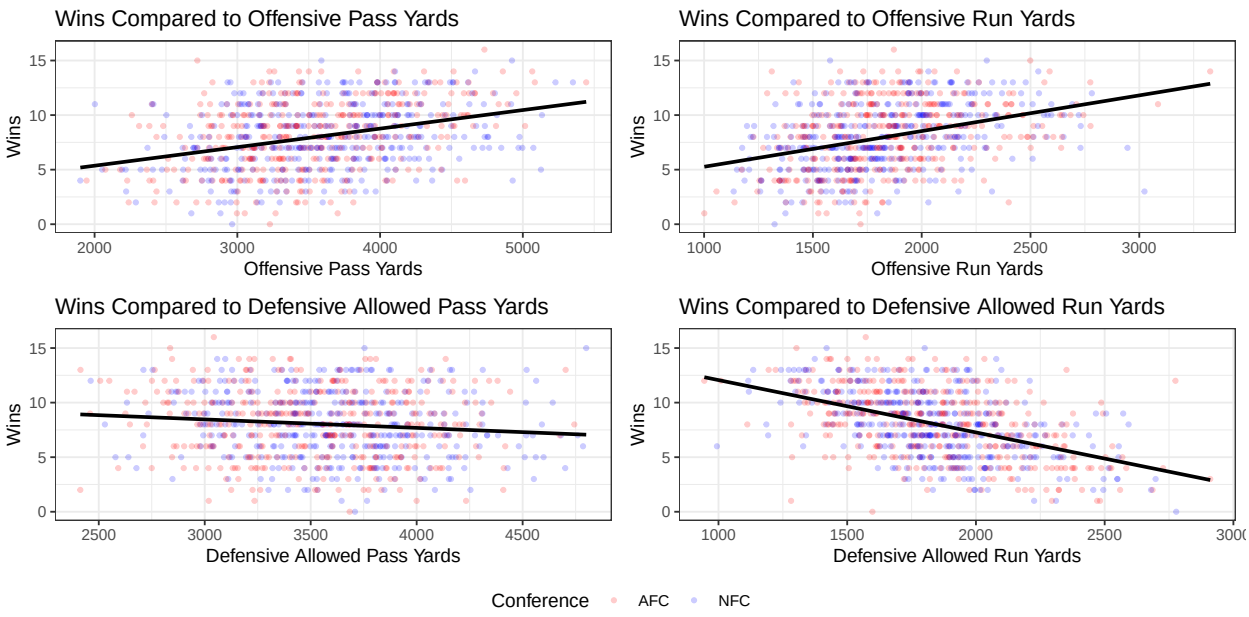
NFL_Data_Analysis

Isaac Swope Kyle Barber Tim Smith

2025-12-03

Data Overview



?@fig-Win-association-1

Table 1

Teams	Average Wins	Mean Offensive Pass Yards	Mean Offense Run Yards	Mean Defensive Pass Yards	Mean Defensive Run Yards
ARI	6.916667	3598.417	1559.417	3589.958	1932.250
ATL	7.791667	3623.417	1858.667	3792.958	1916.625
BAL	9.458333	3247.500	2084.083	3407.500	1501.792
BUF	7.708333	3255.375	1941.000	3221.000	1968.667
CAR	7.583333	3308.417	1915.333	3519.833	1830.167
CHI	7.500000	3182.750	1775.917	3548.667	1835.250
CIN	7.291667	3482.042	1766.792	3616.208	1887.458
CLE	5.291667	3175.083	1711.042	3457.750	2170.792
DAL	8.583333	3679.125	1950.875	3536.792	1782.583

Table 1

Teams	Average Wins	Mean Offensive Pass Yards	Mean Offense Run Yards	Mean Defensive Pass Yards	Mean Defensive Run Yards
DEN	8.541667	3689.208	2003.583	3432.250	1781.417
DET	5.791667	3728.750	1527.250	3817.000	1940.750
GB	9.958333	3960.792	1817.250	3531.542	1847.375
HOU	6.761905	3501.048	1803.762	3648.286	1947.190
IND	9.750000	3988.958	1735.125	3528.333	1957.333
JAX	6.416667	3342.375	1890.958	3528.833	1854.958
KC	8.958333	3721.917	1979.125	3667.083	1998.042
LA	7.666667	3733.750	1728.917	3533.833	1890.708
LAC	8.166667	3892.083	1742.625	3569.333	1764.667
LV	6.541667	3541.250	1797.792	3629.833	1995.250
MIA	7.750000	3363.583	1718.250	3510.583	1857.417
MIN	8.458333	3578.417	2011.667	3768.750	1733.958
NE	11.250000	3931.292	1864.833	3585.875	1756.208
NO	8.916667	4153.792	1813.708	3656.417	1868.917
NYG	7.666667	3622.708	1832.167	3676.292	1820.167
NYJ	7.041667	3168.583	1853.833	3468.083	1837.750
PHI	9.250000	3701.000	1990.417	3531.208	1801.333
PIT	10.041667	3672.917	1864.958	3295.708	1582.000
SEA	9.125000	3446.958	1981.708	3632.792	1834.292
SF	7.541667	3326.375	1969.667	3618.833	1747.417
TB	7.541667	3679.042	1662.875	3486.292	1787.375
TEN	8.458333	3333.250	1964.042	3754.167	1703.875
WAS	6.833333	3450.458	1857.292	3537.042	1858.875

Team Wins from 1999-2022

Figure 1: Boxplots of Team Wins

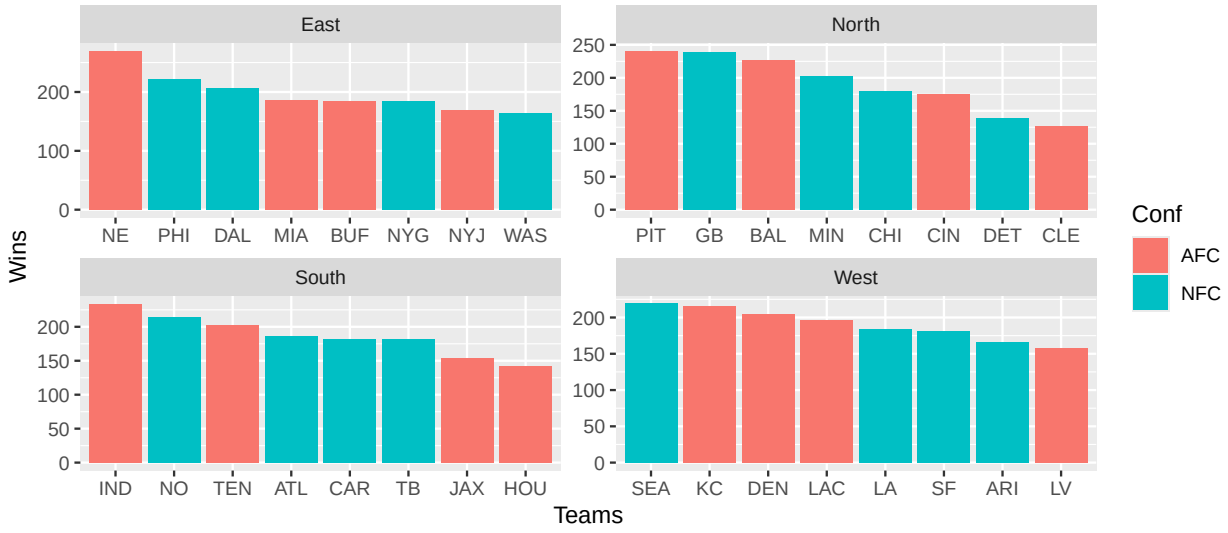


Table 2: Team Wins Statistics

Team	Seasons	Minimum	Quartile 1	Median	Quartile 3	Max	Mean	SD
ARI	24	3	5.00	6.5	8.25	13	6.92	2.72
ATL	24	4	5.75	7.0	10.00	13	7.79	2.78
BAL	24	5	8.00	10.0	11.00	14	9.46	2.34
BUF	24	3	6.00	7.0	9.00	13	7.71	2.56
CAR	24	1	6.00	7.0	8.75	15	7.58	3.17
CHI	24	3	5.00	7.0	9.25	13	7.50	2.95
CIN	24	2	4.00	7.5	10.00	12	7.29	3.10
CLE	24	0	4.00	5.0	7.00	11	5.29	2.68
DAL	24	4	6.00	8.5	10.25	13	8.58	2.69
DEN	24	4	6.75	8.0	10.25	13	8.54	2.75
DET	24	0	3.00	6.0	8.25	11	5.79	2.90
GB	24	4	8.00	10.0	12.00	15	9.96	2.71
HOU	21	2	4.00	7.0	9.00	12	6.76	3.03
IND	24	2	8.00	10.5	12.00	14	9.75	3.22
JAX	24	1	4.75	6.0	8.25	14	6.42	3.24
KC	24	2	7.00	9.5	12.00	14	8.96	3.51
LA	24	1	5.75	7.0	10.25	14	7.67	3.71
LAC	24	1	6.50	8.5	9.25	14	8.17	3.10
LV	24	2	4.00	6.0	8.00	12	6.54	2.89
MIA	24	1	6.00	8.0	9.25	11	7.75	2.44
MIN	24	3	6.75	8.0	10.00	13	8.46	2.59
NE	24	5	10.00	12.0	13.00	16	11.25	2.56
NO	24	3	7.00	8.5	11.00	13	8.92	2.87
NYG	24	3	6.00	7.5	10.00	12	7.67	2.66
NYJ	24	2	4.75	7.5	9.00	11	7.04	2.56
PHI	24	4	7.75	9.5	11.00	14	9.25	2.75

Table 2: Team Wins Statistics

Team	Seasons	Minimum	Quartile 1	Median	Quartile 3	Max	Mean	SD
PIT	24	6	8.75	10.0	12.00	15	10.04	2.24
SEA	24	4	7.00	9.0	10.25	13	9.12	2.36
SF	24	2	5.00	7.0	10.25	13	7.54	3.46
TB	24	2	5.00	7.5	10.00	13	7.54	3.08
TEN	24	2	6.75	9.0	11.00	13	8.46	3.13
WAS	24	3	5.00	7.0	8.00	10	6.83	2.12

#Code Appendix

```
#Tidyverse Coding Styling


library(tidyverse)
library(dbplyr)
library(kableExtra)
#install.packages("patchwork")
library(patchwork)
#install.packages("ggpubr")
library(ggpubr)
NFL_raw <- read.csv("nfl-team-statistics (1).csv")

NFC_East_data <- NFL_raw %>%
  filter(team %in% c("DAL", "NYG", "PHI", "WAS"))%>%
  mutate(Conf = 'NFC_East')

NFC_West_data <- NFL_raw %>%
  filter(team %in% c("LA", "SEA", "SF", "ARI"))%>%
  mutate(Conf = 'NFC_West')

NFC_North_data <- NFL_raw %>%
  filter(team %in% c("CHI", "GB", "DET", "MIN"))%>%
  mutate(Conf = 'NFC_North')

NFC_South_data <- NFL_raw %>%
  filter(team %in% c("TB", "CAR", "ATL", "NO"))%>%
  mutate(Conf = 'NFC_South')

AFC_East_data <- NFL_raw %>%
  filter(team %in% c("NE", "BUF", "MIA", "NYJ"))%>%
  mutate(Conf = 'AFC_East')

AFC_West_data <- NFL_raw %>%
  filter(team %in% c("DEN", "LAC", "KC", "LV"))%>%
  mutate(Conf = 'AFC_West')

AFC_North_data <- NFL_raw %>%
  filter(team %in% c("BAL", "PIT", "CIN", "CLE"))%>%
  mutate(Conf = 'AFC_North')

AFC_South_data <- NFL_raw %>%
  filter(team %in% c("JAX", "IND", "HOU", "TEN"))%>%
  mutate(Conf = 'AFC_South')

NFL_Clean <- bind_rows(NFC_East_data, NFC_North_data, NFC_South_data, NFC_West_data, AFC_East_
  arrange(season, team)%>%
```

```
    separate_wider_delim(
    cols = 'Conf',
    delim = '_',
    names = c('Conf', 'Div')
  )

#|label: fig-Win-association
#|fig-cap: "Win Association of Offensive and Defensive Pass Yards"
#| fig-subcap:
#|   - "first"
#|   - "second"
#|   - "third"
#|   - "fourth"
#|fig-alt: "Four scatter plots showing NFL data, with axes for wins and various yard metrics. T

#author: Timothy Smith
plot1<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(
    mapping = aes(
      x = offense_total_yards_gained_pass,
      y = wins,
      color = Conf
    )
  ) +
  geom_point(size = 1, alpha=0.2) +
  labs(#Step 4: add labels and title to the data visualization and create the colors for the po
    x = "Offensive Pass Yards",
    y = "Wins",
    color = "Conference",
    title = "Wins Compared to Offensive Pass Yards"
  ) +
  scale_color_manual(
    values = c("red", "blue")
  )+
  theme_bw() +
  theme(
    legend.position = "bottom"
  )+
  geom_smooth(method = "lm", se = FALSE, color = "black")

  plot2<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(
    mapping = aes(
      x = offense_total_yards_gained_run,
      y = wins,
```

```
      color = Conf
    )
  ) +
  geom_point(size=1, alpha=0.2) +
  labs(#Step 5: add labels and title to the data visualization and create the colors for the po
    x = "Offensive Run Yards",
    y = "Wins",
    color = "Conference",
    shape = "team",
    title = "Wins Compared to Offensive Run Yards"
  ) +
  scale_color_manual(
    values = c("red", "blue")
  )+
  theme_bw() +
  theme(
    legend.position = "bottom"
  )+
  geom_smooth(method = "lm", se = FALSE, color = "black")

  plot3<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(
    mapping = aes(
      x = defense_total_yards_gained_pass,
      y = wins,
      color = Conf
    )
  ) +
  geom_point(size=1, alpha=0.2) +
  labs(#Step 6: add labels and title to the data visualization and create the colors for the li
    x = "Defensive Allowed Pass Yards",
    y = "Wins",
    color = "Conference",
    title = "Wins Compared to Defensive Allowed Pass Yards"
  ) +
  scale_color_manual(
    values = c("red", "blue")
  )+
  theme_bw() +
  theme(
    legend.position = "bottom"
  )+
  geom_smooth(method = "lm", se = FALSE, color = "black")


  plot4<-NFL_Clean%>%
```

```
  group_by(Div)%>%
ggplot(
    mapping = aes(
      x = defense_total_yards_gained_run,
      y = wins,
      color = Conf
    )
  ) +
  geom_point(size=1, alpha=0.2) +
  labs(#Step 7: add labels and title to the data visualization and create the colors for the li
    x = "Defensive Allowed Run Yards",
    y = "Wins",
    color = "Conference",
    title = "Wins Compared to Defensive Allowed Run Yards"
  ) +
  scale_color_manual(
    values = c("red", "blue")
  )+
  theme_bw() +
  theme(
    legend.position = "bottom"
  )+
  geom_smooth(method = "lm", se = FALSE, color = "black")

  ggarrange(plot1, plot2, plot3, plot4, nrow = 2, ncol = 2, common.legend = TRUE, legend = "bot




#|tbl-cap: "NFL Offensive and Drfensive Statistics"

NFL_O_D <- NFL_Clean %>%
  group_by(team)%>%
  summarize(
    Mean_Wins = mean(wins, na.rm = TRUE),
    Mean_O_pass_yards = mean(offense_total_yards_gained_pass, na.rm = TRUE),
    Mean_O_run_yards = mean(offense_total_yards_gained_run, na.rm = TRUE),
    Mean_D_pass_yards = mean(defense_total_yards_gained_pass, na.rm = TRUE),
    Mean_D_run_yards = mean(defense_total_yards_gained_run, na.rm = TRUE)
  )

    names(NFL_O_D) <- c(
    "Teams",
    "Average Wins",
    "Mean Offensive Pass Yards",
    "Mean Offense Run Yards",
    "Mean Defensive Pass Yards",
```

```
    "Mean Defensive Run Yards"
    )

NFL_O_D %>%
  kable(
    booktabs = TRUE,
    align = c("l", rep("c",10)))

#author: Isaac Swope


ggplot(NFL_Clean) +
  aes(x = fct_reorder(team, wins, .fun=sum, .desc=TRUE), y = wins, fill = Conf) +
  geom_col(color = NA) +
  facet_wrap("Div", scales = "free") + #eliminate spacing between columns due to scale
  labs(y = "Wins",
       x = "Teams",
       )

#author: Isaac Swope

winsTable <- NFL_Clean %>%
  group_by(team)%>%
  summarize(
    Count=n(),
    Min = min(wins, na.rm = TRUE),
    Quartile_1 = quantile(wins, probs = 0.25, na.rm = TRUE),
    Median = median(wins, na.rm = TRUE),
    Quintile_3 = quantile(wins, probs = 0.75, na.rm = TRUE),
    Max = max(wins, na.rm = TRUE),
    Mean = mean(wins, na.rm = TRUE),
    SD = sd(wins, na.rm = TRUE),
  )%>%
  mutate(
    across(
      .cols = where(is.numeric),
      .fns = ~round(.x, digits = 2)
    )
  )

names(winsTable) <- c(
    "Team",
    "Seasons",
    "Minimum",
    "Quartile 1",
    "Median",
    "Quartile 3",
```

```
    "Max",
    "Mean",
    "SD"
    )

winsTable %>%
  kable(
    booktabs = TRUE,
    align = c("l", rep("c",10))
  )
```