Trends in NFL Player Representation from SEC Schools (2011–2025)

Omar Sepulveda, Riccardo Gutierrez, Naomi Menard

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? 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 distinguishing which players came from which college program, seeing if the player was drafted, then drafted within the years 2011-2025.

Introduction

The purpose of this analysis 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. This will answer the question: "How has the number of NFL players from each SEC school changed over time (e.g., 2011–2015, 2016–2020, and 2021–2025), and which schools have shown the most growth or decline in player representation?" This analysis may benefit various groups, including NFL scouts/teams, sports analysts/journalists, or other researchers looking to see which college improved the most. The goal is to understand patterns and changes throughout the 2011 - 2015, 2016-2020, and 2021-2025 time periods.

Data

```
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'
```

```
# 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 in NFL", "Years Played in NFL (From-To)",

# 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 in NFL' Years Played in NFL ~1 School
##   <chr>    <chr>                            <dbl> <chr>                  <chr>
## 1 DT       Jonathan Allen                       9 2017-2025              Alaba~
## 2 CB       Trey Amos                            1 2025-2025              Alaba~
## 3 OLB      Ryan Anderson                        5 2017-2022              Alaba~
## 4 LB       Will Anderson                        3 2023-2025              Alaba~
## 5 DB       Jalyn Armour-Dav~                    4 2022-2025              Alaba~
## 6 DB       Terrion Arnold                       2 2024-2025              Alaba~
## # i abbreviated name: 1: 'Years Played in NFL (From-To)'
```

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.

Visualization & Analysis

```r
# Load necessary libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v forcats   1.0.1      v readr     2.1.5
## v ggplot2   4.0.0      v stringr   1.5.2
## v lubridate 1.9.4      v tibble    3.3.0
## v purrr     1.1.0      v tidyr     1.3.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(gridExtra)  # for arranging multiple plots
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
# Read the CSV file (update the path if needed)
df <- read.csv("my_updated_dataset.csv")

# Convert 'Years Played (From-To)' into numeric start and end years
df <- df %>%
  separate(`Years.Played..From.To.`, into = c("Start", "End"), sep = "-", convert = TRUE)

# Define year ranges
df <- df %>%
  mutate(YearRange = case_when(
    Start >= 2011 & Start <= 2015 ~ "2011-2015",
    Start >= 2016 & Start <= 2020 ~ "2016-2020",
    Start >= 2021 & Start <= 2025 ~ "2021-2025",
    TRUE ~ NA_character_
  ))

# Count players per team and year range
team_summary <- df %>%
  filter(!is.na(YearRange)) %>%
  group_by(School, YearRange) %>%
  summarise(Players = n(), .groups = "drop")

# List of teams
teams <- unique(team_summary$School)

# Find the global maximum Players count for consistent y-axis
y_max <- max(team_summary$Players)
```

```r
# Create plots and print 4 per group
plots <- list()

for (team in teams) {
  team_data <- team_summary %>% filter(School == team)

  p <- ggplot(team_data, aes(x = YearRange, y = Players, fill = YearRange)) +
    geom_bar(stat = "identity", width = 0.6) +
    scale_y_continuous(limits = c(0, y_max)) +  # consistent y-axis
    labs(title = team, x = "Year Range", y = "Number of Players") +
    theme_minimal(base_size = 12) +
    theme(
      legend.position = "none",
      plot.title = element_text(hjust = 0.5, face = "bold")
    )

  plots[[team]] <- p
}

# Print 4 plots per image
for (i in seq(1, length(plots), by = 4)) {
  grid.arrange(
    grobs = plots[i:min(i + 3, length(plots))],
    ncol = 2, nrow = 2
  )
}
```
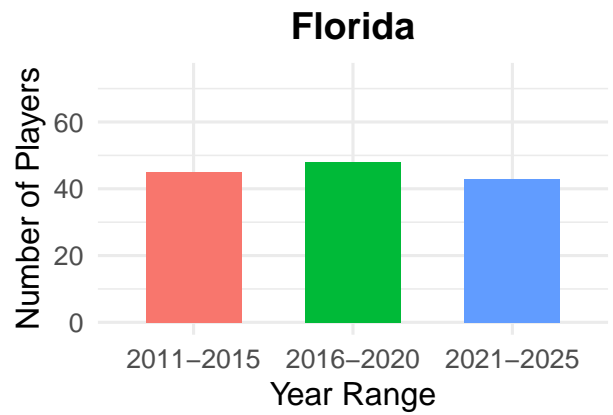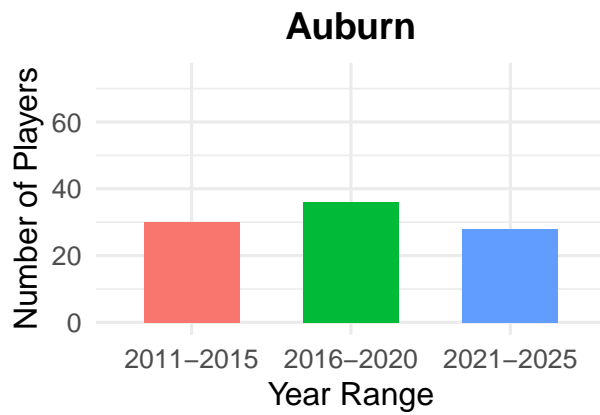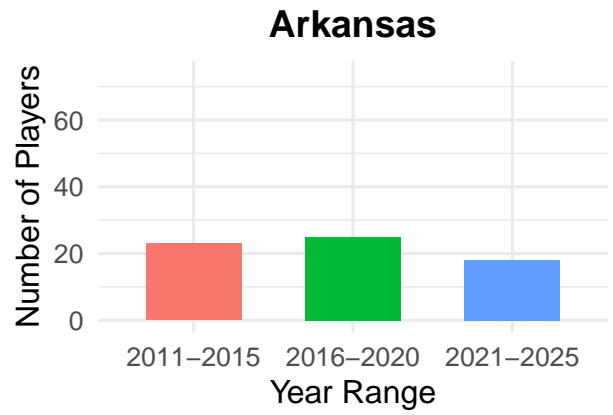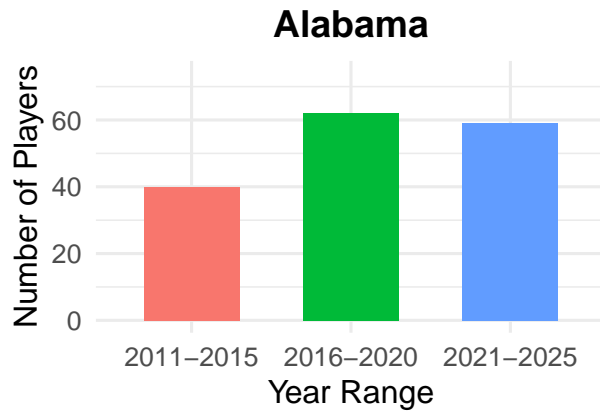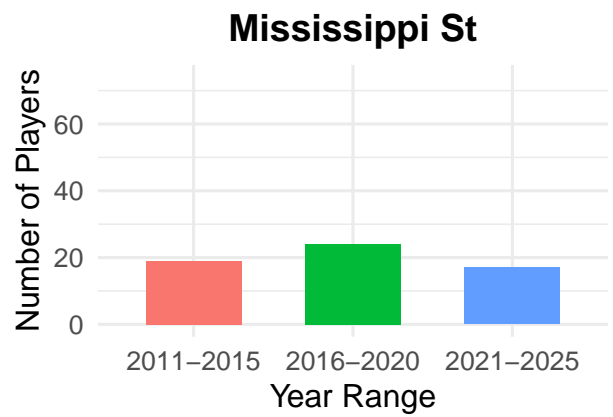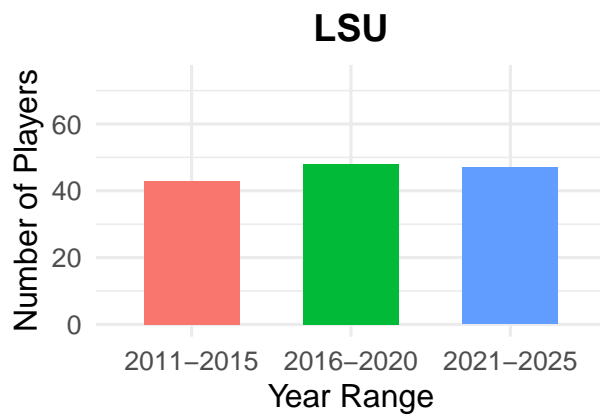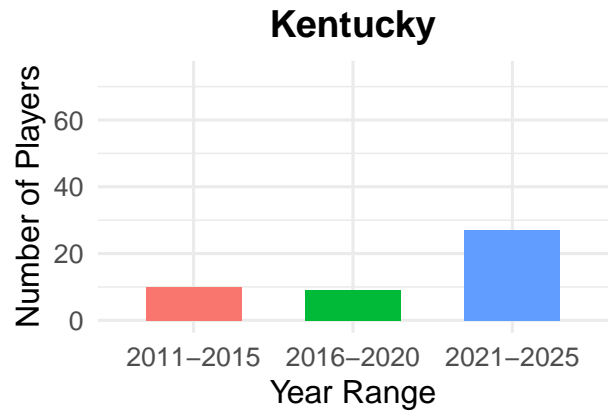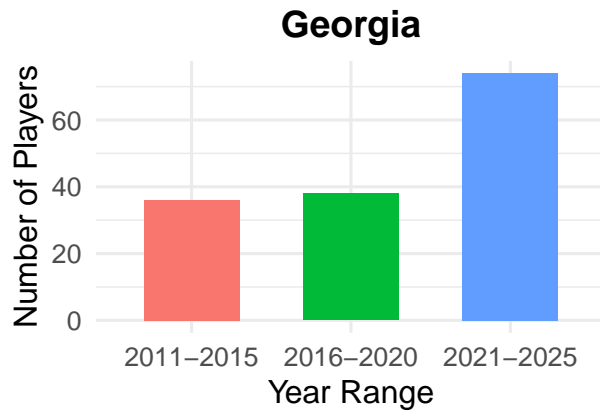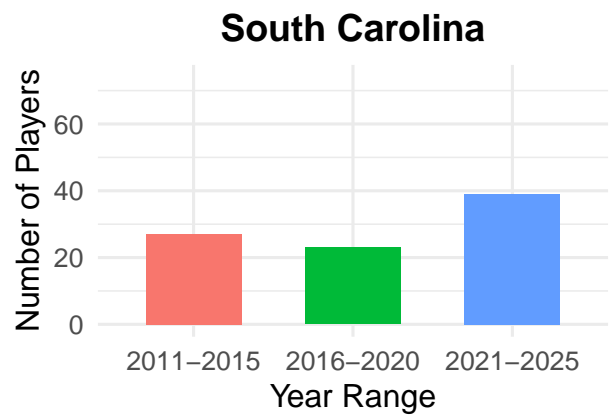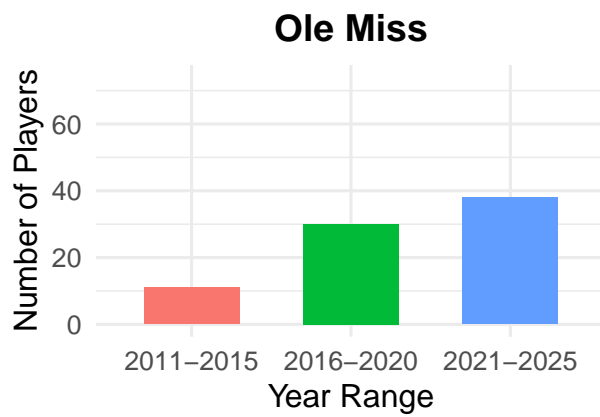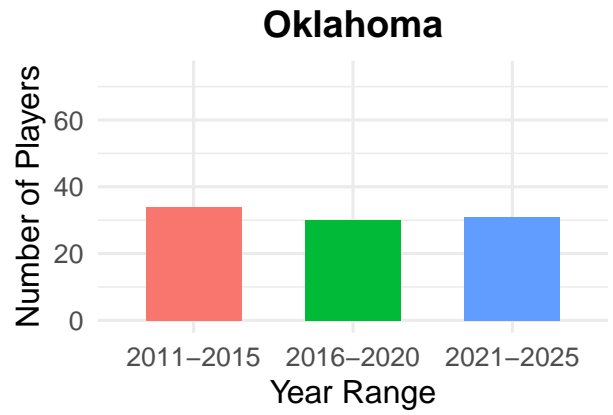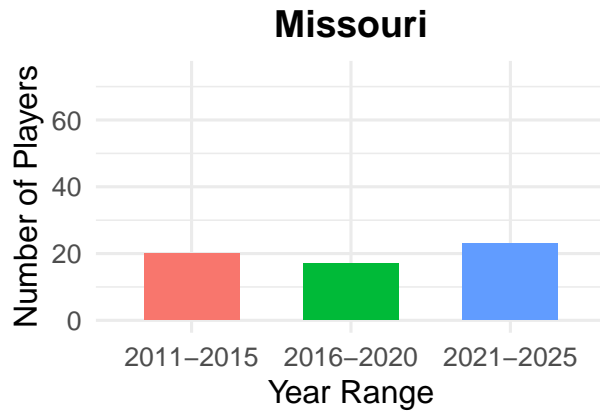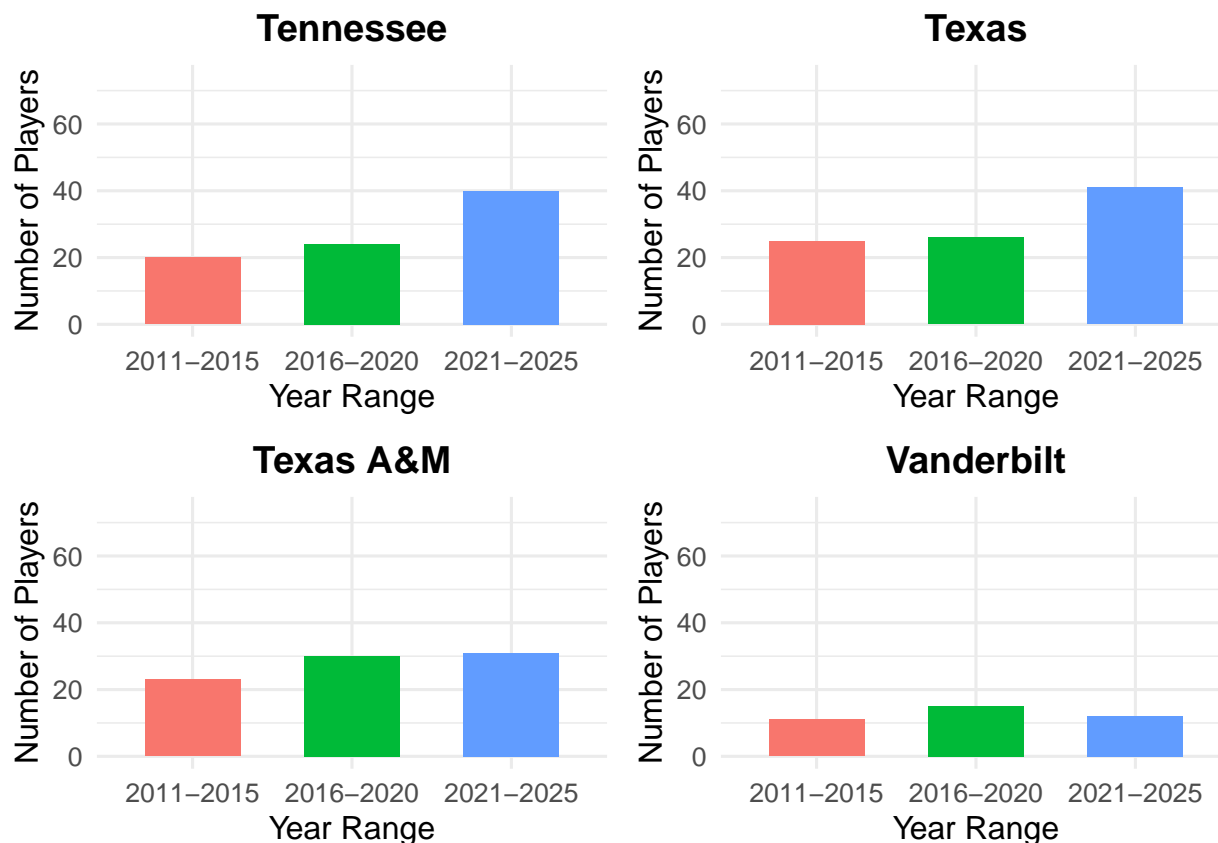
Missouri

Oklahoma

Ole Miss

South Carolina

Tennessee

Texas

Texas A&M

Vanderbilt


```r
# Load necessary libraries
library(tidyverse)
library(knitr)

# Read the CSV file (update the path if needed)
df <- read.csv("my_updated_dataset.csv")

# Convert 'Years Played (From-To)' into numeric start and end years
df <- df %>%
  separate(`Years.Played..From.To.`, into = c("Start", "End"), sep = "-", convert = TRUE)

# Define year ranges
df <- df %>%
  mutate(YearRange = case_when(
    Start >= 2011 & Start <= 2015 ~ "2011-2015",
    Start >= 2016 & Start <= 2020 ~ "2016-2020",
    Start >= 2021 & Start <= 2025 ~ "2021-2025",
    TRUE ~ NA_character_
  ))

# Count players per team and year range
team_summary <- df %>%
  filter(!is.na(YearRange)) %>%
  group_by(School, YearRange) %>%
  summarise(Players = n(), .groups = "drop")
```

```
# Pivot to wide format for numerical display (schools as rows, year ranges as columns)
team_summary_wide <- team_summary %>%
  pivot_wider(names_from = YearRange, values_from = Players, values_fill = 0) %>%
  arrange(School)  # Sort by school name for better readability

# Print the numerical data as a table
print(team_summary_wide)
```

```
## # A tibble: 16 x 4
##    School         '2011-2015' '2016-2020' '2021-2025'
##    <chr>                <int>       <int>       <int>
##  1 Alabama                 40          62          59
##  2 Arkansas                23          25          18
##  3 Auburn                  30          36          28
##  4 Florida                 45          48          43
##  5 Georgia                 36          38          74
##  6 Kentucky                10           9          27
##  7 LSU                     43          48          47
##  8 Mississippi St          19          24          17
##  9 Missouri                20          17          23
## 10 Oklahoma                34          30          31
## 11 Ole Miss                11          30          38
## 12 South Carolina          27          23          39
## 13 Tennessee               20          24          40
## 14 Texas                   25          26          41
## 15 Texas A&M               23          30          31
## 16 Vanderbilt              11          15          12
```

These preliminary visualizations support our future analysis by allowing us to compare the number of NFL players produced by various SEC colleges across different time periods. By examining the bar plots, we can identify early patterns of growth or decline in player representation for each school. In later stages of the analysis, we will compare these bar plots in greater detail to draw conclusions about which colleges have shown the most progress or decline in producing NFL players over the years.

Analysis

For the analysis, we plan to use descriptive statistics and visual models to track changes in the number of NFL players from each SEC school over the three time periods. We'll look at the overall trend for each school using a simple linear regression, which will help show whether the number of NFL players from that school is going up or down over time. We'll also use percentage change calculations to measure growth rates and rank the schools based on overall gains or declines. These methods should make it clear which schools have built a stronger NFL presence over time. We expect that powerhouse programs like Alabama and Georgia will show a steady growth, while schools like Vanderbilt and Auburn may show slight growth or even a decline.

Contributions

Omar Sepulveda: Gathered data sets, completed data section, code used to clean data set, code for the visuals, and completed analysis section