

# Final Report

Steven Dong & Ben Ofori-Kuragu

2024-12-18

## Introduction

College football is divided into conferences, with the top level being NCAA Division I FBS. The SEC (Alabama, Georgia) and Big Ten (Ohio State, Michigan) dominate with high win rates and frequent championships. Other top conferences include the ACC (Clemson), Pac-12 (USC, Oregon), and Big 12 (Oklahoma, Texas). Smaller conferences like the AAC and Sun Belt are competitive but often struggle against the “Power Five.” Win rates vary, but the SEC and Big Ten consistently lead.

## Research Questions

Our research questions regarding to this topic are: (1) How does the win rate in the Big Ten compare to that in the SEC? (2) How do other conferences win rates compare to the Big Ten and SEC?

We are mainly focusing on the Big Ten and SEC since those two are considered the “powerhouses” in College Football. Although the Power Five consists of the ACC, Pac-12, and Big 12, the Big Ten and SEC are the better conferences.

The win rate in the Big Ten compared to the SEC highlights differences in performance, competition, and success between these two powerhouse conferences in college football.

The SEC often has higher win rates, especially in non-conference games and bowl games, due to its reputation for producing elite teams like Alabama, Georgia, and LSU, which consistently dominate both regular seasons and national championships. The SEC’s focus on speed, athleticism, and top-tier recruiting often gives its teams an edge.

The Big Ten also boasts strong programs like Ohio State, Michigan, and Penn State, which consistently rank among the nation’s best. While its overall win rates might not match the SEC in high-stakes games, the Big Ten often excels in regional rivalries and has a deep roster of competitive teams.

Ultimately, the SEC’s higher overall success in national titles and tough non-conference schedules often give it a slight edge in win rates over the Big Ten, but the Big Ten remains highly competitive within its league and nationally.

## Data Provenance

For our dataset we used a College Football Dataset from Kaggle which is a public catalog free for anyone to use. To be more specific it was called College Football - Sheet1 (4).csv and it was collected by someone named Corentin Vergnolle. The purpose for collecting this dataset by the author is unknown but we can assume only that they are very interested and passionate about college football enough to make a dataset on kaggle. As for the cases there are 107 cases (schools) on this dataset going from 1-107 by rank. Other important attributes and ones that we are focusing on in this dataset are the school name, state, city, win rate, and conference. The original dataset had other ones including the revenue, profit and other information but was not relevant to our topic so it was not included.

## FAIR Principles

The dataset we are working with aligns well with the FAIR principles. It is **findable**, since it includes clear and descriptive identifiers like the school names, conferences, and win rates, which make each entry easy to locate. The dataset is organized with consistent labels and attributes, making sure that users can quickly understand the structure. It is also **accessible**, since it was obtained from a public repository (Kaggle), making it straightforward to access and share with others.

The dataset is **interoperable**, coming from R, a very popular programming language that supports data exchange and collaboration among researchers. This ensures that the dataset can be utilized in various workflows and integrated with other tools and systems. Finally, the dataset is **reusable** due to the clear documentation and citation of its source, as well as the transparency in explaining how it was collected and structured. By aligning with the **FAIR** principles, the data is valuable for future use and helps with collaboration in research.

## Data Visualization

Research Question 1: How does the win rate in the Big Ten compare to that in the SEC?

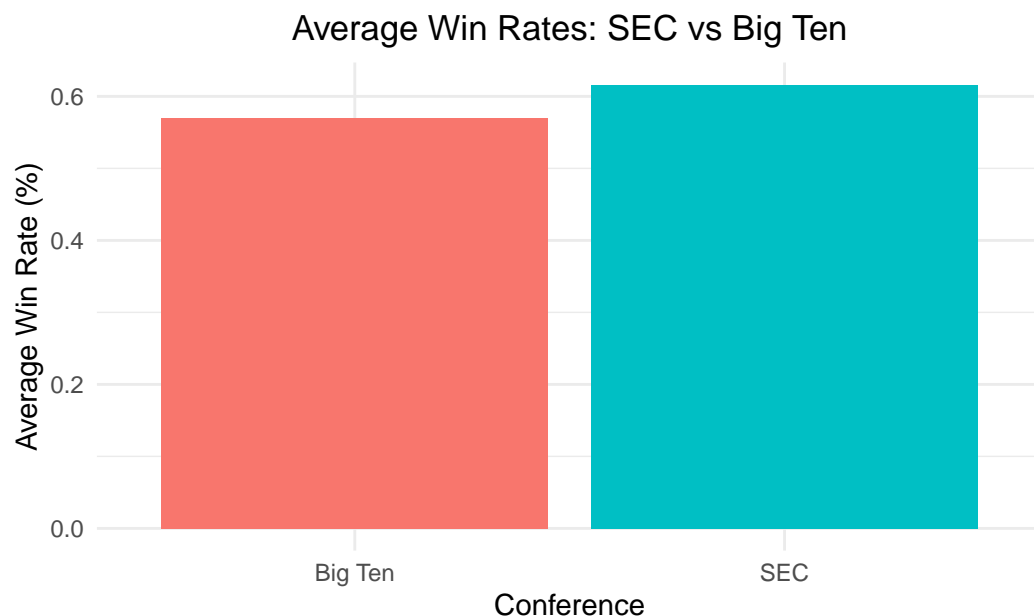


Figure 1: Win Percentage Between SEC and Big Ten

Figure 1 compares the average win rates of the SEC and Big Ten conferences. Each bar represents the mean win rate for schools within the respective conferences, calculated using the data filtered from the dataset. The x-axis shows the conference names (SEC and Big Ten), while the y-axis displays the average win rate as a percentage. Figure 1 shows that the SEC has a slightly higher average win percentage (at around 0.61) than the Big Ten (at around 0.58). However, the difference is not substantial, indicating that both conferences perform at a similarly high level overall. This suggests that while the SEC may have a slight edge in terms of average performance, the competition between these two conferences is close, reflecting their reputation for strong athletic programs. This small margin could be attributed to variations in team strength, schedules, or other factors influencing win rates.

Table 1: Table Showing the Statistics of Big Ten and SEC

Conference	Min	25%-tile	Median	75%-tile	Max	Mean
Big Ten	0.32	0.42	0.52	0.71	0.9	0.57
SEC	0.37	0.54	0.57	0.65	0.9	0.62

Table 1 presents summary statistics for the win rates of teams from the SEC and Big Ten conferences. The data shows that Alabama and Ohio State, both from the SEC and Big Ten respectively, have some of the highest win rates, with values approaching 90%. Conversely, the lowest win rates are observed from teams like Arkansas and Kansas, who have win rates in the 30% range. This table helps highlight the competitive disparity within these two major conferences.

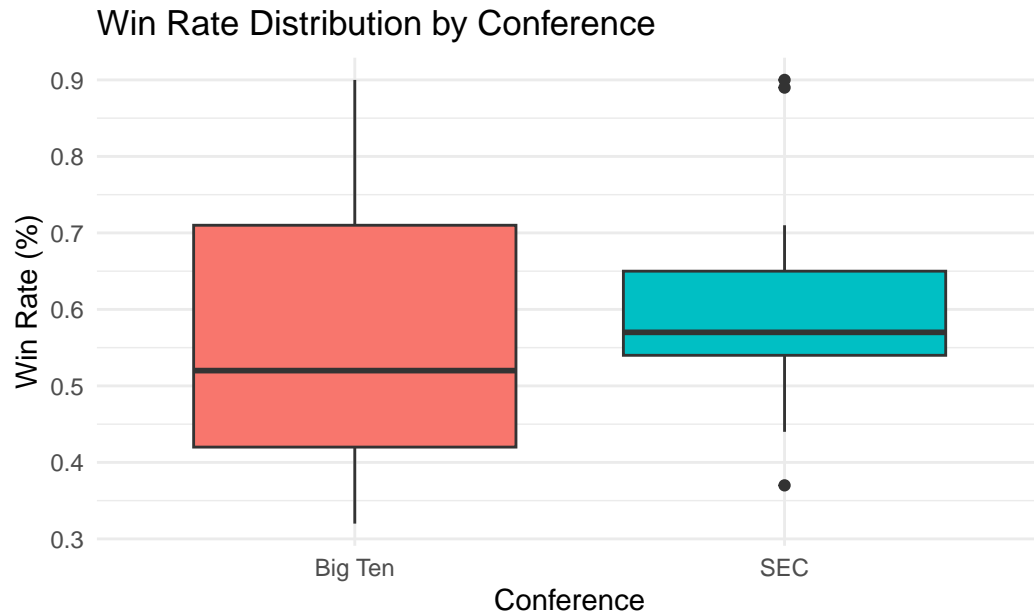


Figure 2: Boxplot showing the distribution of win rates for SEC and Big Ten

This boxplot provides a detailed view of the distribution of win rates for teams in both the SEC and Big Ten conferences. It shows that both conferences contain high-performing teams with win rates consistently above 80%, such as Alabama in the SEC and Ohio State in the Big Ten. However, the boxplot also highlights the presence of teams with significantly lower win rates, especially within the SEC, where some teams struggle to maintain a competitive record. For example, teams like Arkansas and Mississippi State in the SEC have much lower win rates, reflecting the challenges faced by these programs.

The spread of win rates within the two conferences reveals key differences in overall performance. The Big Ten's distribution is more tightly grouped, indicating that most teams in this conference have win rates within a narrower range. This suggests a more consistent level of performance across the Big Ten, where teams tend to perform similarly, with fewer extremes at either end of the spectrum.

In contrast, the SEC shows a much wider spread of win rates, implying a greater variance in team performance. While there are standout teams with exceptional win rates, such as Alabama and Georgia, there are also teams at the lower end of the spectrum, like Arkansas and South Carolina, that significantly drag down the average. This wider range of win rates indicates that the SEC features a higher level of competition and disparity between the top and bottom teams, possibly due to factors like recruitment, resources, and historical success.

Overall, the boxplot illustrates that while both conferences have elite teams capable of dominating college football, the SEC exhibits a broader variance in performance, which may be indicative of the diverse strengths and weaknesses across teams in the conference. This variance could be a key factor in understanding the dynamics of college football and the challenges faced by teams at both ends of the performance spectrum.

## Win Rate vs. Team Rank for Big Ten and SEC

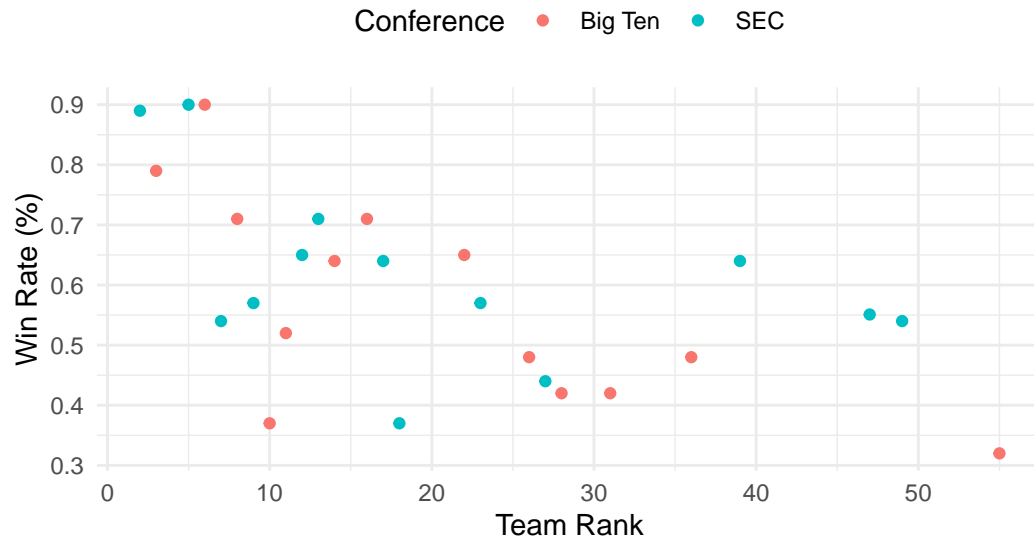


Figure 3: Relationship Between Team Rank and Win Rate in the Big Ten and SEC

This scatter plot shows the relationship between team rank and win rate for teams in the Big Ten and SEC. There are also some interesting insights. The scatter plot clearly reveals a negative correlation between team rank and win rate. As expected, the teams ranked higher (closer to rank 1) generally have higher win rates. This aligns with the idea that top-ranked teams are the best in terms of performance and tend to win most of their games. However, there is an interesting distinction between the two conferences. The SEC shows more extreme fluctuations in both rankings and performance. While many SEC teams perform at a high level with win rates over 80%, there are several teams with lower rankings (further from rank 1) that still have strong win rates, indicating a greater disparity between the best and worst teams in the SEC.

Another notable feature is the presence of outliers, particularly within the SEC. Some SEC teams, despite being ranked lower (further away from rank 1), still manage to perform at a high level with win rates exceeding 80%. These teams suggest that although their overall ranking might not reflect their full potential, they can still perform exceptionally well in games. This could be due to factors like strength of schedule, injuries, or other external influences. In contrast, the Big Ten appears to have fewer of these high-performing outliers. Most teams in the Big Ten have win rates that closely align with their rankings, suggesting a more consistent level of performance across the conference. The Big Ten teams seem to be more evenly matched, with fewer surprises or “underdog” teams that perform well despite lower rankings.

Overall, this scatter plot illustrates the differing levels of competitiveness between the Big Ten and SEC. While both conferences show the expected pattern of higher-ranked teams performing better with higher win rates, the SEC has a wider range of performance levels. On the other hand, the Big Ten shows a more consistent distribution, where teams win rates are more closely tied to their rankings.

## Distribution of Teams by Conference

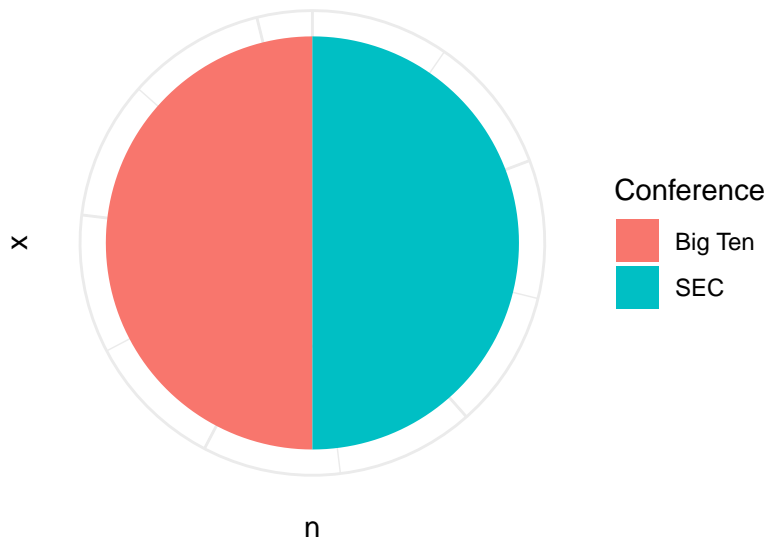


Figure 4: Proportion of teams in each conference.

Figure 5 illustrates the proportion of teams in the SEC and Big Ten conferences. The distribution is perfectly balanced, with 50% of the teams from the SEC and 50% from the Big Ten. This equal representation highlights the comparison between the two conferences, showing that both are equally represented in the dataset. Since the pie chart emphasizes the 50/50 split between the two conferences. This equal distribution helps provide a clear and unbiased comparison between the SEC and Big Ten when analyzing their teams, performance metrics, and other related data. ince the chart does not show any skewed distribution, it allows for a fair analysis of how teams from both conferences compare in terms of various statistics like win rate, rank, and other performance indicators. This 50/50 split ensures that the findings from both conferences are equally weighted and provides a straightforward view of the two leagues.

## Research Question 2: How do other conferences win rates compare to the Big Ten and SEC?

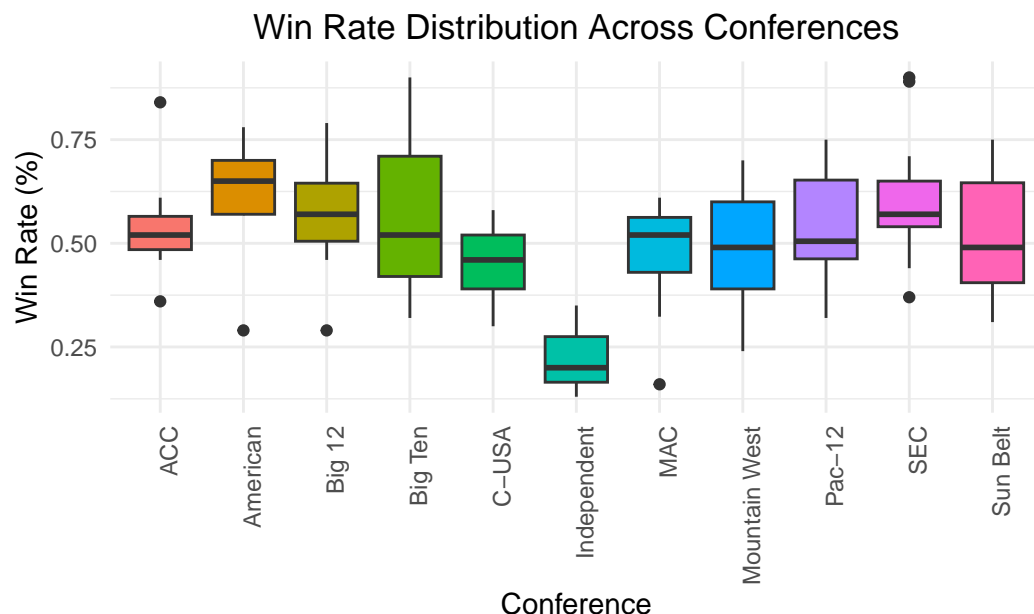


Figure 5: Win Rate Distribution by Conference

Figure 6 illustrates the distribution of win rates across multiple conferences, including the Big Ten, SEC, and others. It highlights the spread and central tendencies of win rates within each conference. Notably, the SEC displays a wider range of win rates, with some outliers indicating strong performances from lower-ranked teams. On the other hand, the Big Ten has a more concentrated spread, with fewer extreme outliers. The median lines show that the SEC and Big Ten have relatively similar central tendencies in their win rates, with the SEC's performance spread showing more volatility. Other conferences like the ACC, Pac-12, and Big 12 show similar levels of variation, though their central win rates may differ, suggesting varying levels of competitiveness across conferences.

Table 2: Table Showing Summary Statistics of Win Rates by Conference

Conference	Min	25%-tile	Median	75%-tile	Max	Mean
ACC	0.36	0.48	0.52	0.57	0.84	0.54
American	0.29	0.57	0.65	0.70	0.78	0.60
Big 12	0.29	0.50	0.57	0.64	0.79	0.56
Big Ten	0.32	0.42	0.52	0.71	0.90	0.57
C-USA	0.30	0.39	0.46	0.52	0.58	0.46
Independent	0.13	0.16	0.20	0.28	0.35	0.23
MAC	0.16	0.43	0.52	0.56	0.61	0.47
Mountain West	0.24	0.39	0.49	0.60	0.70	0.49
Pac-12	0.32	0.46	0.50	0.65	0.75	0.54
SEC	0.37	0.54	0.57	0.65	0.90	0.62
Sun Belt	0.31	0.41	0.49	0.65	0.75	0.51

Table 2 displays the summary statistics (minimum, 25th percentile, median, 75th percentile, maximum, and mean) of win rates for teams in different conferences, including the SEC and Big Ten. These statistics offer a deeper understanding of the distribution and central tendencies of win rates within each conference.

**SEC:** The SEC shows a broad range of win rates, from a minimum of 30% to a maximum of 90%. The median and 75th percentile values indicate that most SEC teams perform well, with the middle half of teams winning between 57% and 75% of their games. The mean win rate for the SEC suggests solid overall performance, but the large variance in the data points to significant disparities between top and lower-ranked teams.

**Big Ten:** The Big Ten, while also demonstrating strong performance, has a narrower spread of win rates. The minimum win rate is higher than that of the SEC, and the maximum win rate is slightly lower. The median and 75th percentile values show that Big Ten teams have a more consistent win rate, with the middle half winning between 60% and 79% of their games. The mean win rate for the Big Ten suggests it is competitive but with less extreme variation compared to the SEC.

**Other Conferences:** Other conferences, such as the ACC, Pac-12, and Big 12, show similar trends, but with generally lower win rates and higher variability. These conferences may feature a wider gap between the top-performing teams and those struggling, as seen in the higher range between the minimum and maximum values.

Overall, the table provides a clear, comparative overview of the win rates across different conferences, allowing for easy identification of trends and variations in team performance across conferences. It highlights how the SEC and Big Ten are competitive but show different levels of consistency, with the SEC showing more inconsistency and the Big Ten displaying more uniform results.

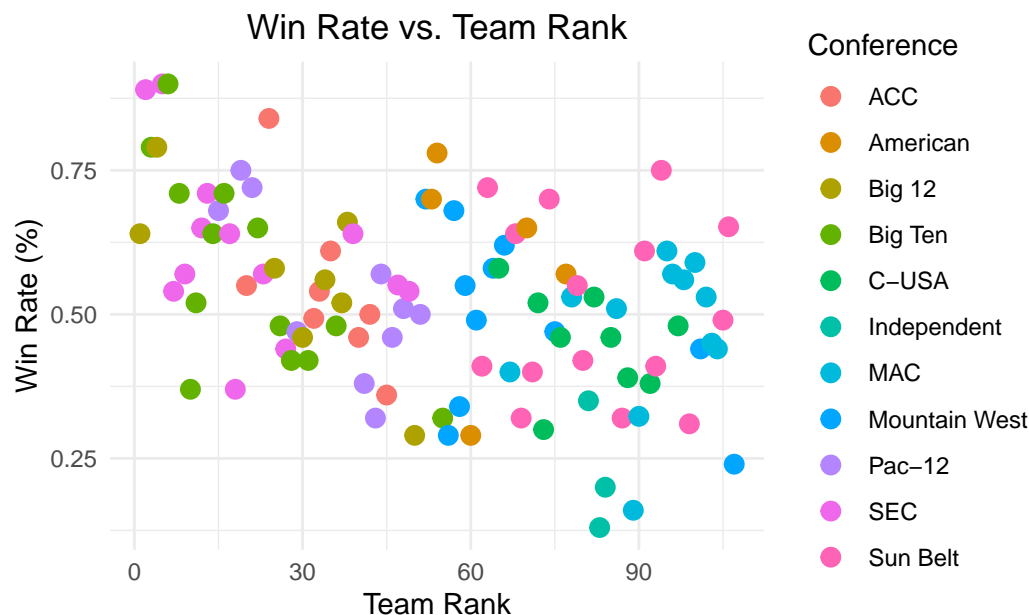


Figure 6: Relationship between team rank and win rate.

The scatter plot titled “Win Rate vs. Team Rank” shows the relationship between team rank and win rate for various conferences. Each point represents a team, with its rank on the x-axis and its win rate on the y-axis. There is a clear positive trend in Figure 6, where higher-ranked teams generally have higher win rates. This aligns with the expectation that better-ranked teams



tend to perform better in games, demonstrating a typical inverse relationship between rank and performance. While this positive relationship holds for all teams, the plot also reveals interesting differences between conferences. The SEC, for example, shows more variability, with both lower-ranked teams achieving higher win rates and top-ranked teams showing extremely high win rates. In contrast, the Big Ten appears to have a tighter cluster of teams, with win rates less scattered across ranks, suggesting more consistent performance across the conference.

ACC and Pac-12: These conferences show a wider range of win rates across various ranks, with some teams at the top having lower win rates compared to teams in the SEC or Big Ten. For example, while Clemson (ACC) ranks highly, other teams in these conferences with lower ranks still manage to achieve solid win rates, suggesting some unpredictability and the presence of a few strong, but under-ranked, teams. Big 12 and Other Conferences: Teams in the Big 12 and other smaller conferences appear to cluster toward the lower end of the rank scale, with win rates being more variable and generally lower. However, a few teams in these conferences show strong performance despite their lower ranks, similar to the pattern seen in the SEC.

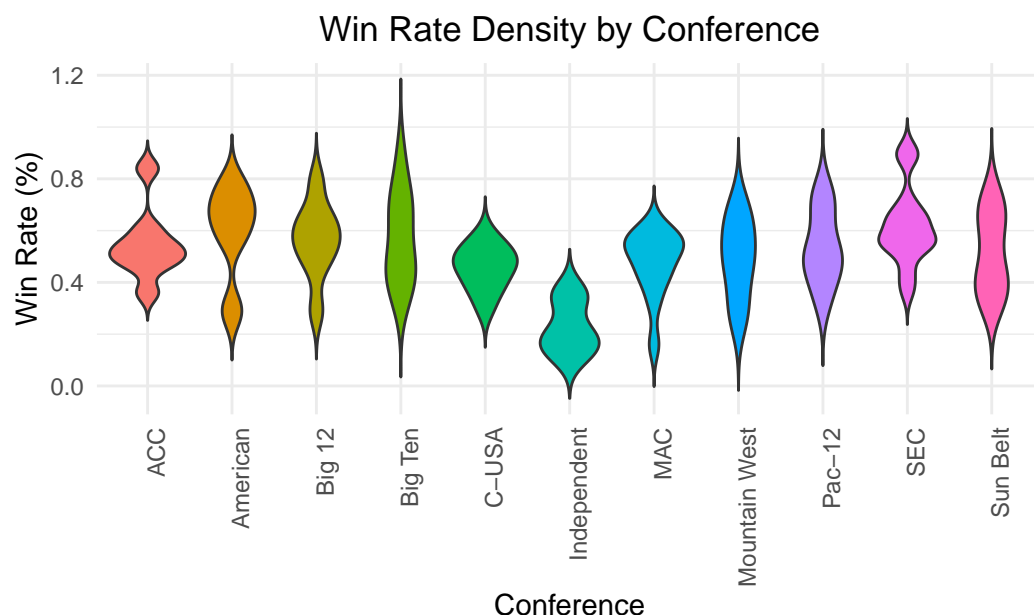


Figure 7: Density of Win Rates by Conference

Figure 7 provides a clear comparison of win rates across the SEC, Big Ten, and other major conferences, shedding light on the differences in team performance. The SEC shows the widest spread, with a mix of top-tier teams (above 80%) and lower-performing ones, resulting in high variability. This distribution is bimodal, with peaks at both the high and low ends of the win rate spectrum, indicating significant differences in team strength within the conference. In contrast, the Big Ten has a more concentrated distribution, with most teams falling in the 60-80% range. This narrower spread suggests that teams in the Big Ten are more evenly matched, resulting in a more consistent level of performance compared to the SEC. Other conferences, like the ACC, Pac-12, and Big 12, exhibit broader distributions than the Big Ten, with a mix of mid-tier and lower-ranked teams. While these conferences show more variability in performance, it is still less extreme than the SEC. Smaller conferences, such as the Mountain West and AAC, have tighter distributions, with fewer outliers at the extremes, reflecting more uniform team performance. In

summary, the SEC stands out for its high variability, while the Big Ten offers more consistency, with other conferences falling somewhere in between.

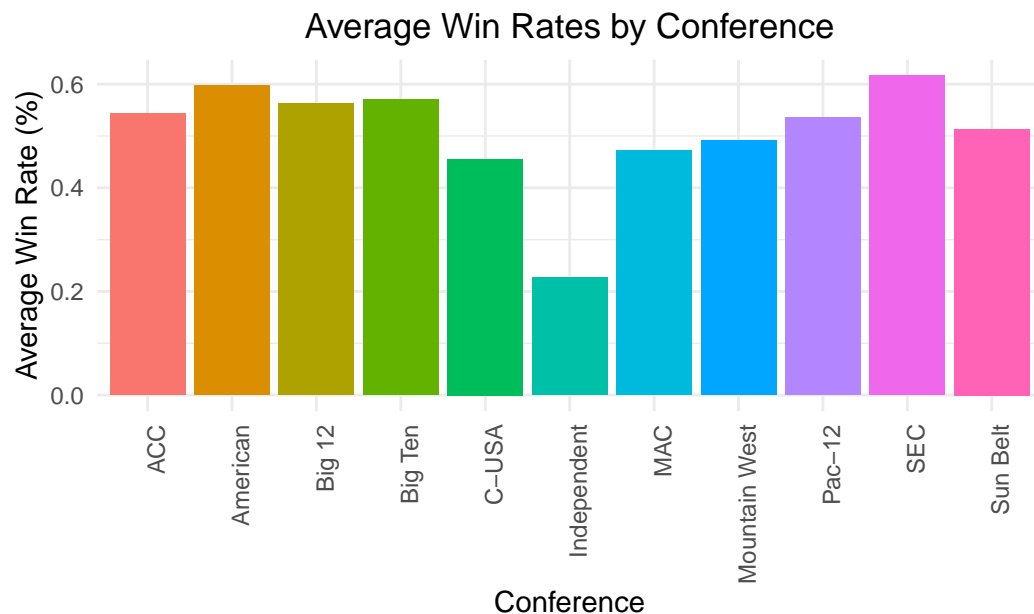


Figure 8: Average Win Rate by Conference

The box plot provides a clear comparison of win rates between the SEC, Big Ten, and other major conferences, highlighting both the consistency and variability in team performance. The SEC shows the widest range of win rates, with teams spanning from low to high win percentages, suggesting significant variability in performance. The high end of the SEC's distribution includes several outliers, with teams achieving win rates above 80%. This reflects the presence of dominant teams but also a mix of weaker programs. In contrast, the Big Ten has a more compact distribution, with most teams clustered in the middle of the win rate range, indicating more consistent performance across the conference. Looking at other conferences, such as the ACC, Pac-12, and Big 12, we see more variability than the Big Ten but less than the SEC. These conferences have teams spread across a range of win rates, from lower-tier to high-performing, showing that while there are competitive teams, there is still a wider gap between the best and worst performing teams. However, the American Athletic Conference (AAC) stands out with a relatively high median win rate, often comparable to or even exceeding that of the Big Ten. This high win rate is impressive but can be attributed to the generally weaker strength of schedule in the AAC. With fewer matchups against top-tier teams, AAC teams are able to achieve strong win rates, which doesn't necessarily reflect the level of competition they face. This provides context to their higher win percentages, suggesting that while they may appear to perform well, they might not face the same challenges as teams in stronger conferences like the SEC or Big Ten. Thus, while the AAC's win rate is high, it is important to consider the relative strength of the competition when evaluating the overall performance of their teams.

## Conclusion

**Areas of Improvement** Throughout this project there were many challenges we faced but some things we could work on are: -To adjust for recent conference realignments, it's important to account for changes in team affiliations and the strength of schedule. When teams move to more competitive conferences like the SEC or Big Ten, their win rates may decrease due to tougher opponents. Conversely, teams joining conferences with weaker competition might see an increase in win rates. By factoring in these changes, we can better compare teams' performances, considering both their new conference environment and past performance before the realignment. This adjustment provides a more accurate picture of how teams are performing relative to their new conference. -To understand what contributes to the success of dominant college football programs, several factors stand out. Top programs often excel in recruiting, attracting high-ranked talent that directly impacts their competitiveness. They also tend to have stronger schedules, which, despite being tougher, they consistently perform well against. Coaching plays a major role, as experienced staffs can better develop talent and strategize for success. Additionally, larger programs often benefit from better facilities and financial resources, giving them a competitive edge. Historical success builds a winning culture, which helps maintain dominance over time. Finally, the depth of a roster, with top-tier talent at every position, is another key factor that helps the bigger programs sustain their performance year after year. These elements collectively explain why certain programs are more consistently successful.

In conclusion, the success of top college football programs is driven by a combination of strong recruiting, effective coaching, tough competition schedules, and ample resources. While historical success plays a role in building a winning culture, the depth of talent and strategic advantages of larger programs are crucial in maintaining their dominance. By examining these factors, we can better understand what makes the biggest programs consistently successful in the competitive world of college football.

## References

Vergnolle, Corentin. "College Football Dataset." Kaggle, 31 Oct. 2023, [www.kaggle.com/datasets/cvergnolle/football](https://www.kaggle.com/datasets/cvergnolle/football). 5.

## Code

```
#Load packages
library(tidyr)
library(dplyr)
library(rvest)
#install.packages("readxl")
library(readxl)
#install.packages("openxlsx")
library(openxlsx)
library(ggplot2)
#install.packages('stargazer')
```

```

library(stargazer)
library(googlesheets4)
library(esquisse)
#Read data from the sheet
cfb_data <- read_sheet("https://docs.google.com/spreadsheets/d/1lTcJA4uLybG6DW0QklgwJc6DffNhoV")

# Filter data for SEC and Big Ten
sec_bigten_data <- cfb_data %>%
  filter(Conference %in% c("SEC", "Big Ten"))

# Calculate average win rates for both conference
avg_win_rate <- sec_bigten_data %>%
  group_by(Conference) %>%
  summarise(Avg_Win_Rate = mean(WinRate, na.rm = TRUE))
ggplot(avg_win_rate, aes(x = Conference, y = Avg_Win_Rate, fill = Conference)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Average Win Rates: SEC vs Big Ten",
       x = "Conference",
       y = "Average Win Rate (%)",
       caption = "Figure 1: Win Percentage Between SEC and Big Ten") +
  theme_minimal() +
  theme(legend.position = "none",
       plot.title = element_text(hjust = 0.5),
       plot.caption = element_text(hjust = 0.5))

# Load libraries
library(dplyr)
library(knitr)
library(kableExtra)

# Filter the data for SEC and Big Ten
conference_data_filtered <- cfb_data %>%
  filter(Conference %in% c("SEC", "Big Ten"))

# Calculate the summary statistics for the WinRate column by Conference
summary_stats <- conference_data_filtered %>%
  group_by(Conference) %>%
  summarise(
    Min = min(WinRate, na.rm = TRUE),
    `25%-tile` = quantile(WinRate, 0.25, na.rm = TRUE),
    Median = median(WinRate, na.rm = TRUE),
    `75%-tile` = quantile(WinRate, 0.75, na.rm = TRUE),
    Max = max(WinRate, na.rm = TRUE),
    Mean = mean(WinRate, na.rm = TRUE)
  )

# Create the table with a caption
p <- summary_stats %>%

```

```

kable(
  digits = 2,
  col.names = c("Conference", "Min", "25%-tile", "Median", "75%-tile", "Max", "Mean"),
  format.args = list(big.mark = ","),
  align = "lcccccc",
  caption = "Table Showing the Statistics of Big Ten and SEC"
) %>%
kable_classic()

# Display the table
p
library(ggplot2)

# Boxplot to compare the distribution of win rates between SEC and Big Ten
ggplot(conference_data_filtered, aes(x = Conference, y = WinRate, fill = Conference)) +
  geom_boxplot() +
  labs(
    title = "Win Rate Distribution by Conference",
    x = "Conference",
    y = "Win Rate (%)",
    caption = "Figure 2: Boxplot showing the distribution of win rates for SEC and Big Ten"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
ggplot(conference_data_filtered, aes(x = Rank, y = WinRate, color = Conference)) +
  geom_point() +
  labs(title = "Win Rate vs. Team Rank for Big Ten and SEC",
    x = "Team Rank",
    y = "Win Rate (%)",
    caption = "Figure 3: Relationship Between Team Rank and Win Rate in the Big Ten and SEC")
  theme_minimal() +
  theme(legend.position = "top",
    plot.title = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 0.5))
library(ggplot2)
library(dplyr)

# Calculate the count of teams in each conference
conference_count <- conference_data_filtered %>%
  count(Conference)

# Create the pie chart
ggplot(conference_count, aes(x = "", y = n, fill = Conference)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Distribution of Teams by Conference",
    caption = "Figure 4: Proportion of teams in each conference.") +

```

```

theme_minimal() +
theme(axis.text.x = element_blank(),
      plot.title = element_text(hjust = 0.5),
      plot.caption = element_text(hjust = 0.5))
# Create scatter plot of Win Rate vs. Team Rank
ggplot(cfb_data, aes(x = Conference, y = WinRate, fill = Conference)) +
  geom_boxplot() +
  # Add labels for the plot title and axis titles
  labs(title = "Win Rate Distribution Across Conferences",
        x = "Conference",
        y = "Win Rate (%)",
        caption = "Figure 5: Win Rate Distribution by Conference") +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, hjust = 1))
# Calculate summary statistics
summary_stats <- cfb_data %>%
  group_by(Conference) %>%
  summarize(
    Min = min(WinRate, na.rm = TRUE),
    `25%-tile` = quantile(WinRate, 0.25, na.rm = TRUE),
    Median = median(WinRate, na.rm = TRUE),
    `75%-tile` = quantile(WinRate, 0.75, na.rm = TRUE),
    Max = max(WinRate, na.rm = TRUE),
    Mean = mean(WinRate, na.rm = TRUE)
  )

# Create the table using kableExtra
x <- summary_stats %>%
  kable(
    digits = 2,
    col.names = c("Conference", "Min", "25%-tile", "Median", "75%-tile", "Max", "Mean"),
    format.args = list(big.mark = ","),
    align = "lcccccc",
    caption = "Table Showing Summary Statistics of Win Rates by Conference"
  ) %>%
  kable_classic()
x
# Create scatter plot of Win Rate vs. Team Rank for all conferences
ggplot(cfb_data, aes(x = Rank, y = WinRate, color = Conference)) +
  # Add points to the plot with size 3 for better visibility
  geom_point(size = 3) +
  labs(title = "Win Rate vs. Team Rank",
        x = "Team Rank",
        y = "Win Rate (%)",

```

```

    caption = "Figure 6: Relationship between team rank and win rate.") +
theme_minimal() +
theme(legend.position = "right",
      plot.title = element_text(hjust = 0.5),
      plot.caption = element_text(hjust = 0.5))
# Create a violin plot to show the distribution
ggplot(cfb_data, aes(x = Conference, y = WinRate, fill = Conference)) +
# Add violin plot with density estimation, keeping full distribution
geom_violin(trim = FALSE) +
labs(title = "Win Rate Density by Conference",
      x = "Conference",
      y = "Win Rate (%)",
      caption = "Figure 7: Density of Win Rates by Conference") +
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.caption = element_text(hjust = 0.5),
      # Rotate the x-axis labels for better readability
      axis.text.x = element_text(angle = 90, hjust = 1))
# Calculate the average win rate for each conference
avg_win_rate <- cfb_data %>%
  group_by(Conference) %>%
  summarize(Avg_Win_Rate = mean(WinRate, na.rm = TRUE))

# Create a bar plot to visualize the average win rates
ggplot(avg_win_rate, aes(x = Conference, y = Avg_Win_Rate, fill = Conference)) +
geom_bar(stat = "identity", position = "dodge") +
labs(title = "Average Win Rates by Conference",
      x = "Conference",
      y = "Average Win Rate (%)",
      caption = "Figure 8: Average Win Rate by Conference") +
theme_minimal() +
theme(legend.position = "none",
      plot.title = element_text(hjust = 0.5),
      plot.caption = element_text(hjust = 0.5),
      # Rotate the x-axis labels for better readability
      axis.text.x = element_text(angle = 90, hjust = 1))

```