Lab 5

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────────────────────────────────────── tidyv
erse 1.2.1 ──
```

```
## ✔ ggplot2 3.2.1     ✔ purrr   0.3.2
## ✔ tibble  2.1.3     ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0     ✔ stringr 1.4.0
## ✔ readr   1.3.1     ✔ forcats 0.4.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────────────────────── tidyverse_c
onflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
colleges <- read_csv("https://www.dropbox.com/s/1ja0pnpiaoh9ykl/colleges1995.csv?dl=1")
```

```
## Parsed with column specification:
## cols(
##   Name = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   Avg.SAT = col_double(),
##   Size = col_double(),
##   Acceptance.Rate = col_double(),
##   Tuition = col_double()
## )
```

```
colleges <- colleges %>% mutate(
  Name = gsub("California State Univ. at", "CSU", Name),
  Name = gsub("California State University at", "CSU", Name),
  Name = gsub("California Polytechnic", "Cal Poly", Name),
  Name = gsub("California Poly", "Cal Poly", Name),
  Name = gsub("University of California at", "UC", Name)
)
```

Adjusting and Cleaning Data

```
CA_public <- colleges %>%
  filter(State == "CA" & Type == "Public") %>%
  select(Name, Avg.SAT, Acceptance.Rate, Size, Tuition) %>%
  drop_na()
```

drop_na() in the fourth line removed all the times in the dataset where there were rows containing missing values.

Querying Data

```
colleges %>% arrange(Tuition)
```

```
## # A tibble: 469 x 7
##    Name                  State Type   Avg.SAT  Size Acceptance.Rate Tuition
##    <chr>                 <chr> <chr>    <dbl> <dbl>           <dbl>   <dbl>
##  1 Boise State Universi… ID    Public     899  7948           0.851    2093
##  2 Troy State Universit… AL    Public     980  4343           0.871    2883
##  3 College of the South… NM    Priva…     850   352           0.792    3120
##  4 Savannah State Colle… GA    Public     709  2726           0.851    3200
##  5 University of Arkans… AR    Public     694  2993           0.938    3216
##  6 Texas College         TX    Priva…     620   503           0.744    3605
##  7 Southern College of … GA    Public     940  2229           0.771    4023
##  8 University of North … TX    Public     993 14047           0.620    4104
##  9 University of Housto… TX    Public     984 15340           0.640    4104
## 10 West Texas A&M Unive… TX    Public     890  4111           0.772    4104
## # … with 459 more rows
```

Franklin and Marshall has the highest out of state tuition in this data at $24,940.

```
colleges %>% arrange(Acceptance.Rate)
```

```
## # A tibble: 469 x 7
##    Name                  State Type   Avg.SAT  Size Acceptance.Rate Tuition
##    <chr>                 <chr> <chr>    <dbl> <dbl>           <dbl>   <dbl>
##  1 Stanford University   CA    Priva…    1401  6573           0.215   18669
##  2 Amherst College       MA    Priva…    1324  1593           0.231   19760
##  3 Brown University      RI    Priva…    1300  5643           0.257   19528
##  4 Duke University       NC    Priva…    1302  6188           0.282   18590
##  5 Howard University     DC    Priva…     883  6705           0.324    7128
##  6 Kent State University OH    Public     886 15615           0.330    7854
##  7 Spelman College       GA    Priva…    1002  1971           0.333    7000
##  8 Massachusetts Instit… MA    Priva…    1381  4481           0.334   20100
##  9 Sonoma State Univers… CA    Public     968  4239           0.360    7830
## 10 Pomona College        CA    Priva…    1340  1500           0.372   17720
## # … with 459 more rows
```

Stanford University has the lowest acceptance rate in this data at 21.5%.

```
CA_public %>% arrange(Size)
```

```
## # A tibble: 14 x 5
##    Name                          Avg.SAT Acceptance.Rate  Size Tuition
##    <chr>                           <dbl>           <dbl> <dbl>   <dbl>
##  1 Westmont College                 1051           0.751  1276   14320
##  2 CSU Stanislaus                    849           0.797  2895    8868
##  3 Sonoma State University           968           0.360  4239    7830
##  4 Humboldt State University         956           0.632  5637    9979
##  5 UC Riverside                      982           0.772  6922    7699
##  6 CSU Chico                         899           0.814 11822    5904
##  7 CSU Fullerton                     871           0.746 12147    8990
##  8 Cal Poly-San Luis                1002           0.489 12911    7380
##  9 San Francisco State University    816           0.657 13165    8964
## 10 UC San Diego                     1131           0.590 13738    7699
## 11 San Jose State University         838           0.786 14392    8820
## 12 UC Davis                         1070           0.697 15429    7699
## 13 CSU Northridge                    809           0.695 15567    5904
## 14 San Diego State University        851           0.747 16407    8384
```

San Diego State is the largest public university in California.

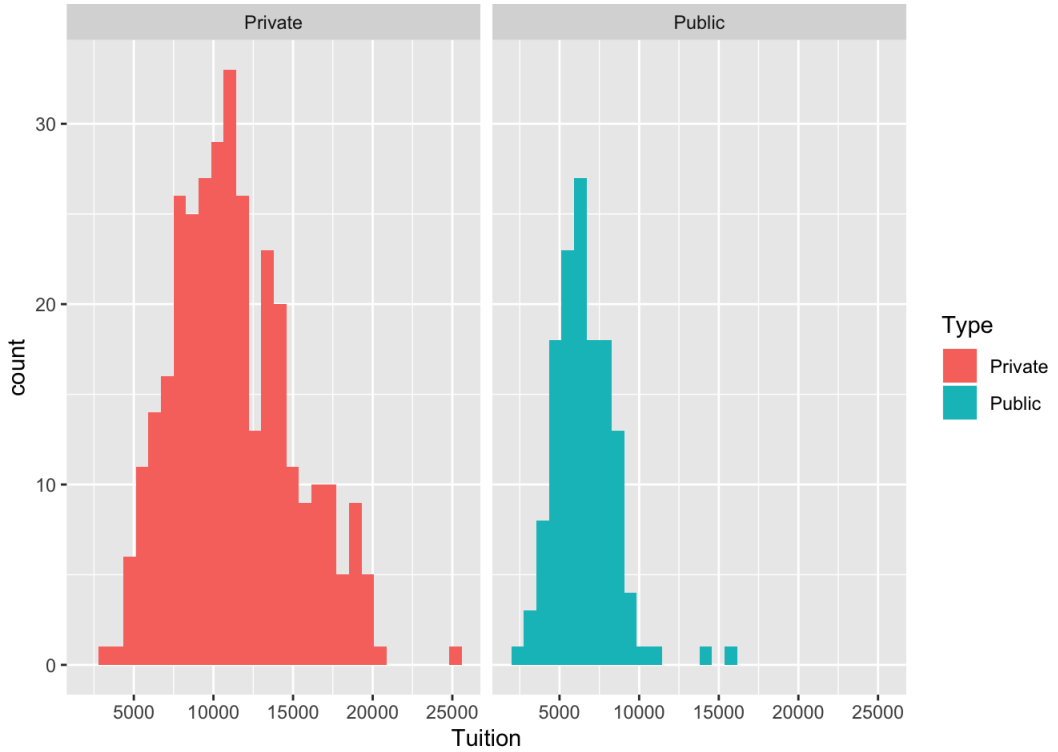
```
colleges <- colleges %>%
  mutate(
    BFB = factor((Avg.SAT) / (Tuition * Acceptance.Rate)))

colleges %>% arrange(BFB)
```

```
## # A tibble: 469 x 8
##    Name          State Type    Avg.SAT  Size Acceptance.Rate Tuition BFB
##    <chr>         <chr> <chr>     <dbl> <dbl>           <dbl>   <dbl> <fct>
##  1 Pine Manor … MA    Priva…      764   372           0.912   15190 0.05512…
##  2 Azusa Pacif… CA    Priva…      902  2032           0.878   15500 0.06629…
##  3 Menlo Colle… CA    Priva…      829   553           0.873   13975 0.06794…
##  4 Franklin an… PA    Priva…     1100  1748           0.648   24940 0.06803…
##  5 Hartwick Co… NY    Priva…     1002  1464           0.828   17480 0.06921…
##  6 Albertus Ma… CT    Priva…      890   416           0.963   13290 0.06952…
##  7 Antioch Uni… OH    Priva…     1004   712           0.927   15476 0.06997…
##  8 Woodbury Un… CA    Priva…      822   765           0.897   13020 0.07039…
##  9 Chapman Uni… CA    Priva…      957  1662           0.804   16624 0.07160…
## 10 Heidelberg … OH    Priva…      920   964           0.931   13780 0.07171…
## # … with 459 more rows
```

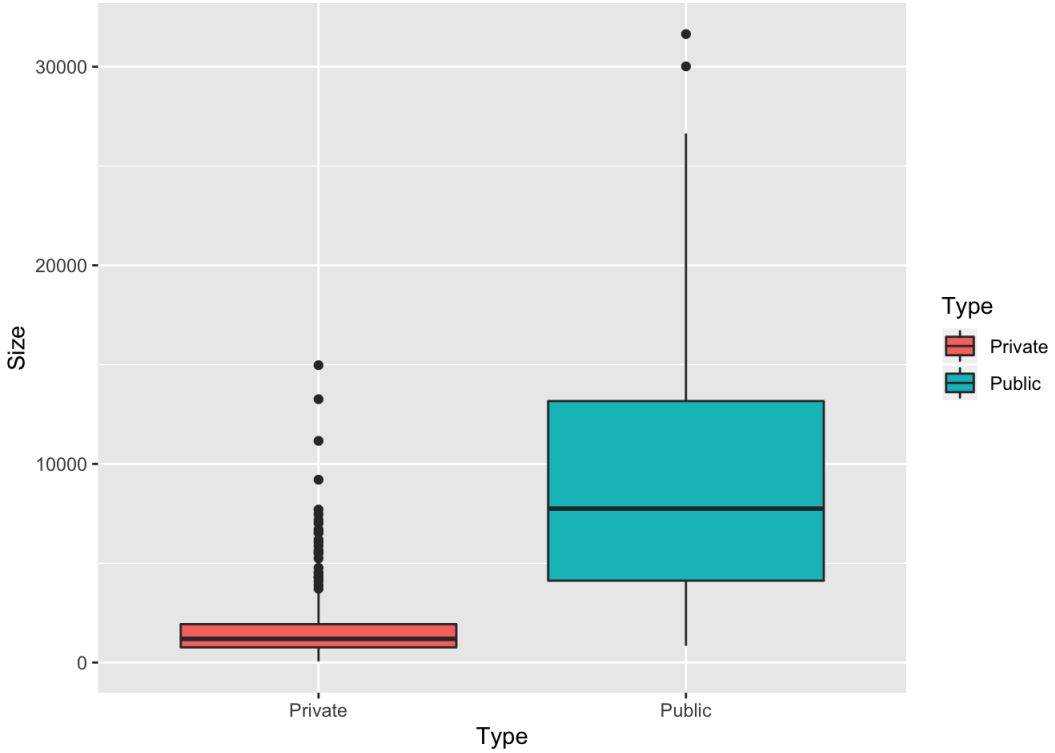
In my opnion, Boise State University has the best “bang for your buck.”

Plot

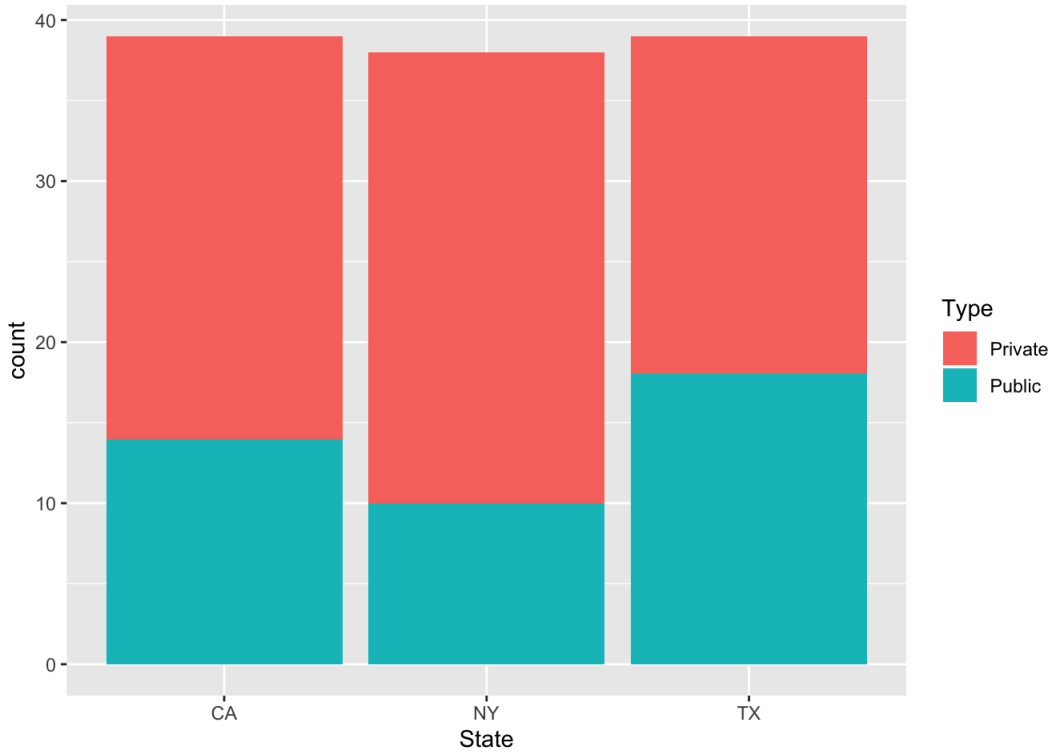
```
ggplot(colleges, aes(x = Tuition, fill = Type)) +
  geom_bar() +
  facet_grid(~Type) +
  stat_bin(binwidth = 787.76)
```



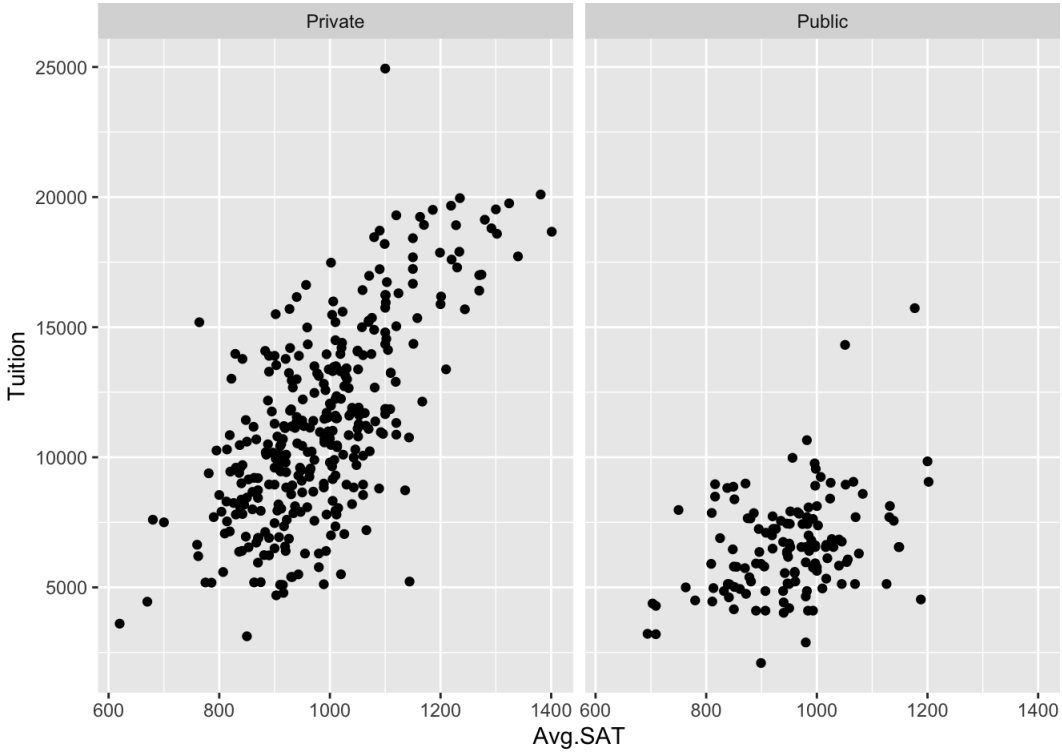

```
ggplot(colleges, aes(x = as.factor(Type), y = Size, fill = Type)) +
  geom_boxplot() +
  xlab("Type") +
  ylab("Size")
```




```
States <- filter(colleges, State == "CA"| State == "NY"| State == "TX")

ggplot(States, aes(x = State, fill = Type)) +
  geom_bar()
```




```
ggplot(colleges, aes(x = Avg.SAT, y = Tuition)) +
  geom_point() +
  facet_grid(~Type)
```



Questions

1. There are more private schools than public schools.
2. Yes, private schools tend to charge more than public schools.
3. The higher the average SAT score of admitted students, the higher the out of state tuition.
4. The relationship is a lot stronger for private schools than for public schools since, at the same average SAT score, private schools are more expensive than public schools.
5. Out of Texas, California, and New York, Texas most favors Public schools.
6. No, public schools area a lot larger than private schools.