Stats 107 Project

Omar Sepulveda, Riccardo Gutierrez

2025-11-06

###ABSTRACT - The NCAA SEC is one of the premier conferences in college football from recent years, but how much is each program represented in the draft? What we are looking to accomplish is to highlight which schools in the NCAA SEC (South Eastern Conference) have gained representation in the draft and which programs have lost representation in the draft over the last 15 years. We collected our data from Pro-Football-Reference.com and combined each program in the SEC data sets into a singular Excel file and analyzed the data by distinhuishing which players came from the which college program, seeing if the player was drafted, then drafted within the years 2011-2025.

Data

The data set used for this analysis, "College_Dataset.xlsx," contains information on NFL players and the colleges they attended. The data was obtained from Pro-Football-Reference, a reliable source that collects and publishes official NFL statistics, including player backgrounds and school affiliations. Relevant variables include each player's name, the college they attended, and the range of years they played in the NFL. During the cleaning process, unnecessary columns such as AP1, PB, St, wAV, Ht, and Wt were removed, keeping only the relevant variables. Additionally, since the data set included players from earlier eras, we extracted the starting year from each player's career range and filtered the data to include only those who began their NFL careers in 2011 or later. The data is entirely real and will not be generated through any randomized simulation.

```
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Read the Excel file
data <- read_excel("College Dataset.xlsx")
```

```
## Warning: Expecting date in H2740 / R2740C8: got '0-0'
```

```
## New names:
## * '' -> '...3'
## * '' -> '...11'
```

```r
# Rename columns for clarity (based on the header)
colnames(data) <- c("Pos", "Player", "College_Stats", "AP1", "PB", "St", "wAV", "Ht", "Wt", "Yrs_Num",

# Extract the start year from Years_Range
data <- data %>%
  mutate(start_year = as.integer(sub("-.*", "", Years_Range)))

# Filter to keep players who started in 2011 or later
filtered_data <- data %>%
  filter(start_year >= 2011)

# Select only the specified columns
fixed_data <- filtered_data %>%
  select(Pos, Player, Yrs_Num, Years_Range, School)

# Rename columns as per user labels if desired
colnames(fixed_data) <- c("Position", "Player", "Years Played", "Years Played (From-To)", "School")

# View the selected data (or you can use write_excel_csv to save it)
head(fixed_data)
```

```
## # A tibble: 6 x 5
##   Position Player          'Years Played' 'Years Played (From-To)' School
##   <chr>    <chr>                    <dbl> <chr>                    <chr>
## 1 DT       Jonathan Allen               9 2017-2025                Alabama
## 2 CB       Trey Amos                    1 2025-2025                Alabama
## 3 OLB      Ryan Anderson                5 2017-2022                Alabama
## 4 LB       Will Anderson                3 2023-2025                Alabama
## 5 DB       Jalyn Armour-Davis           4 2022-2025                Alabama
## 6 DB       Terrion Arnold               2 2024-2025                Alabama
```