
## + Private      1   0.01709 45.788 -2188.0
##
## Step:  AIC=-2208.71
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad
##
##              Df Sum of Sq    RSS     AIC
## + Room.Board  1   0.53677 44.047 -2216.1
## + Expend      1   0.47715 44.107 -2215.1
## + Personal    1   0.44653 44.138 -2214.5
## + Private     1   0.23306 44.351 -2210.8
## + PhD         1   0.19653 44.388 -2210.2
## <none>                    44.584 -2208.7
## + S.F.Ratio   1   0.09239 44.492 -2208.3
## + Books       1   0.08016 44.504 -2208.1
## + Terminal    1   0.06417 44.520 -2207.8
## + Elite10     1   0.00468 44.580 -2206.8
##
## Step:  AIC=-2216.12
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad + Room.Board
##
##             Df Sum of Sq    RSS     AIC
```

```
## + Expend     1   0.52950 43.518 -2223.5
## + Personal   1   0.38784 43.660 -2221.0
## + Private    1   0.17487 43.873 -2217.2
## + PhD        1   0.16029 43.887 -2217.0
## + Books      1   0.12290 43.925 -2216.3
## <none>                   44.047 -2216.1
## + S.F.Ratio  1   0.10797 43.939 -2216.0
## + Terminal   1   0.02887 44.019 -2214.6
## + Elite10    1   0.00569 44.042 -2214.2
##
## Step:  AIC=-2223.52
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad + Room.Board + Expend
##
##             Df Sum of Sq    RSS     AIC
## + Personal   1  0.295342 43.223 -2226.8
## + PhD        1  0.242133 43.276 -2225.9
## + Private    1  0.163862 43.354 -2224.4
## <none>                   43.518 -2223.5
## + Elite10    1  0.107978 43.410 -2223.4
## + Books      1  0.098685 43.419 -2223.3
## + Terminal   1  0.063646 43.454 -2222.7
## + S.F.Ratio  1  0.000864 43.517 -2221.5
##
## Step:  AIC=-2226.81
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad + Room.Board + Expend + Personal
##
##             Df Sum of Sq    RSS     AIC
## + PhD        1  0.238862 42.984 -2229.1
## + Private    1  0.169940 43.053 -2227.9
## <none>                   43.223 -2226.8
## + Elite10    1  0.110612 43.112 -2226.8
## + Terminal   1  0.059691 43.163 -2225.9
## + Books      1  0.050175 43.172 -2225.7
## + S.F.Ratio  1  0.000101 43.222 -2224.8
##
## Step:  AIC=-2229.12
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad + Room.Board + Expend + Personal + PhD
##
##             Df Sum of Sq    RSS     AIC
## + Private    1   0.33284 42.651 -2233.2
## <none>                   42.984 -2229.1
## + Elite10    1   0.08893 42.895 -2228.7
## + Terminal   1   0.04108 42.943 -2227.9
## + Books      1   0.03370 42.950 -2227.7
## + S.F.Ratio  1   0.00344 42.980 -2227.2
##
## Step:  AIC=-2233.16
## Grad.Rate ~ Outstate + perc.alumni + Accept.pct + P.Undergrad +
##     F.Undergrad + Room.Board + Expend + Personal + PhD + Private
##
##             Df Sum of Sq    RSS     AIC
## <none>                   42.651 -2233.2
## + Elite10    1  0.105038 42.546 -2233.1
```

```
## + Books      1  0.043282 42.608 -2231.9
## + Terminal   1  0.020120 42.631 -2231.5
## + S.F.Ratio  1  0.000759 42.650 -2231.2
```

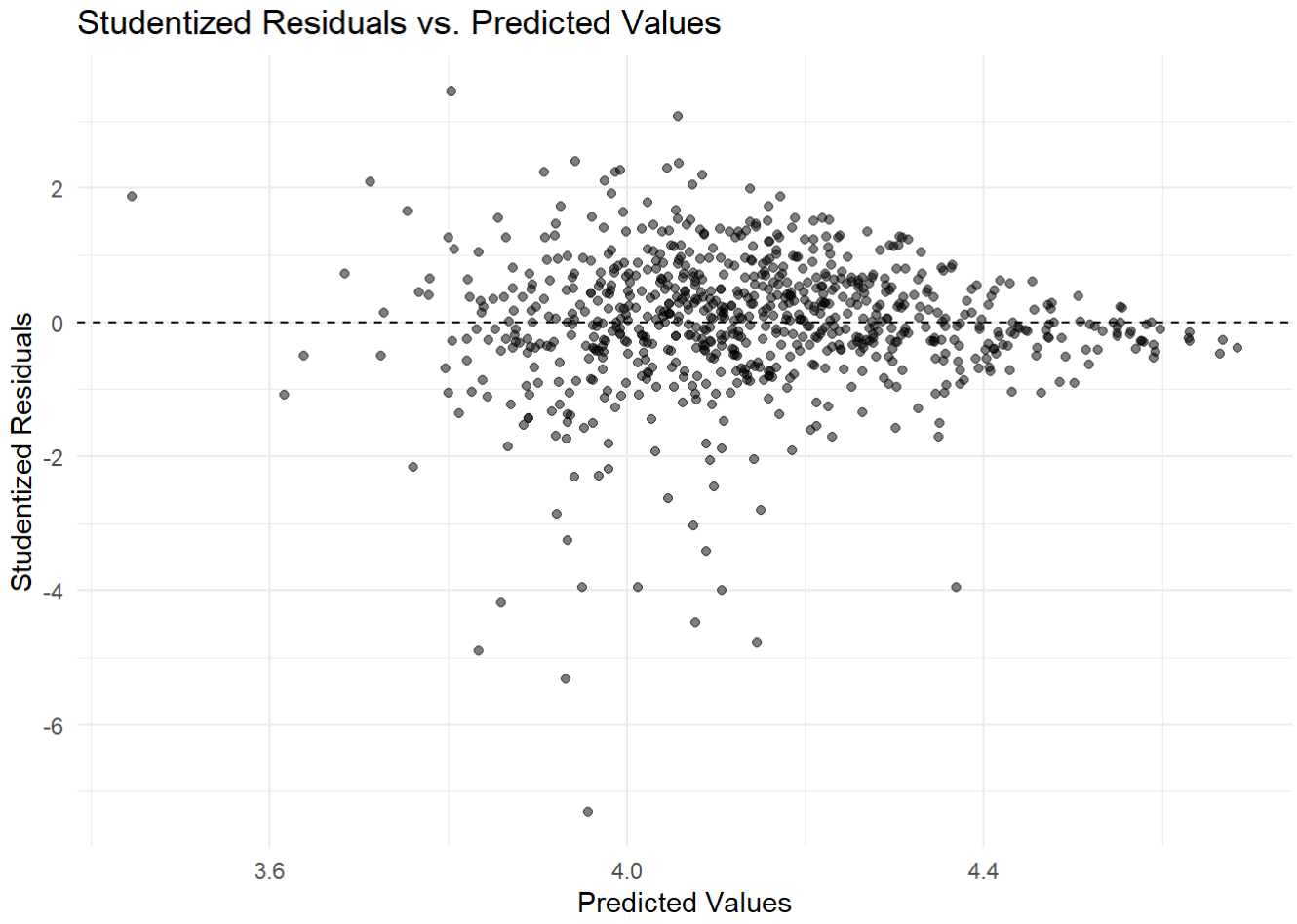
#Problem1(g)

```
#g) Fit a final regression model M1
#final Model based on the backward selection as observed in the output
M1 <- lm(Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad + P.Undergrad +
           Outstate + Room.Board + Personal + PhD + perc.alumni + Expend, data=data)
summary(M1)
```

```
##
## Call:
## lm(formula = Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad +
##     P.Undergrad + Outstate + Room.Board + Personal + PhD + perc.alumni +
##     Expend, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65407 -0.09599  0.00627  0.13276  0.80185
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.764e+00  8.441e-02  44.591  < 2e-16 ***
## PrivateYes   7.728e-02  3.085e-02   2.505 0.012457 *
## Accept.pct  -2.094e-01  6.936e-02  -3.019 0.002620 **
## Elite10      5.026e-02  3.657e-02   1.374 0.169757
## F.Undergrad  1.274e-05  2.578e-06   4.942 9.50e-07 ***
## P.Undergrad -3.700e-05  7.126e-06  -5.192 2.67e-07 ***
## Outstate     2.051e-05  4.146e-06   4.947 9.27e-07 ***
## Room.Board   2.913e-05  1.074e-05   2.713 0.006822 **
## Personal    -3.255e-05  1.395e-05  -2.333 0.019901 *
## PhD          1.764e-03  6.769e-04   2.606 0.009336 **
## perc.alumni  5.004e-03  8.778e-04   5.700 1.71e-08 ***
## Expend      -8.500e-06  2.540e-06  -3.347 0.000857 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2358 on 765 degrees of freedom
## Multiple R-squared:  0.389,  Adjusted R-squared:  0.3802
## F-statistic: 44.27 on 11 and 765 DF,  p-value: < 2.2e-16
```

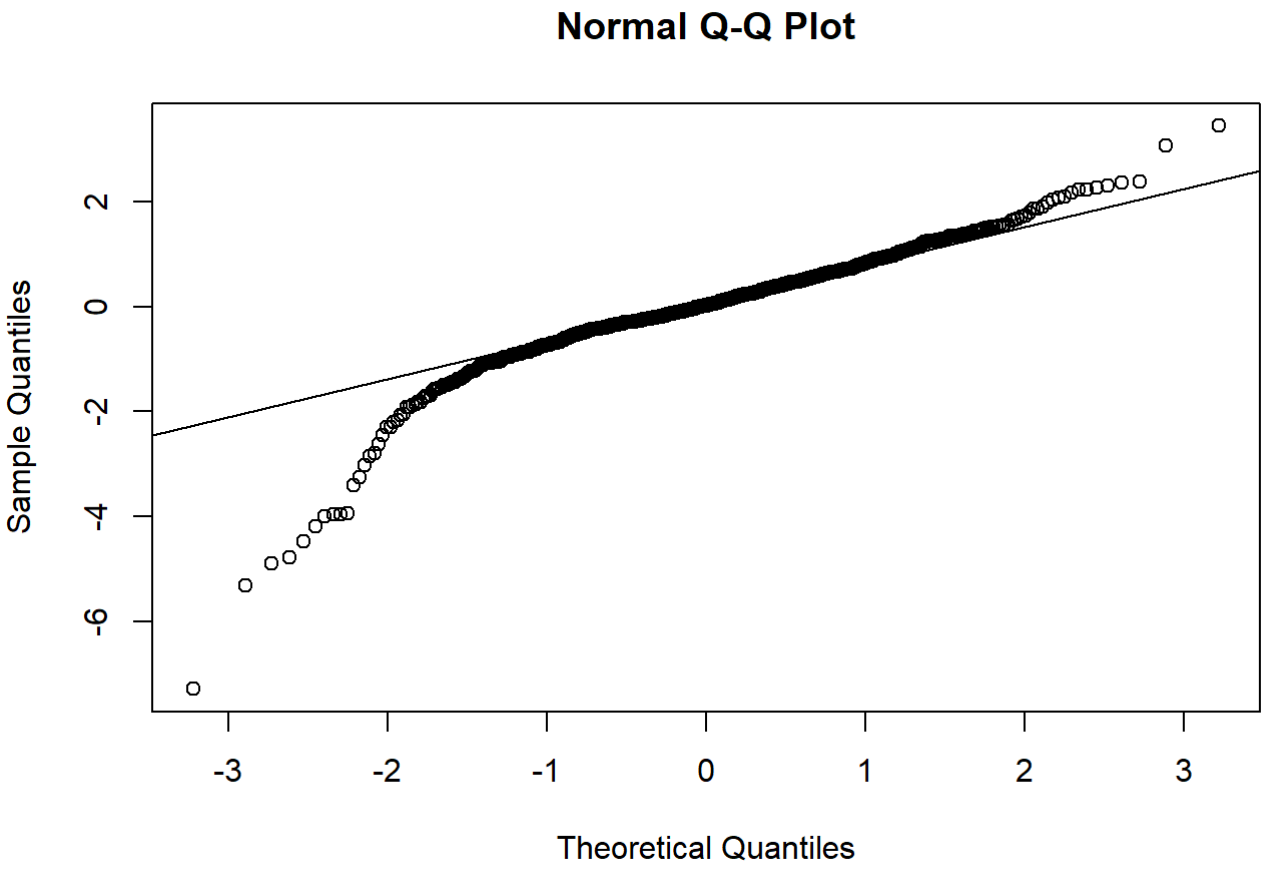
#Problem1(h)

```
#h) Scatter plot of studentized residuals against predicted values
# Compute studentized residuals
studentized_residuals <- rstudent(M1)

# Scatter plot
ggplot(data, aes(x=predict(M1), y=studentized_residuals)) +
  geom_point(alpha=0.5) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(title="Studentized Residuals vs. Predicted Values", x="Predicted Values", y="Studentiz
ed Residuals") +
  theme_minimal()
```



#Problem1(i)

```
#i) Normal probability plot of residuals
qqnorm(studentized_residuals)
qqline(studentized_residuals)
```



#Problem1(j)

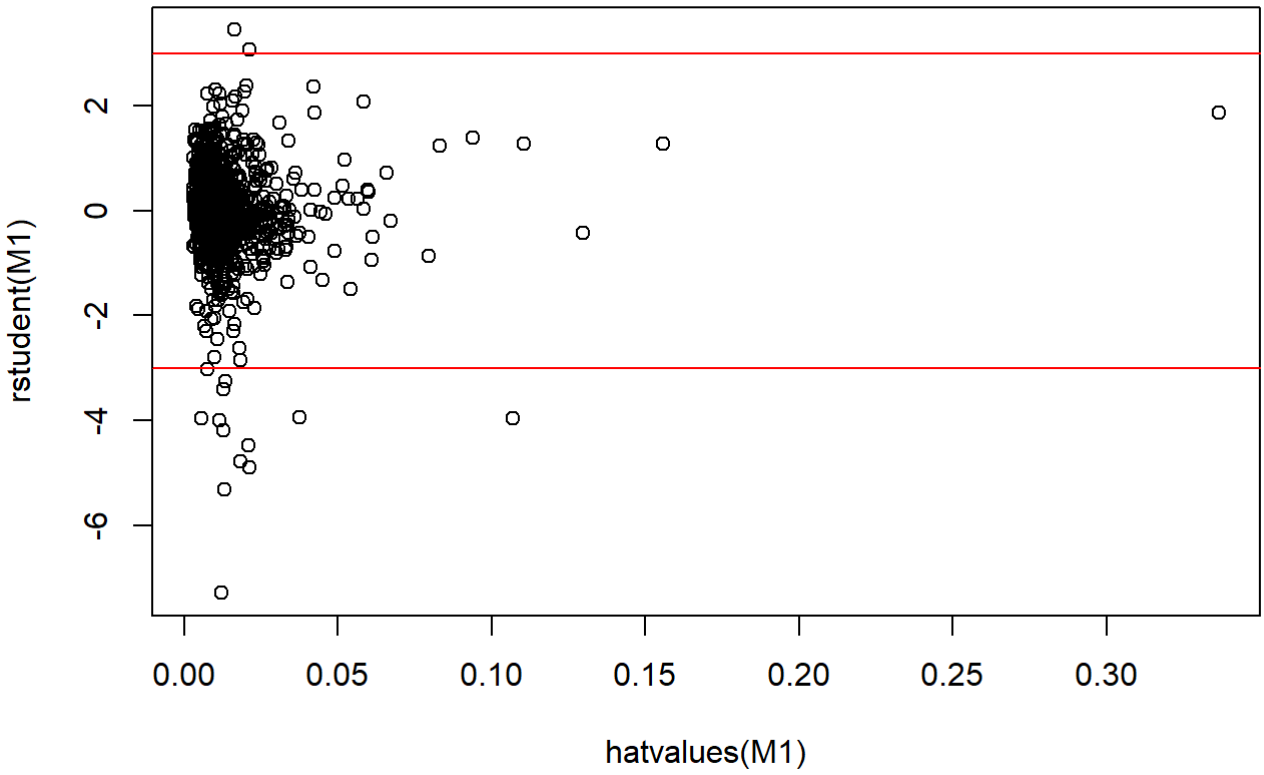
```
#j) Outliers or Influential Points
# Cook's distance to detect influential observations
cook_d <- cooks.distance(M1)

a<-influence.measures(M1)
summary(a)
```

```
## Potentially influential observations of
##   lm(formula = Grad.Rate ~ Private + Accept.pct + Elite10 + F.Undergrad +      P.Undergrad
+ Outstate + Room.Board + Personal + PhD + perc.alumni +      Expend, data = data) :
##
##     dfb.1_ dfb.PrvY dfb.Acc. dfb.El10 dfb.F.Un dfb.P.Un dfb.Otst dfb.Rm.B
## 5    0.01  -0.28     0.02     0.05     0.14    -0.03     0.24     0.04
## 17  -0.02   0.01     0.03    -0.01     0.01     0.00    -0.01     0.01
## 21   0.01   0.02    -0.04     0.05     0.01     0.00     0.02     0.04
## 24   0.02  -0.02    -0.01    -0.01    -0.03    -0.04     0.00    -0.01
## 38   0.01   0.01     0.00    -0.03     0.00     0.00     0.00    -0.03
## 48  -0.17   0.13     0.07     0.08    -0.01    -0.01    -0.27     0.14
## 67  -0.04  -0.29    -0.01    -0.08     0.02    -0.03     0.17     0.19
## 70   0.06  -0.81     0.08     0.11    -1.15_*   0.36     0.66    -0.09
## 96   0.12  -0.06     0.09     0.04     0.08    -0.01     0.00     0.14
## 99  -0.05   0.08    -0.10    -0.01     0.00    -0.04    -0.02    -0.19
## 101 -0.06   0.03     0.02     0.01    -0.01     0.00     0.00    -0.07
## 107  0.03  -0.01    -0.04    -0.06     0.01    -0.01     0.03     0.00
## 114 -0.30  -0.31     0.33     0.10    -0.03     0.04     0.36     0.21
## 145  0.03   0.00    -0.03     0.01    -0.01    -0.01    -0.01     0.00
## 198 -0.18   0.19     0.04     0.03     0.13     0.02    -0.08     0.18
## 199 -0.04  -0.07    -0.03    -0.02    -0.01     0.06    -0.02     0.00
## 202  0.04  -0.01    -0.06     0.03    -0.16     0.44     0.08    -0.16
## 216 -0.02  -0.11    -0.03    -0.04    -0.03     0.00     0.08    -0.03
## 224 -0.01   0.00     0.01    -0.01     0.05    -0.10    -0.02     0.04
## 239  0.18   0.16    -0.13     0.25     0.03     0.02    -0.11    -0.07
## 251  0.01   0.00    -0.02    -0.01     0.00    -0.01    -0.01     0.00
## 254  0.00   0.00     0.00    -0.01     0.00     0.00     0.00     0.00
## 265 -0.64  -0.15     0.60     0.16    -0.03     0.09    -0.05     0.23
## 266  0.04   0.04     0.05     0.06     0.01     0.02     0.02     0.02
## 275  0.00   0.00     0.00     0.00    -0.02     0.00     0.00     0.00
## 276 -0.21  -0.08     0.18     0.03    -0.05     0.00    -0.03     0.07
## 282 -0.18   0.13     0.19     0.08     0.10     0.03     0.16    -0.17
## 285 -0.02  -0.02     0.05    -0.07    -0.06     0.04    -0.18     0.01
## 318 -0.01   0.09    -0.16    -0.01    -0.02    -0.11     0.05     0.07
## 355  0.01   0.00    -0.01     0.00     0.00     0.00     0.00    -0.01
## 358 -0.09  -0.10     0.07     0.00     0.00     0.01     0.04     0.05
## 367  0.00   0.01     0.01     0.00     0.04     0.00     0.01    -0.01
## 369 -0.01   0.00     0.01     0.01     0.00    -0.01     0.00     0.01
## 378  0.19  -0.19     0.07     0.08    -0.11     0.14    -0.03    -0.05
## 379 -0.15  -0.02     0.05    -0.04    -0.06     0.03     0.02    -0.17
## 385 -0.06  -0.07    -0.05     0.03     0.06     0.11    -0.05     0.13
## 395 -0.16   0.10    -0.06    -0.03    -0.06    -0.03    -0.05     0.00
## 419  0.05  -0.05    -0.02    -0.02     0.02    -0.21     0.00    -0.07
## 427 -0.05   0.02    -0.05    -0.04    -0.02    -0.02     0.00    -0.11
## 431  0.00   0.00     0.00     0.01     0.00     0.00     0.00     0.00
## 446 -0.04  -0.03     0.06     0.05    -0.21     0.09    -0.05     0.02
## 452  0.06  -0.18    -0.04     0.00     0.03     0.02     0.23     0.01
## 457  0.02   0.00    -0.02    -0.05     0.01     0.00    -0.03     0.02
## 460 -0.01   0.00     0.02     0.00     0.00     0.00     0.00     0.00
## 462  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 498 -0.18   0.04     0.05    -0.05    -0.04    -0.06     0.00     0.03
## 507  0.12   0.01    -0.01     0.00     0.02    -0.01     0.06    -0.05
## 543  0.00   0.01    -0.01     0.00     0.00     0.00    -0.01     0.01
## 582  0.00  -0.02     0.00     0.01    -0.04     0.01     0.01     0.00
## 586 -0.24   0.48     0.04    -0.09     0.27    -0.38    -0.34     0.13
```

```
## 591  0.01   0.00    -0.01    -0.02     0.00     0.00     0.01     0.00
## 606  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 610 -0.01   0.00     0.01     0.00    -0.01     0.00    -0.02     0.00
## 620  0.00   0.00     0.00    -0.01    -0.01     0.00     0.00     0.00
## 624 -0.02   0.02     0.02     0.05     0.07    -0.03    -0.01     0.01
## 638  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 641 -0.04   0.06     0.02     0.06    -0.37     1.29_*   0.03    -0.16
## 645 -0.01  -0.02     0.00     0.00    -0.03     0.03     0.01     0.01
## 677  0.01   0.02    -0.01     0.01    -0.03     0.18     0.00    -0.03
## 686 -0.01   0.05    -0.01    -0.01     0.09     0.00    -0.01    -0.01
## 688  0.01   0.00     0.00     0.00     0.00     0.00     0.01     0.00
## 692  0.03   0.00    -0.03    -0.01     0.02    -0.08     0.00     0.00
## 701  0.00   0.00     0.00     0.00     0.01     0.00     0.00     0.00
## 713 -0.03   0.17    -0.06    -0.01     0.06     0.05     0.07    -0.16
## 715 -0.02  -0.07    -0.03    -0.03    -0.02     0.03     0.02     0.03
## 721  0.03   0.04    -0.01    -0.01    -0.02     0.00    -0.15    -0.07
## 729 -0.08  -0.01     0.13     0.01    -0.03     0.04    -0.10    -0.02
## 732  0.22   0.15    -0.24    -0.03     0.00    -0.05    -0.03    -0.11
## 736  0.03  -0.02     0.04    -0.02     0.03    -0.01    -0.03     0.08
## 763  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00
## 766  0.01   0.10     0.00     0.01     0.00    -0.03    -0.03    -0.06
## 776  0.01   0.00    -0.02    -0.01     0.00    -0.01    -0.02     0.00
##     dfb.Prsn dfb.PhD dfb.prc. dfb.Expn dffit   cov.r   cook.d hat
## 5    0.02    -0.27    0.38    -0.23    -0.61_*  0.67_*  0.03   0.01
## 17  -0.01     0.01   -0.02     0.01    -0.05    1.05_*  0.00   0.03
## 21   0.02     0.02    0.01    -0.16    -0.17    1.16_*  0.00   0.13_*
## 24   0.01     0.00    0.01     0.02    -0.08    1.05_*  0.00   0.04
## 38   0.00     0.01    0.00     0.02    -0.05    1.05_*  0.00   0.03
## 48   0.04     0.24    0.03    -0.02    -0.36    1.04    0.01   0.05_*
## 67  -0.10    -0.14    0.11    -0.03    -0.47_*  0.78_*  0.02   0.01
## 70   0.20    -0.04   -0.41    -0.08    -1.37_*  0.89_*  0.15   0.11_*
## 96  -0.14    -0.37    0.02     0.05     0.45_*  0.90_*  0.02   0.02
## 99   0.10     0.32    0.01    -0.06    -0.43_*  0.80_*  0.02   0.01
## 101  0.03     0.14    0.03    -0.04    -0.17    1.06_*  0.00   0.05_*
## 107  0.00    -0.02   -0.05     0.01    -0.09    1.05_*  0.00   0.04
## 114  0.10    -0.16   -0.17    -0.08    -0.65_*  0.76_*  0.03   0.02
## 145 -0.04     0.00   -0.05     0.06     0.12    1.05_*  0.00   0.04
## 198  0.15    -0.05    0.11    -0.04    -0.38_*  0.87_*  0.01   0.01
## 199 -0.06     0.10    0.17    -0.02    -0.30    0.80_*  0.01   0.01
## 202  0.10     0.01    0.08    -0.05     0.52_*  1.01    0.02   0.06_*
## 216  0.15    -0.04    0.11     0.03    -0.26    0.89_*  0.01   0.01
## 224 -0.04    -0.01    0.00     0.01    -0.13    1.08_*  0.00   0.06_*
## 239 -0.12     0.03   -0.04    -0.12     0.40_*  1.00    0.01   0.04
## 251  0.01    -0.01    0.01     0.03     0.05    1.07_*  0.00   0.05_*
## 254  0.00     0.00    0.00     0.00    -0.01    1.05_*  0.00   0.03
## 265 -0.09     0.32    0.21    -0.08    -0.78_*  0.83_*  0.05   0.04
## 266  0.01    -0.13   -0.10    -0.01     0.24    0.95_*  0.00   0.01
## 275  0.00     0.01    0.00     0.00    -0.02    1.05_*  0.00   0.04
## 276  0.05     0.10    0.06     0.04    -0.26    0.94_*  0.01   0.01
## 282  0.01     0.07   -0.02    -0.08    -0.39_*  0.91_*  0.01   0.02
## 285 -0.06    -0.05   -0.03     0.52     0.55_*  1.17_*  0.03   0.16_*
## 318  0.42     0.02   -0.05    -0.18     0.50_*  0.97    0.02   0.04
## 355  0.01     0.00   -0.01     0.03     0.05    1.05_*  0.00   0.03
## 358 -0.01     0.00    0.07     0.00    -0.18    0.95_*  0.00   0.01
## 367 -0.03     0.00   -0.01     0.00     0.06    1.07_*  0.00   0.05_*
## 369  0.05    -0.02    0.01     0.00     0.06    1.08_*  0.00   0.06_*
```

```
## 378 -0.27    -0.13   -0.04     0.05     0.45_*  0.86_*  0.02   0.02
## 379  0.06     0.26    0.01     0.09    -0.39_*  0.86_*  0.01   0.01
## 385 -0.42     0.24    0.26    -0.27    -0.72_*  0.72_*  0.04   0.02
## 395 -0.03     0.48   -0.40     0.04    -0.65_*  0.73_*  0.03   0.02
## 419  0.01     0.03   -0.01     0.03    -0.25    1.09_*  0.01   0.08_*
## 427  0.09     0.17   -0.06     0.07    -0.28    0.91_*  0.01   0.01
## 431  0.01     0.00    0.00     0.00     0.01    1.08_*  0.00   0.06_*
## 446 -0.02     0.06   -0.01     0.03    -0.24    1.07_*  0.00   0.06_*
## 452  0.11    -0.25    0.05    -0.05    -0.35    0.93_*  0.01   0.02
## 457 -0.03    -0.01    0.03     0.02    -0.08    1.05_*  0.00   0.03
## 460 -0.01     0.01   -0.01    -0.01    -0.03    1.05_*  0.00   0.03
## 462  0.00     0.00    0.00     0.00     0.01    1.06_*  0.00   0.04
## 498  0.42     0.04    0.01     0.08     0.45_*  1.11_*  0.02   0.11_*
## 507 -0.10    -0.12   -0.13     0.04     0.23    0.94_*  0.00   0.01
## 543  0.01     0.00    0.00     0.00     0.02    1.05_*  0.00   0.03
## 582  0.00     0.00   -0.01     0.00    -0.05    1.09_*  0.00   0.07_*
## 586 -0.06     0.23   -0.20     0.30    -0.81_*  0.46_*  0.05   0.01
## 591  0.00     0.00   -0.01     0.00    -0.03    1.05_*  0.00   0.03
## 606  0.00     0.00    0.00     0.00     0.00    1.05_*  0.00   0.03
## 610  0.00     0.00   -0.01     0.07     0.08    1.06_*  0.00   0.04
## 620  0.00     0.00    0.00     0.00    -0.01    1.06_*  0.00   0.05
## 624  0.00    -0.01   -0.01    -0.02     0.09    1.08_*  0.00   0.06_*
## 638  0.00     0.00    0.00     0.00    -0.01    1.06_*  0.00   0.04
## 641 -0.04    -0.08    0.27     0.10     1.34_*  1.45_*  0.15   0.34_*
## 645  0.06    -0.01    0.01    -0.01     0.08    1.05_*  0.00   0.04
## 677 -0.02     0.00   -0.01     0.02     0.19    1.08_*  0.00   0.07_*
## 686  0.02     0.00    0.00    -0.01     0.11    1.07_*  0.00   0.05_*
## 688  0.00    -0.01    0.00     0.00     0.02    1.05_*  0.00   0.03
## 692 -0.02    -0.01    0.00    -0.01    -0.10    1.05_*  0.00   0.04
## 701  0.00     0.00    0.00     0.00     0.01    1.05_*  0.00   0.03
## 713  0.11     0.06    0.01    -0.02    -0.29    0.95_*  0.01   0.02
## 715 -0.04     0.04    0.09    -0.02    -0.20    0.94_*  0.00   0.01
## 721 -0.03     0.01   -0.01     0.33     0.37    1.08_*  0.01   0.08_*
## 729 -0.02    -0.05   -0.06     0.41     0.45_*  1.09_*  0.02   0.09_*
## 732  0.07     0.00   -0.12    -0.04     0.35    0.95_*  0.01   0.02
## 736 -0.04    -0.17   -0.03     0.12     0.23    1.06_*  0.00   0.05_*
## 763 -0.01     0.00    0.00     0.00    -0.01    1.05_*  0.00   0.03
## 766  0.01     0.07   -0.12     0.01     0.19    0.95_*  0.00   0.01
```

```
plot(rstudent(M1)~hatvalues(M1))
abline(a=3, b=0, col= 'red')
abline(a=-3, b=0,col='red')
```




```
# Flag observations with Cook's distance > 4/n
influential_points <- as.numeric(names(cook_d)[(cook_d > 4/length(cook_d))])
print(influential_points)
```

```
##  [1]   5  48  67  70  96  99 114 143 153 170 198 199 202 216 239 265 276 282 285
## [20] 304 318 320 378 379 385 395 419 427 452 498 586 588 592 604 629 641 713 721
## [39] 729 732 777
```

```
# For outliers, you can inspect large studentized residuals
outliers <- which(abs(studentized_residuals) > 2) # Adjust the threshold as necessary
print(outliers)
```

```
##   5  67  70  96  99 114 170 198 199 202 216 265 266 273 276 282 304 318 320 358
##   5  67  70  96  99 114 170 198 199 202 216 265 266 273 276 282 304 318 320 358
## 378 379 385 395 427 440 452 507 586 590 713 715 732 766 777
## 378 379 385 395 427 440 452 507 586 590 713 715 732 766 777
```

#Problem1(k)

```
#k) adjusted r2 value
r_squared <- summary(M1)$r.squared
print(r_squared)
```

```
## [1] 0.3889722
```

#problem 1(L) Being an Elite10 university significantly affects the relationship between graduation rates and the predictors. Specifically:

The negative effect of acceptance percentage on graduation rate is stronger for Elite10 universities.

The positive effect of out-of-state tuition on graduation rate is weaker for Elite10 universities.

The relationship between the percentage of alumni who donate and graduation rate doesn’t change significantly for Elite10 universities.

The negative relationship between expenditure per student and graduation rate reverses for Elite10 universities.

This means that the factors influencing graduation rates differ somewhat for Elite10 universities compared to other institutions.