

# College Graduation Correlation

Josh Sutton

2024-02-11

## College Selected Info Table

```
unis <- read.csv("College_Data.csv") %>%  
  select(X,Private, Accept, Enroll, Top10perc,Grad.Rate) %>%  
  rename(College = X) %>%  
  arrange(Grad.Rate)  
head(unis,15)
```

##	College	Private	Accept	Enroll	Top10perc	Grad.Rate
## 1	Texas Southern University	No	3245	2604	15	10
## 2	Alaska Pacific University	Yes	146	55	16	15
## 3	Montreat-Anderson College	Yes	223	103	10	15
## 4	Brewton-Parker College	Yes	1228	1202	10	18
## 5	Claflin College	Yes	697	499	21	21
## 6	Huron University	Yes	197	124	3	21
## 7	Mount Saint Clare College	Yes	284	95	16	21
## 8	Ferrum College	Yes	1107	336	12	22
## 9	Centenary College	Yes	312	90	12	24
## 10	Fayetteville State University	No	1064	452	1	24
## 11	Jersey City State College	No	1423	691	10	26
## 12	Lamar University	No	1725	1043	10	26
## 13	Missouri Valley College	Yes	983	316	5	27
## 14	Gardner Webb University	Yes	930	332	18	29
## 15	University of Texas at Arlington	No	2559	1448	19	29

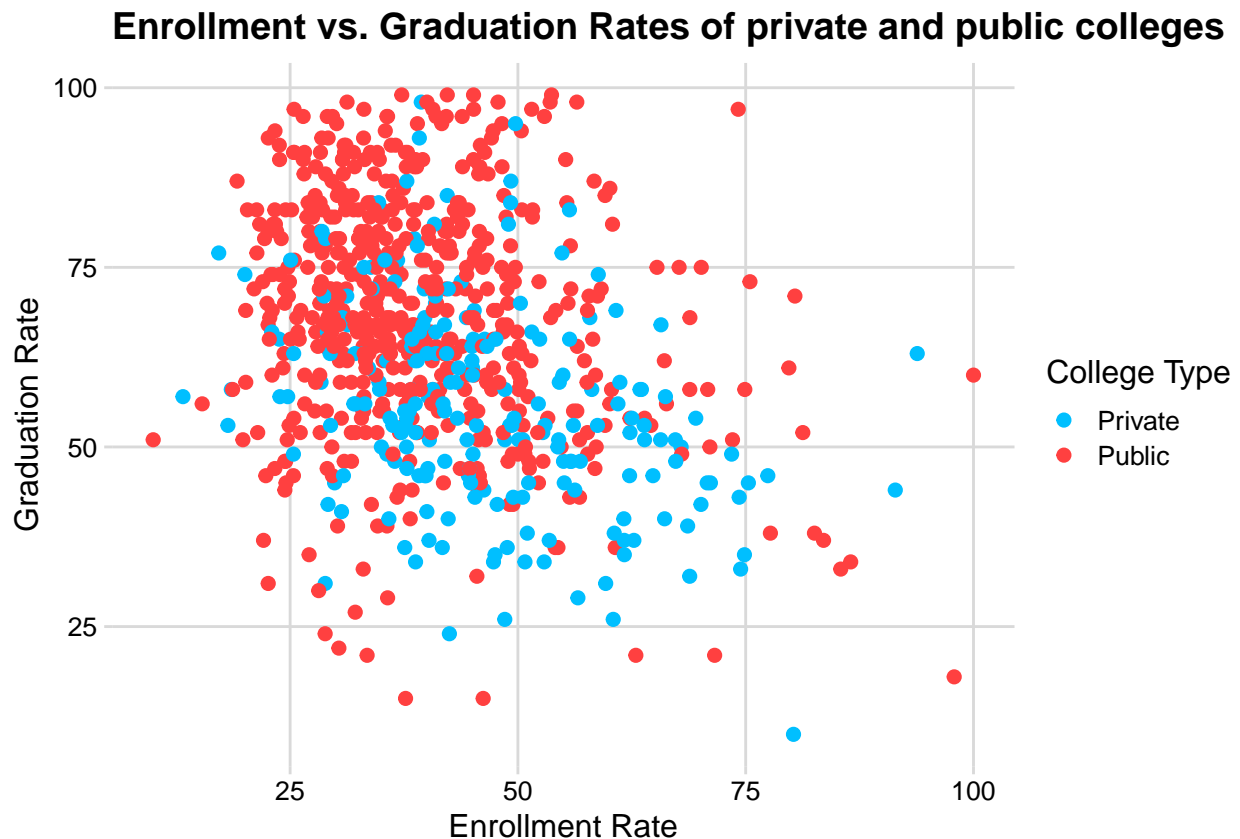
## College Modified Info Table with Tibble

```
unisMod <- read.csv("College_Data.csv") %>%  
  mutate(Enroll.Rate = (Enroll / Accept)*100) %>%  
  mutate(Accept.Rate = (Accept / Apps)*100) %>%  
  select(X, Private, Apps, Accept.Rate, Enroll, Enroll.Rate, Grad.Rate)  
as_tibble(unisMod)
```

```
## # A tibble: 777 x 7  
##       X                Private  Apps Accept.Rate Enroll Enroll.Rate Grad.Rate  
##   <chr>          <chr>    <int>      <dbl>  <int>      <dbl>      <int>  
## 1 Abilene Christian Uni~ Yes      1660      74.2    721        58.5        60  
## 2 Adelphi University    Yes      2186      88.0    512        26.6        56  
## 3 Adrian College       Yes      1428      76.8    336        30.6        54  
## 4 Agnes Scott College   Yes       417      83.7    137        39.3        59  
## 5 Alaska Pacific Univer~ Yes       193      75.6     55        37.7        15  
## 6 Albertson College     Yes       587      81.6    158        33.0        55  
## 7 Albertus Magnus Colle~ Yes       353      96.3    103        30.3        63  
## 8 Albion College        Yes      1899      90.6    489        28.4        73  
## 9 Albright College      Yes      1038      80.8    227        27.1        80  
## 10 Alderson-Broadbudds Col~ Yes       582      85.6    172        34.5        52  
## # i 767 more rows
```

## Visualizing College Data as Scatter Plot

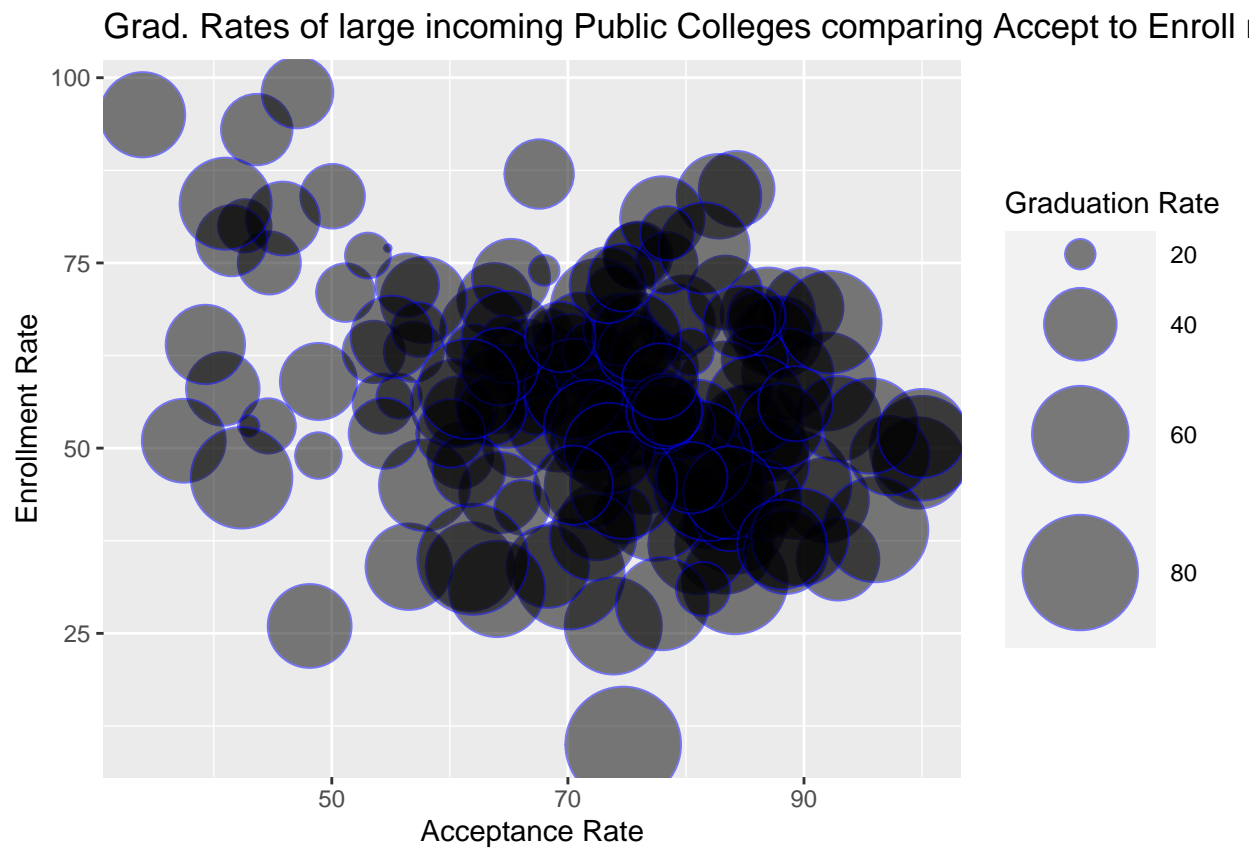
```
unisMod %>% ggplot(aes(Enroll.Rate, Grad.Rate, color = Private)) + geom_point(size=2) +  
  scale_color_manual(labels = c("Private", "Public"),  
    values = c("deepskyblue", "brown1")) +  
  labs(  
    title = "Enrollment vs. Graduation Rates of private and public colleges",  
    x = "Enrollment Rate",  
    y = "Graduation Rate",  
    color = "College Type"  
  ) + theme_minimal_grid(12)
```



This visualization coming from Lab Journal 3 is a scatter plot of private and public colleges and the comparison between enrollment rates and graduation rates of each college. From this graph we can see public colleges tend to have a lower enrollment rate but a higher graduation rate while private colleges tend to have a slightly higher enrollment rate and lower graduation rate.

## Visualizing using a Bubble Chart

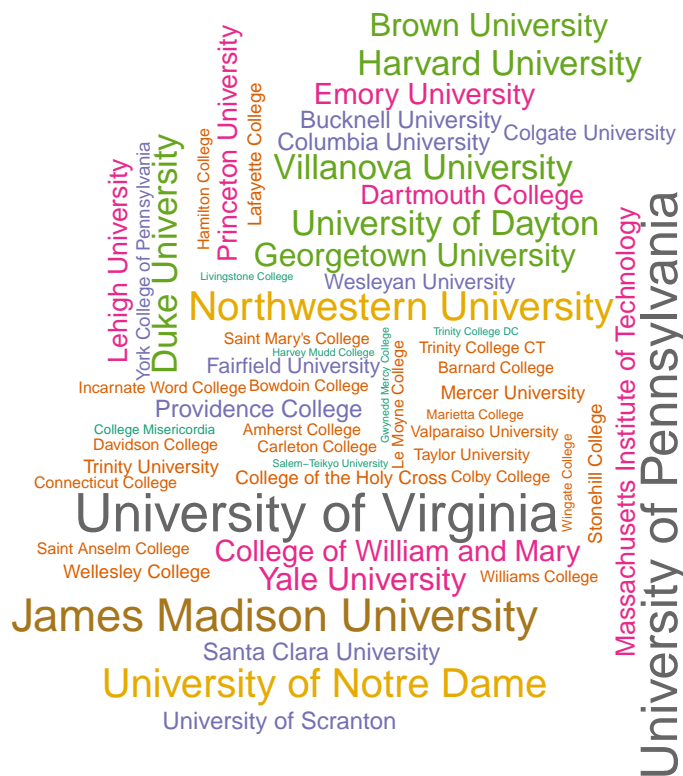
```
unisMod %>% filter(Private=='No') %>% filter(Enroll >=600) %>%  
  ggplot(aes(x = Accept.Rate, y = Grad.Rate, size = Enroll.Rate)) +  
  geom_point(alpha = 0.5, color = "blue", fill = "black", shape = 21) +  
  labs(  
    title = "Grad. Rates of large incoming Public Colleges comparing Accept to Enroll rates",  
    x = "Acceptance Rate",  
    y = "Enrollment Rate",  
    size = "Graduation Rate"  
  ) + scale_size(range = c(1,20))
```



This bubble chart coming from Lab Journal 3 is a 2D visualization of public colleges who enrolled 600 or more students. Point location is based on comparison between acceptance rate and enrollment rates. Each point size is determined by graduation rate. We can see that most colleges are found with a 70-90% acceptance rate and a 50% enrollment rate and roughly the colleges found in this area tend to have a graduation rate of 60% or less.

## Word Cloud

```
#png("wordcloud.png",width=20,height=8,units="in",res=300)
par(mar = rep(0,4))
set.seed(458)
unisHighAccept <- unisMod %>% filter(Grad.Rate > 90)
wordcloud(words = unisHighAccept$X, freq = unisHighAccept$Enroll, scale= c(1.7,0.25),
          max.words = 91, colors = brewer.pal(8,"Dark2"))
```



Visualization coming from Lab Journal 3, word clouds are used to display frequency of words through visualization where bigger words tie to more frequent appearance. This word cloud displays all colleges with a 90% graduation rate or higher and the frequency is the total amount of students enrolled in 2019. With more and more data the more difficult it becomes to see differences in words and more difficult to find desired words.