Capstone_plotting

Grant Cox

November 9, 2018

R Markdown

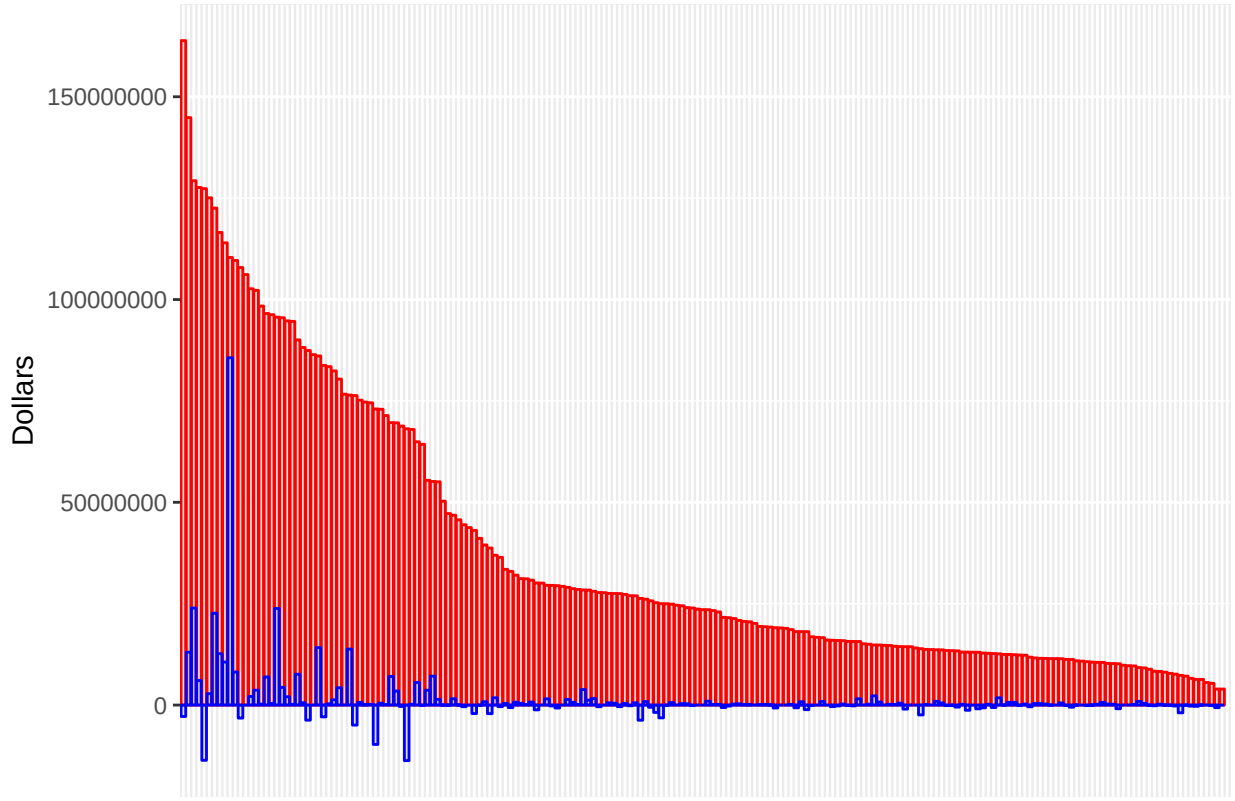
Intro text goes here. Here's a link to the github page where everything can be found https://github.com/gcox32/capstone.

For now, this copy will just be for thinking out loud. Here are the libraries used and the R code for uploading the wrangled data set.

```
library(readxl)
combined_df <- read.csv("C:/Users/gcox3/Desktop/capstone/combined_df.csv")
# View(combined_df)

library(ggplot2)
library(reshape2)

# since I'm going to be using this qualifier of "athletic_expenses > 0" often, I'll just name the new d
# "known_exp" will signify universities with listed athletic expense data to control for zeros in that
# as zeros just represent missing data, not truly $0

known_exp <- combined_df[which(combined_df$athletic_expenses>0),]
```

```
g <- ggplot(known_exp, aes(reorder(instnm, -athletic_expenses, na.rm = TRUE), athletic_expenses))

g2 <- ggplot(combined_df[which(combined_df$year==params$year),], aes(reorder(instnm, -athletic_expenses

g2 +
  geom_bar(aes(y = athletic_expenses), alpha = 0.2, stat = "identity", position = "identity", col = "re
  geom_bar(aes(y = net_revenue), alpha = 0.2, stat = "identity", position = "identity", col = "blue") +
  theme(axis.ticks.x = element_blank(), axis.text.x = element_blank()) +
  labs(x = "", y = "Dollars")
```

```
## Warning: Removed 36 rows containing missing values (geom_bar).

## Warning: Removed 36 rows containing missing values (geom_bar).
```



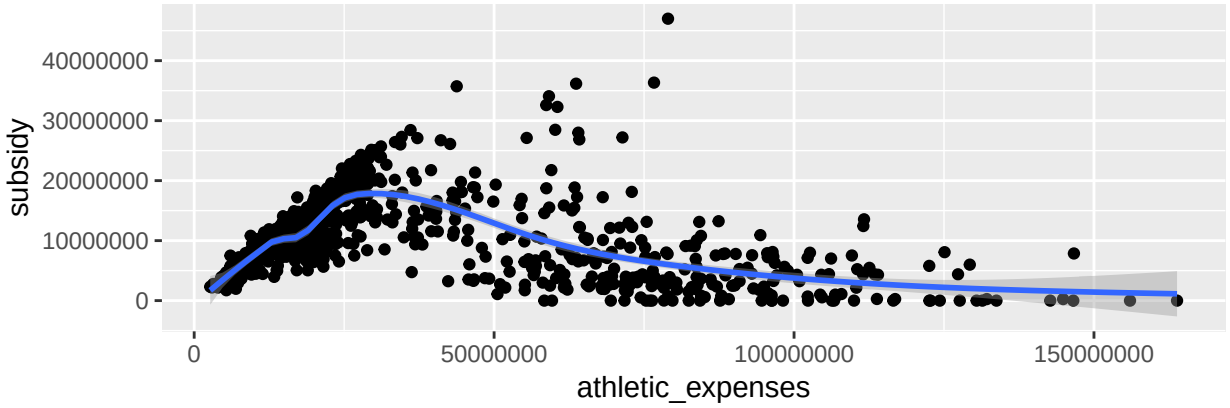

```
  # theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 4)) +
  # scale_x_discrete(labels = combined_df$abbrev_name)
```

Subsidy amount against Athletic Expense

One interesting trend that emmerged when exploring how subsidy amounts were effected by increasing athletic expenses. Essentially, we want to see if pouring more money into the athletic program might result in a decreased need for subsidizing the program. Would increased spending lead to success, increased revenue, and therefore decreased need for subsidizing (i.e. taxing the students)? What we see is an interesting curve:

```
ggplot(known_exp, aes(x = athletic_expenses, y = subsidy, na.rm = TRUE)) +
  geom_point() +
  geom_smooth() +
  coord_equal()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

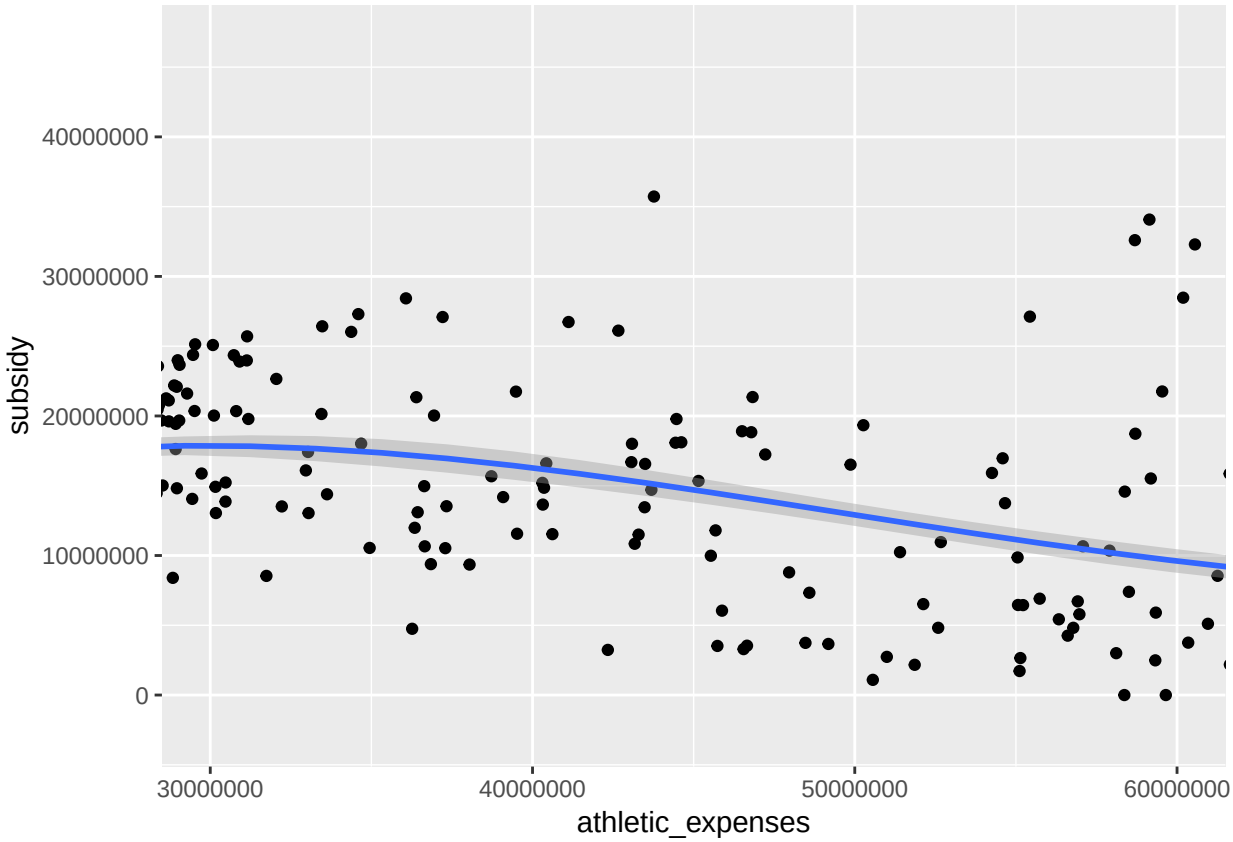


When we control for enrollment, though, we see the trends are unique between the bigger and smaller universities:

```
ggplot(known_exp, aes(x = athletic_expenses, y = subsidy, na.rm = TRUE)) +
  geom_point() +
  geom_smooth() +
  coord_equal() +
  coord_cartesian(if(params$school_size=="small") {
      xlim = c(0, 30000000)
      } else if (params$school_size=="medium") {
        xlim = c(30000000,60000000)
        } else { xlim = c(60000000,90000000)
      }
)
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
# library(DT)
# datatable(head(combined_df, n = nrow(combined_df)), options = list(pageLength = 5))
```

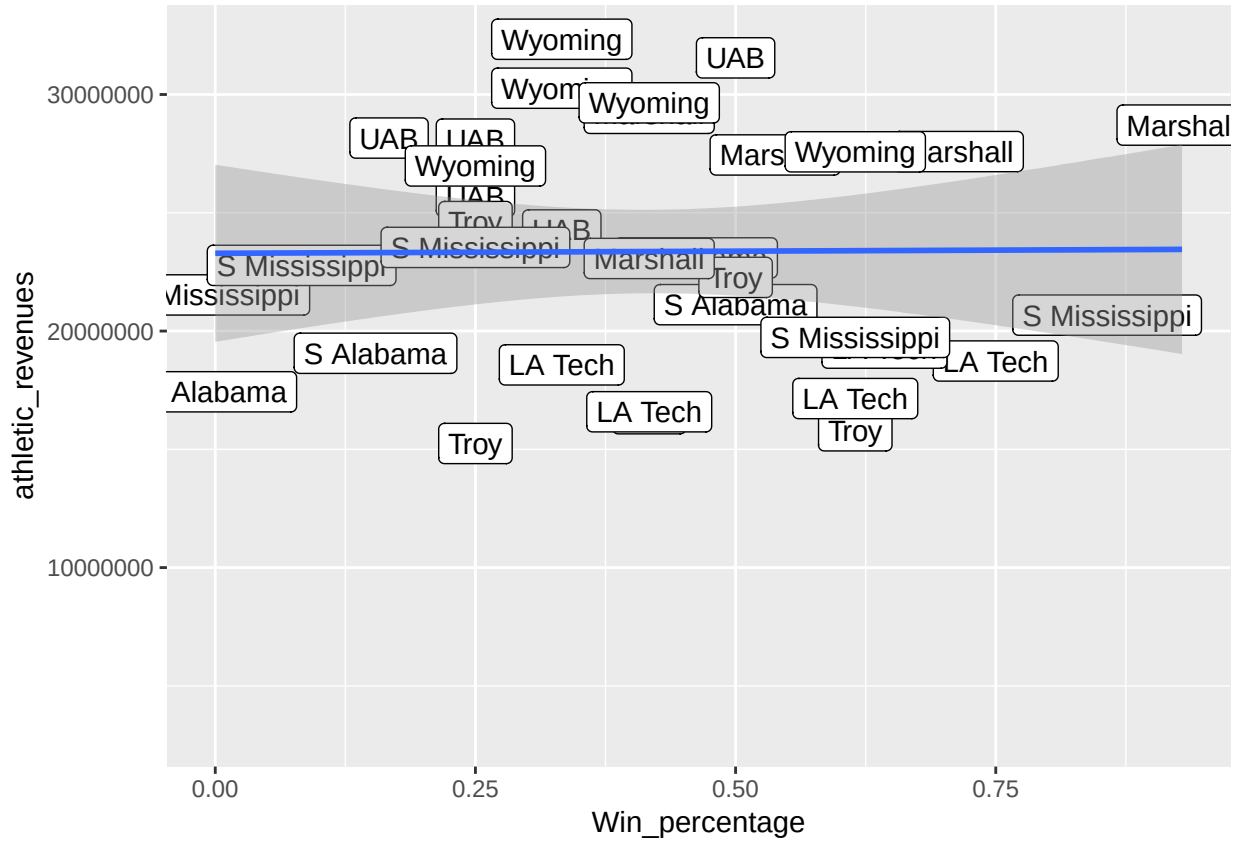
Will your win-loss record predict anything?

```
# simple vars for controlling for enrollment (note: first solution before taking advantage of params fe

gtop <- combined_df[which(combined_df$full_time_enrollment>40000),]
gbot <- combined_df[which(combined_df$full_time_enrollment<12500),]

ggplot(gbot, aes(x = Win_percentage, y = athletic_revenues, na.rm = TRUE)) +
  geom_point() +
  geom_label(aes(label = as.character(abbrev_name))) +
  geom_smooth(method = "lm")
```

```
## Warning: Removed 380 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 380 rows containing missing values (geom_point).
```

```
## Warning: Removed 380 rows containing missing values (geom_label).
```


Wyoming
UAB
Wyoming
Wyoming
Marshall
Troy
S Mississippi
S Alabama
LA Tech
Troy
athletic_revenues
30000000
20000000
10000000
0.00
0.25
0.50
0.75
Win_percentage