

R Football Analytics Basics

Jake Blumengar

2025-07-25

```
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
library(ggplot2)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## vforcats    1.0.0    vstringr    1.5.1
## vlubridate  1.9.4    vtibble      3.2.1
## vpurrr      1.0.4    vtidyrm     1.3.1
## vreadr       2.1.5

## -- 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

library(nflplotR)
library(nflreadr)
library(gganimate)

setwd("/Users/jakeblumengarten/git_repository/nfl_analysis")

season_24 <- read.csv("2024.csv")
```

Yards By week for The Chicago Bears

```
chicago_bears_24 <- season_24 %>%
  filter(posteam == "CHI") %>%
  select(week, yards_gained) %>%
  group_by(week) %>%
  summarize(yards_per_week = sum(yards_gained, na.rm = TRUE))

chicago_bears_24$winperc <- c((1/1), (1/2), (1/3), (2/4), (3/5), (4/6), (4/7), (4/8), (4/9), (4/10), (4/11)
```

```

# Team branding positions
wordmarks <- tibble(
  x = c(3.5, 17.5),
  y = c(100, 400),
  team_abbr = "CHI"
)

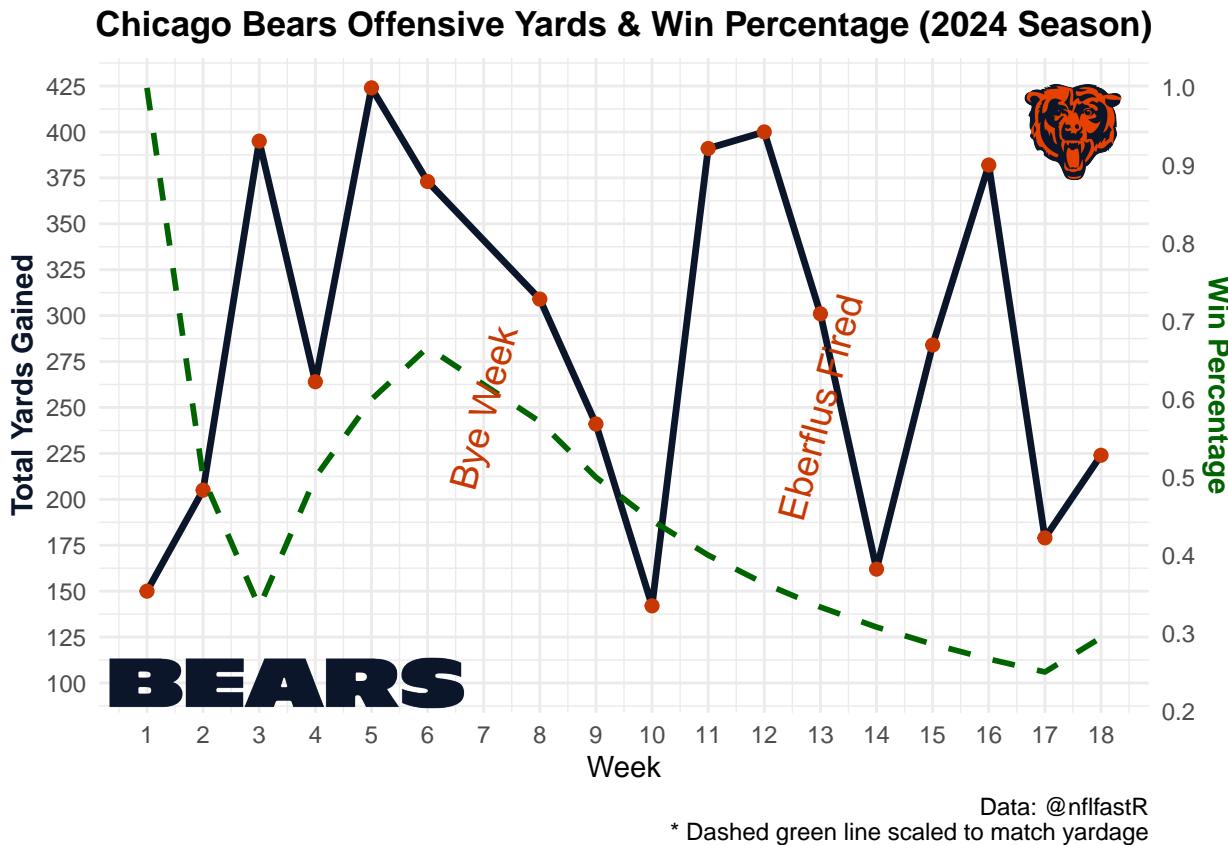
chicagoypw <- ggplot(chicago_bears_24, aes(x = week)) +
  geom_line(aes(y = yards_per_week), color = "#0B162A", size = 1.2) +
  geom_point(aes(y = yards_per_week), color = "#C83803", size = 2) +
  geom_line(aes(y = winperc * max(yards_per_week)),
            color = "darkgreen", linetype = "dashed", size = 1) +
  scale_y_continuous(
    name = "Total Yards Gained",
    breaks = seq(100, 500, 25),
    sec.axis = sec_axis(
      trans = ~ . / max(chicago_bears_24$yards_per_week),
      name = "Win Percentage",
      breaks = seq(0, 1.0, 0.1)
    )
  ) +
  scale_x_continuous(breaks = seq(1, 18, 1)) +
  labs(
    title = "Chicago Bears Offensive Yards & Win Percentage (2024 Season)",
    x = "Week",
    caption = "Data: @nflfastR\n* Dashed green line scaled to match yardage"
  ) +
  annotate("text", x = 13, y = 250, label = "Eberflus Fired",
           color = "#C83803", size = 5, angle = 75) +
  annotate("text", x = 7, y = 250, label = "Bye Week",
           color = "#C83803", size = 5, angle = 75) +
  geom_nfl_wordmarks(data = wordmarks[1, ],
                      aes(x = x, y = y, team_abbr = team_abbr), width = 0.35) +
  geom_nfl_logos(data = wordmarks[2, ],
                 aes(x = x, y = y, team_abbr = team_abbr), width = 0.1) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.title.y.left = element_text(color = "#0B162A", face = "bold"),
    axis.title.y.right = element_text(color = "darkgreen", face = "bold")
  )

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## Warning: The `trans` argument of `sec_axis()` is deprecated as of ggplot2 3.5.0.
## i Please use the `transform` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

```
ggsave("chicago_yards_by_week_R.png",
       plot = chicagopw,
       path = "/Users/jakeblumengarten/git_repository/nfl_analysis/plots/",
       width = 10,
       height = 10,
       dpi = 300,
       bg = "white")
chicagopw
```



Average Yards Per Game by Team

```
library(forcats)
yards_pg_pt <- season_24 %>%
  filter(!is.na(yards_gained), !is.na(postteam)) %>%
  filter(!(week %in% 19:22)) %>%
  group_by(postteam) %>%
  summarize(avgyards = (sum(yards_gained, na.rm = TRUE) / 17), .groups = 'drop') %>%
  filter(postteam != "")
yards_pg_pt$winpct <- c(.471, .471, .706, .765, .294, .294, .529, .176, .412, .588, .882, .647, .588, .4

yards_pg_pt <- yards_pg_pt %>%
  mutate(postteam = fct_reorder(postteam, avgyards, .desc = TRUE))

teamsyypg <- ggplot(yards_pg_pt, aes(x = avgyards, y = postteam)) +
  geom_col(aes(fill = postteam, color = postteam), width = .8) +
  geom_nfl_logos(aes(x = 0, y = postteam, team_abbr = postteam),
```

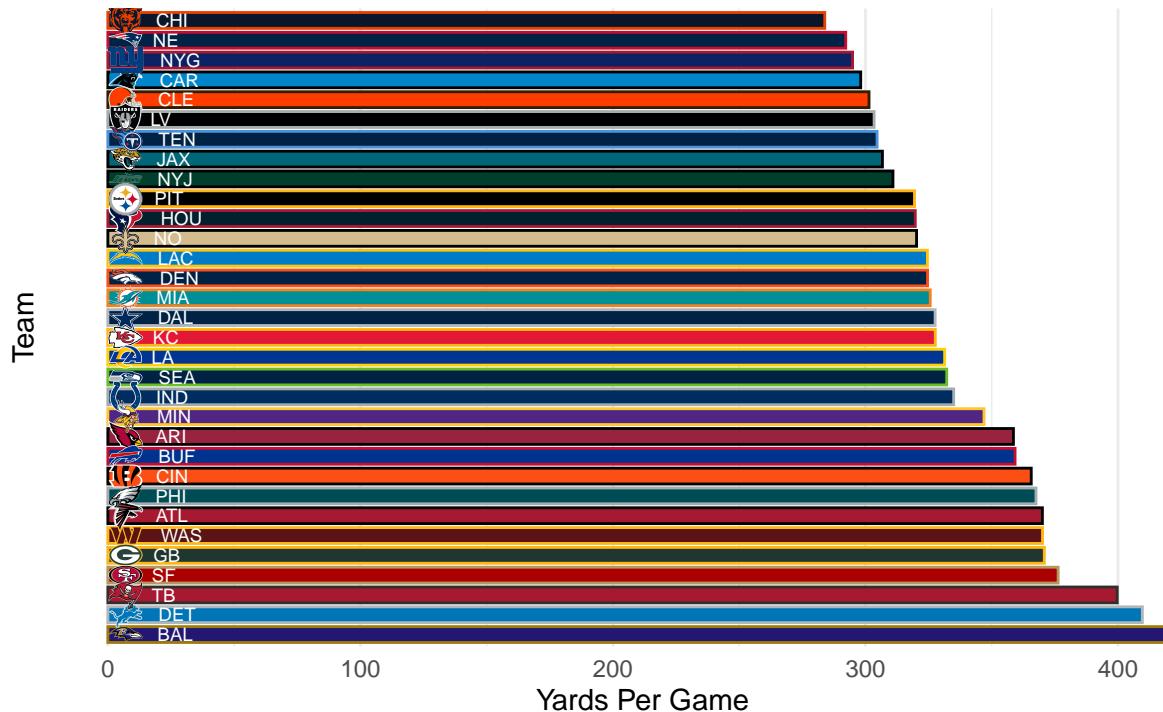
```

        width = 0.03,
        hjust = 0,
        vjust = 0.5) +
geom_text(aes(x = 0, y = posteam, label = posteam),
          hjust = -0.5,
          vjust = 0.5,
          nudge_x = max(yards_pg_pt$avgyards, na.rm = TRUE) * 0.03,
          color = "white", size = 2.5) +
labs(
  x = "Yards Per Game",
  y = "Team",
  title = "2024 NFL Team Average Yards Per Game",
  subtitle = "Data derived from 2024 Season through NFLreadr",
  caption = "Data: @nflfastR"
) +
theme_minimal() +
theme(
  axis.text.y = element_blank(),
  axis.ticks.y = element_blank(),
  panel.grid.major.y = element_blank(),
  panel.grid.minor.y = element_blank(),
  plot.title = element_text(face = "bold", hjust = 0),
  plot.title.position = "plot",
  plot.subtitle = element_text(hjust = 0),
  plot.background = element_rect(fill = "white", color = NA),
  legend.position = "none"
) +
scale_fill_nfl() +
scale_color_nfl(type = "secondary")
teamsypg

```

2024 NFL Team Average Yards Per Game

Data derived from 2024 Season through NFLreadr



Data: @nflfastR

```
ggsave("teams_ypg.png",
       path = "/Users/jakeblumengarten/git_repository/nfl_analysis/plots/",
       plot = teamsypg,
       width = 10,
       height = 10,
       dpi = 300,
       bg = "white")
```

Correlation Between Average Yards and Winning Percentage

```
ypg_winpct <- ggplot(yards_pg_pt, aes(x=avgyards, y= winpct))+
  geom_point()+
  geom_smooth(method = "lm", se = F)+
  geom_nfl_logos(aes(team_abbr = postteam), width = 0.065, alpha = 1) +
  labs(
    x= "Yards Per Game",
    y= "Win Percentage",
    caption = "Data: @nflfastR",
    title = "2024 NFL Correlation Between Avg Yards Per Game and Win Percentage"
  )+
  theme_minimal()+
  theme(
    plot.title = element_text(face = "bold"),
    plot.title.position = "plot"
  )+
  scale_x_continuous(breaks = seq(250, 500, 25))+
```

```

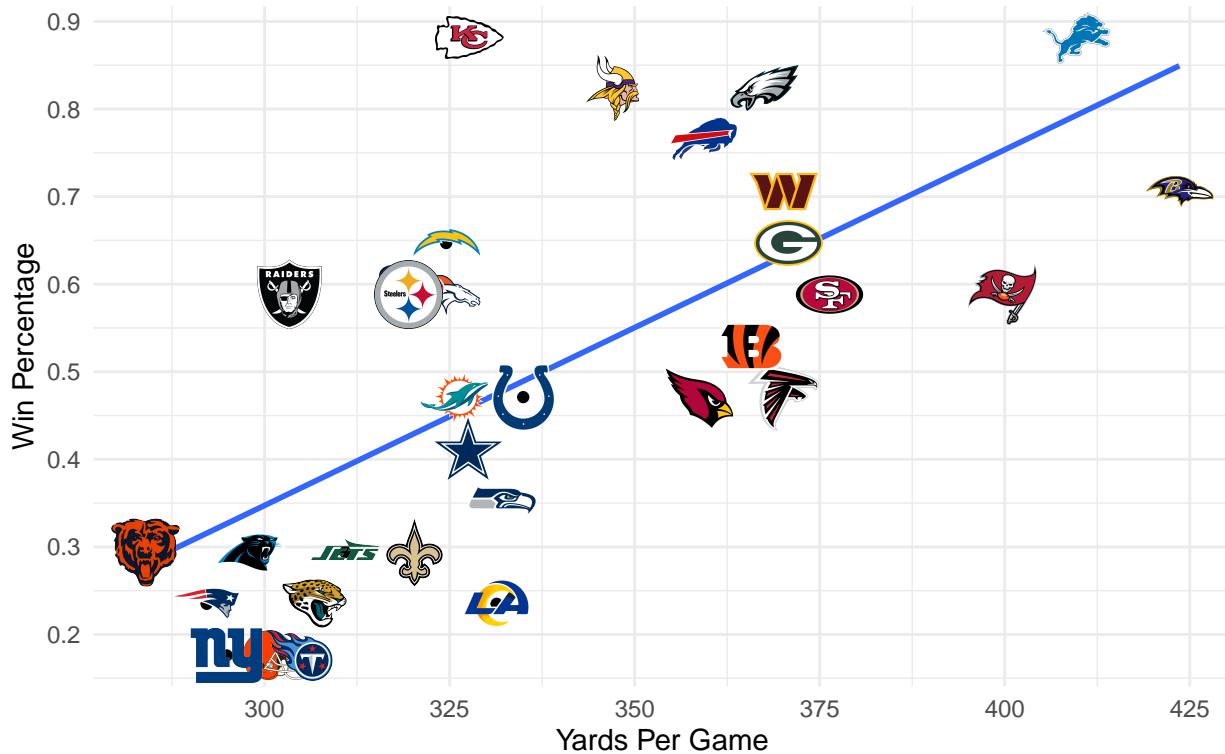
scale_y_continuous(breaks = seq(0,1,0.1))
ggsave("ypg_winpct_corr.png",
       plot = ypg_winpct,
       path = "/Users/jakeblumengarten/git_repository/nfl_analysis/plots/",
       width = 10,
       height = 10,
       dpi = 300,
       bg = "white")

## `geom_smooth()` using formula = 'y ~ x'
ypg_winpct

## `geom_smooth()` using formula = 'y ~ x'

```

2024 NFL Correlation Between Avg Yards Per Game and Win Percentage



Data: @nflfastR

Looking at QB Influence on Win Percentage

```

qbs <- season_24 %>%
  filter(!is.na(passer_player_name),
         !is.na(yards_gained),
         play_type == "pass") %>%
  group_by(passer_player_name, passer_player_id, posteam) %>%
  summarize(avg_yards = mean(yards_gained), plays = n(), .groups = "drop")%>%
  filter(plays > 100, passer_player_name != "")%>%
  arrange(desc(avg_yards))

qbs <- qbs %>%

```

```

mutate(passer_player_name = fct_reorder(passer_player_name, avg_yards, .desc = TRUE))

top_qbs <- qbs %>%
  arrange(desc(avg_yards)) %>%
  slice(1:15)

qbs_bar <- ggplot(top_qbs, aes(x= avg_yards, y= passer_player_id))+ 
  geom_col(aes(color = posteam, fill = posteam), width = 0.5)+ 
  scale_color_nfl(type="secondary")+
  scale_fill_nfl(alpha = 0.4)+ 
  theme_minimal()+
  labs(
    title = "Average Yards by QB",
    x = "Average Yards",
    subtitle = "Top 15",
    caption = "Data: @nflfastR"
  )+
  theme(
    plot.title = element_text(face = "bold"),
    plot.title.position = "plot",
    axis.title.y = element_blank(),
    axis.text.y.left = element_nfl_headshot(size = 1)
  )+
  scale_x_continuous(breaks = seq(1, 10, 1))
ggsave("averageydsqb.png",
       plot = qbs_bar,
       path = "/Users/jakeblumengarten/git_repository/nfl_analysis/plots/",
       width = 10,
       height = 10,
       dpi = 300,
       bg = "white")

## Warning in plot_theme(plot): The `plot.title.position` theme element is not defined in the element
## hierarchy.

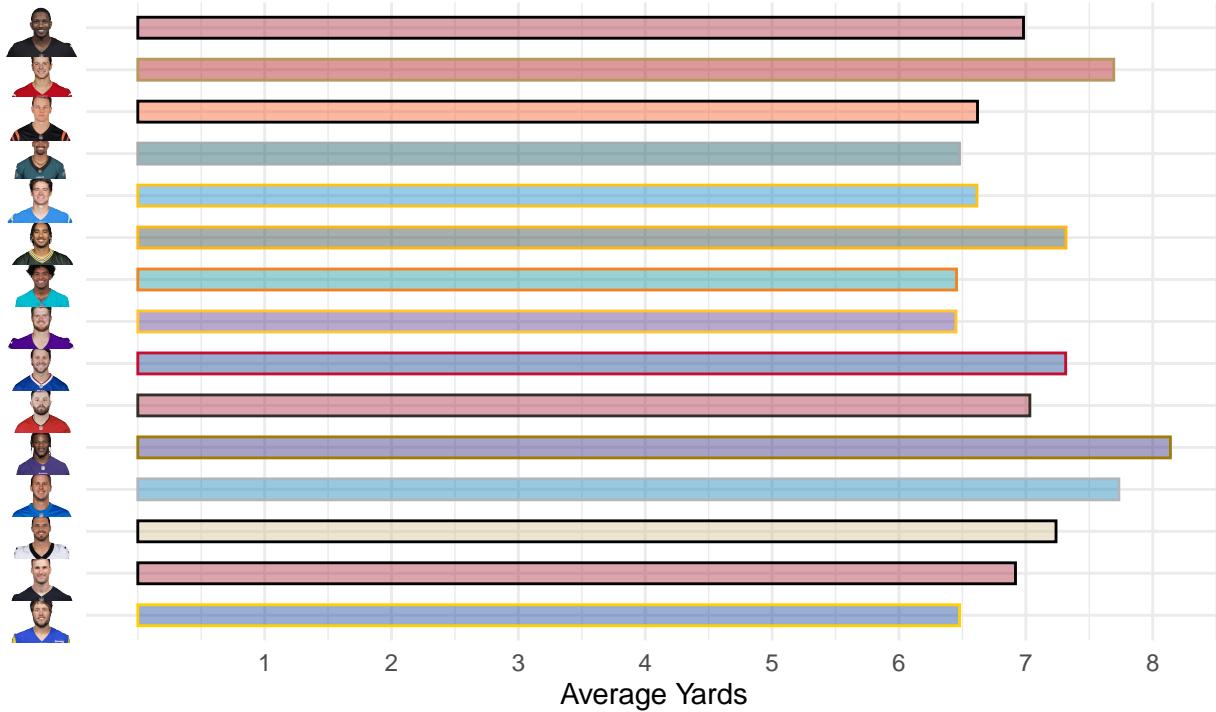
qbs_bar

## Warning in plot_theme(plot): The `plot.title.position` theme element is not defined in the element
## hierarchy.

```

Average Yards by QB

Top 15



Data: @nflfastR