Georgia Recruit Analysis

Jeremy Harris

May 9, 2020

Georgia Football Recruiting Analysis (Past 10 Years)

I'm going to take a look at a few excel sheets that have all of the high school commits to college from the state of Georgia for the past 10 years. The goal is to provide the best potential pipeline for WVU to establish that will yield the highest caliber of athlete that will potentially commit to WVU.

```r
#tidy up the column names
recruiting_data <- orig_data

#change all-conf/nfl to factor with 1 an 0
recruiting_data$`All Conf/NFL` <- sub("X", "1", recruiting_data$`All Conf/NFL`)
recruiting_data$`All Conf/NFL`[is.na(recruiting_data$`All Conf/NFL`)] <- 0

#rename class & year - they had the same name
names(recruiting_data)[1] <- "year"
names(recruiting_data)[6] <- "hs.class"

#make tidy column names and rename columns
c_names <- colnames(recruiting_data) %>%
  make.names()
colnames(recruiting_data) <- c_names

#remove na from dataset on 247 natl rankings
recruiting_data$X247.Nat. <- sub("NA", "0", recruiting_data$X247.Nat.) %>%
  as.numeric()

#create factors for appropriate variables
c_names2 <- colnames(recruiting_data[c(1,4:9,12,17)])
recruiting_data[c_names2] <- lapply(recruiting_data[c_names2], factor)

#remove letter form hs.class
recruiting_data$hs.class <- gsub('.{1}$', '', recruiting_data$hs.class) %>%
  as.factor()

#add in state of college and in state / out of state factor
recruiting_data$coll.state <- as.factor(substr(recruiting_data$Signed, 1, 2))

recruiting_data <- recruiting_data %>%
  mutate(in.state = ifelse((coll.state == "GA"), 1, 0),
         in.sec = ifelse((Conf. == "SEC"), 1, 0))
```
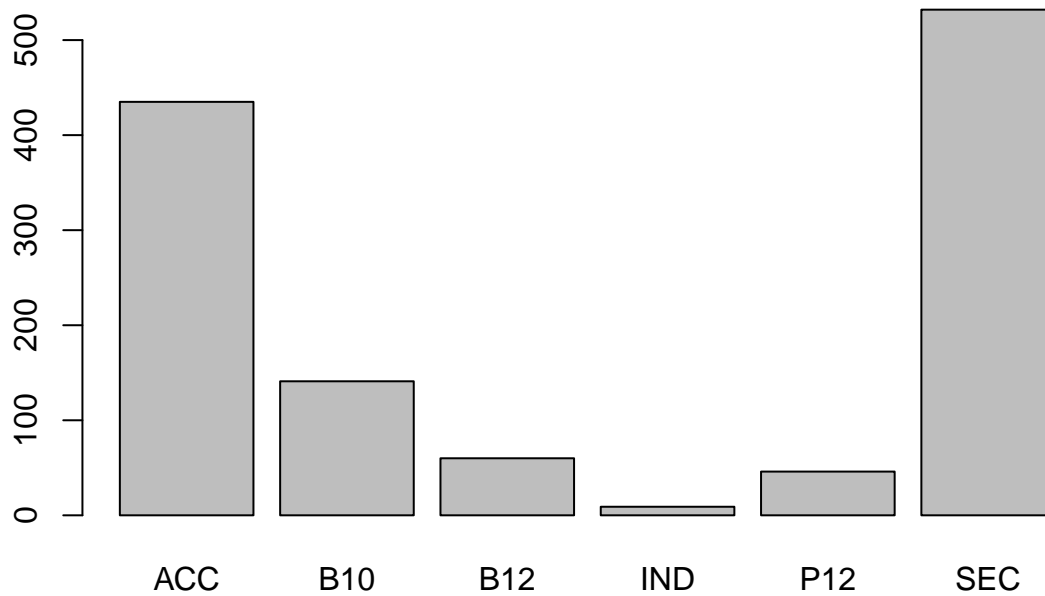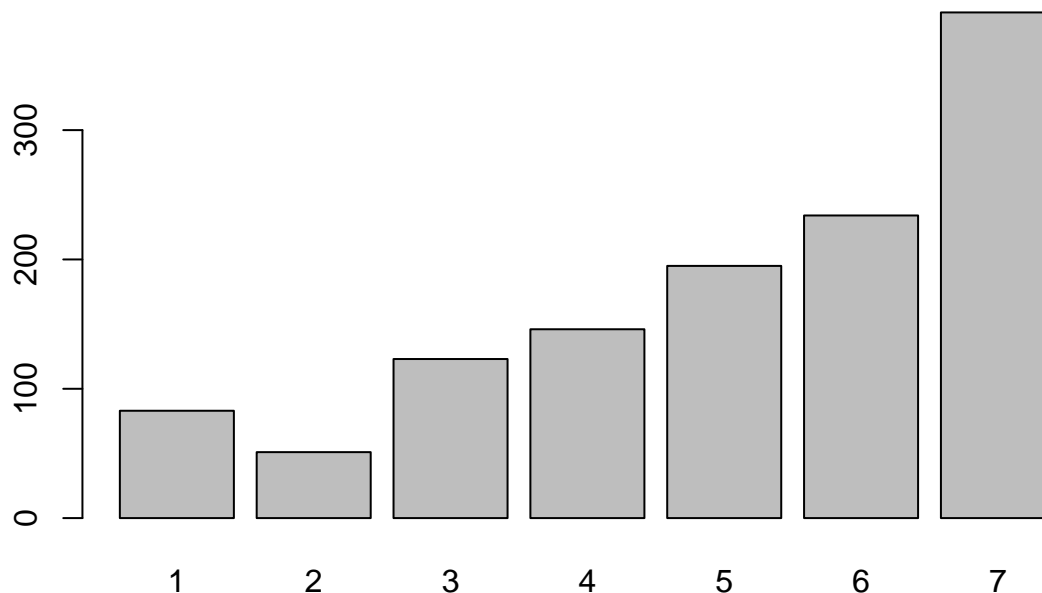
Look At Basic Plots for Inference

I just want to look at rough plots to get an idea of how this data shapes up before I start my analysis.

```r
#Conferences of commits
plot(recruiting_data$Conf.)
```



```r
#hs class of commits
plot(recruiting_data$hs.class)
```

Create "Like" WVU Category

I want to create a category that contains schools that are "like" WVU. I'll use the data I have with a little outside research to lump schools together in terms of location, conference and then final season rankings. I'll use this new category for the analysis to determine which commits would have been more likely to attend WVU with a statistical probability. From there, I will analyze the detailed information of the commits to find commonality with things such as school size, school location and recruit 247 rating.

I can see that there are 198 of the 1,223 total commits narrowed down just by selecting the Big 12 and the surrounding states.

Verify Predictions and Analyze the Variables

Now I will see how well I can build a model that will predict whether a recruit is likely to go to a like_wvu school. I will analyze the findings to determine what the driving factors are as well as look to narrow down locations of interest for WVU.

Looking at the raw numbers and verifying with a variable importance output, I can see that the following variables have the highest influence on my results:

- **HS Size**: We're going to see better results with 4A & 6A. We'll see the worst results with 7A.
- **Positions**: We'll see the best results with QB, RB, OLB & WR. Worst results with FB & PK.
- **247 Rating**: Our best star rating is 3 and 5. Worst rating is 4 star.

More Initial Data Analysis

I create a few dataframes here to take a look at different variables that might be important.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 256  49
##          1   0   0
##
##                Accuracy : 0.8393
##                  95% CI : (0.7932, 0.8787)
##     No Information Rate : 0.8393
##     P-Value [Acc > NIR] : 0.538
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 7.025e-12
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.8393
##          Neg Pred Value :    NaN
##              Prevalence : 0.8393
##          Detection Rate : 0.8393
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : 0
##
```

All-Conf / NFL Feature Importance

I want to gain more insight into the variables that lead to a player becoming an All-Conference or NFL player. There is a lot to take in here that cannot be captured in the data - however, I was able to determine certain attributes that lead to a more likely scenario for success.

Graphically Analyze the "Like WVU" Category

I want to take the commits that went to the like_wvu category and gain insight into the makup of the 198 observations. In particular, I want to see if we can narrow down a geographic region or other valuable insight for the state of Georgia in terms of focusing our recruiting efforts.

Cities in GA with More than 1 Commits in 5 Years
To Schools 'Like' WVU
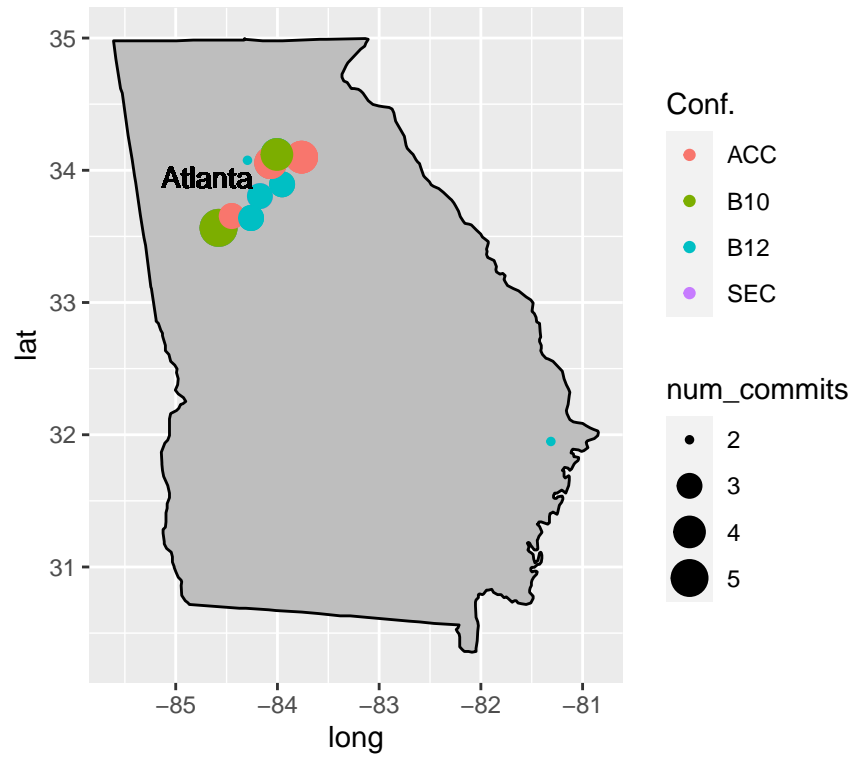
Commits By Town Colored by College
Years 2015 – 2020
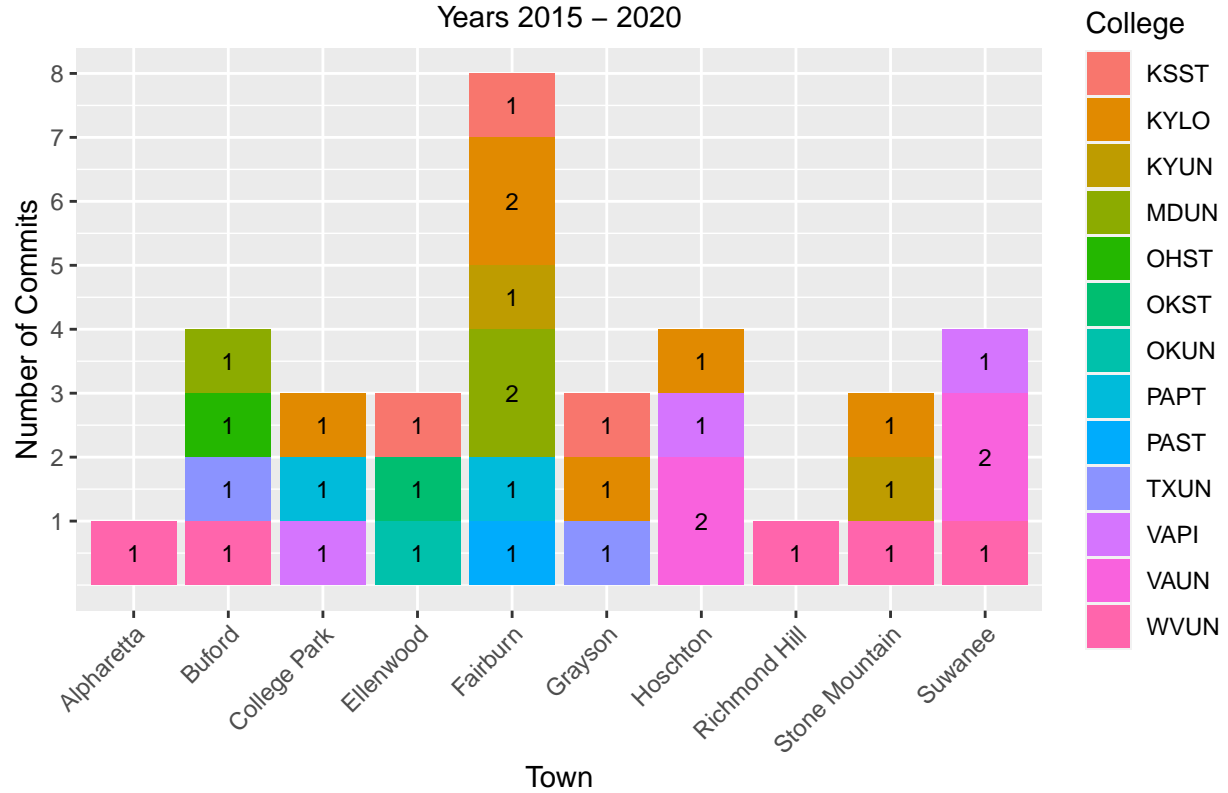
Commits By Town Colored by Position