NFL Data Analysis

Isaac Swope Kyle Barber Tim Smith

2025-12-03

Data Overview

We chose to use the National Football Team Statistics data from data.scorenetwork.org. This data set was created by Ron Yurko and was published on May 24th, 2023. This data sets contains various team statistics from the 1999 regular season to the 2022 regular season for all 32 NFL teams. In this data set a case is an individual NFL team during a specific regular season from 1999 to 2022. This data was collected through the nflreadr package in r. The data contains two categorical variables, team and season. It also contains 54 quantitative variables including offensive passing stats, offensive rushing stats, defensive stats, wins, losses, and much more. This data was produced in a free to use csv file found on the data.scorenetwork.org website, (Yurko, 2023). The data met all of the FAIR requirements as it was easily finadable and accessible. Also, the data was interoperable as all the data was in a csv file. Lastly, it was reproducible as we were able to work with it in R studio. The data also met all CARE principles. The data can be used by anyone to benefit data analysis. The data is accessible by anyone and free to use. Also, the data is responsible and ethical as the data is accurate and up to date.

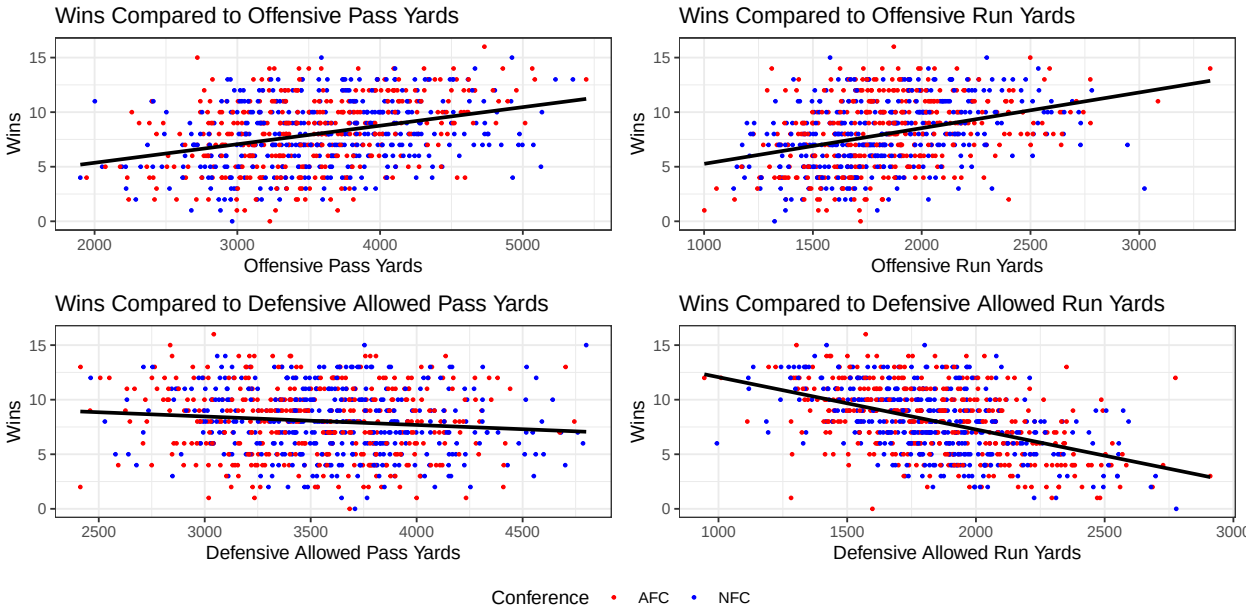
Offensive and Defensive Association with Wins

Scatterplot of Win associations

In Figure 1 the y-axis has the wins that range from 0 to 16 and a linear regression line that wins with four different variables. In Wins Compared to Offensive Pass Yards the x-axis is the Offensive Pass Yards that range from 1000 to 6000. There is a positive linear association between wins and offensive pass yards which shows that as teams pass for more yards in a season they tend to win more games. In Wins Compared to Offensive Run Yards the x-axis is offensive run yards that range from 1000 to 4000 yards. There is a slightly stronger positive linear association between offensive run yards per seasons and offensive pass yards. This leads us to believe that it is slightly more important for teams to focus on good run offense then pass offense. Wins Compared to Defensive Allowed Pass Yards has defensive allowed pass yards as the x-axis with a range of 2000 to 5000 yards. There is a slightly negative linear association between defensive allowed pass yards and team wins. In Wins Compared to Defensive Allowed Run Yards the x-axis is defensive allowed run yards with a range from 500 to 3000 yards. There is a relatively strong negative linear association between defensive run yards allowed and team wins in a season. This leads us to believe that it is much more important to focus on run defense than pass defense as that would lead to more wins. In Figure 1 the colors are sorted by AFC and NFC teams for all of the graphs. In Figure 1 it seems

to be more important to focus on both run offense and run defense to lead to more wins. It is definitely more important to focus on run defense then it is to focus on pass defense. This could be because a strong game allows a team to control the clock and have more possession time then the other team as every run play uses the clock.

Figure 1: Win Association of Offensive and Defensive Pass Yards



Average Wins and Offensive and Defensive Yards Table

Table 1 includes the average team wins, the 32 NFL teams, average Offensive pass yards, average offensive rush yards, average defensive rush yards, and average defensive pass yards. The defensive statistics is the average amount of yards that the defense allowed so the lower the number the better the team is in that particular stat. The team with the most average wins in the patriots which makes sense as they had one of the biggest dynasties in NFL history. What is interesting about this is that they had the highest wins, but in every other category there was at least one team that performed better then them. The highest average offensive pass yards belonged to New Orleans which is 9th in the average wins statistic. The highest average rush yards belongs and lowest allowed rush yards belongs to Baltimore which is 5th in average wins. Buffalo has the lowest allowed pass yards and they are 17th in average wins. This shows that the best team is not the best in each category so it must be a combination of these things that lead to winning football. Cleveland does have the lowest average wins and lowest average allowed rush yards so there could be a stronger correlation with losing and have a bad rush defense which is also supported in Figure 1.

Table 1: NFL Offensive and Defensive Statistics

Teams	Average Wins	Mean Offensive Pass Yards	Mean Offense Run Yards	Mean Defensive Pass Yards	Mean Defensive Run Yards
NE	11.25	3,931.29	1,864.83	3,585.88	1,756.21
PIT	10.04	3,672.92	1,864.96	3,295.71	1,582.00

Table 1: NFL Offensive and Defensive Statistics

Teams	Average Wins	Mean Offensive Pass Yards	Mean Offense Run Yards	Mean Defensive Pass Yards	Mean Defensive Run Yards
GB	9.96	3,960.79	1,817.25	3,531.54	1,847.38
IND	9.75	3,988.96	1,735.12	3,528.33	1,957.33
BAL	9.46	3,247.50	2,084.08	3,407.50	1,501.79
PHI	9.25	3,701.00	1,990.42	3,531.21	1,801.33
SEA	9.12	3,446.96	1,981.71	3,632.79	1,834.29
KC	8.96	3,721.92	1,979.12	3,667.08	1,998.04
NO	8.92	4,153.79	1,813.71	3,656.42	1,868.92
DAL	8.58	3,679.12	1,950.88	3,536.79	1,782.58
DEN	8.54	3,689.21	2,003.58	3,432.25	1,781.42
MIN	8.46	3,578.42	2,011.67	3,768.75	1,733.96
TEN	8.46	3,333.25	1,964.04	3,754.17	1,703.88
LAC	8.17	3,892.08	1,742.62	3,569.33	1,764.67
ATL	7.79	3,623.42	1,858.67	3,792.96	1,916.62
MIA	7.75	3,363.58	1,718.25	3,510.58	1,857.42
BUF	7.71	3,255.38	1,941.00	3,221.00	1,968.67
LA	7.67	3,733.75	1,728.92	3,533.83	1,890.71
NYG	7.67	3,622.71	1,832.17	3,676.29	1,820.17
CAR	7.58	3,308.42	1,915.33	3,519.83	1,830.17
SF	7.54	3,326.38	1,969.67	3,618.83	1,747.42
TB	7.54	3,679.04	1,662.88	3,486.29	1,787.38
CHI	7.50	3,182.75	1,775.92	3,548.67	1,835.25
CIN	7.29	3,482.04	1,766.79	3,616.21	1,887.46
NYJ	7.04	3,168.58	1,853.83	3,468.08	1,837.75
ARI	6.92	3,598.42	1,559.42	3,589.96	1,932.25
WAS	6.83	3,450.46	1,857.29	3,537.04	1,858.88
HOU	6.76	3,501.05	1,803.76	3,648.29	1,947.19
LV	6.54	3,541.25	1,797.79	3,629.83	1,995.25
JAX	6.42	3,342.38	1,890.96	3,528.83	1,854.96
DET	5.79	3,728.75	1,527.25	3,817.00	1,940.75
CLE	5.29	3,175.08	1,711.04	3,457.75	2,170.79

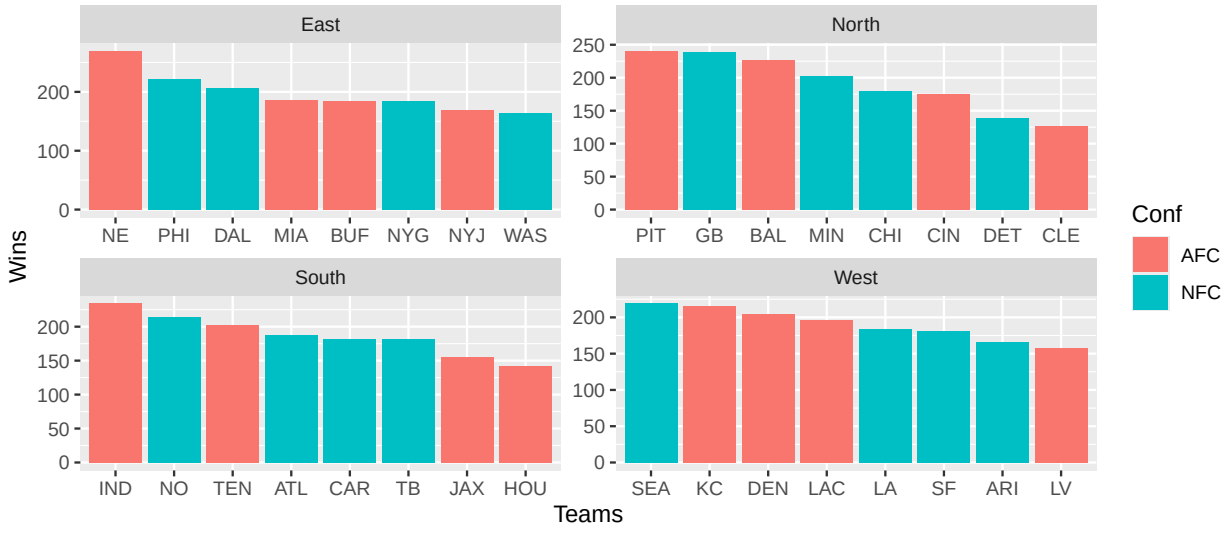
Team Wins from 1999-2022

When people think of the best football team of all time, to many, the first thing to come to mind is wins. In this section we will look at a visualization and some statistics to build an understanding of team wins in the 21st century.

First we have simple bar charts Figure 2 showing every NFL team's total wins since from 1999-2022. The bar charts are separated and colored so one can more easily compare within divisions and across the conferences. Note the thin white lines, which show how the wins are summed up by season. An interesting observation is that it seams as though teams with more wins were also more consistent across seasons as the distance between white lines does not vary as much as those in

teams with less overall wins. This is a observed trend which can be made by looking at the figure, but it will not be shown to be statistically significant in this section, whether or not it truly is.

Figure 2: Barcharts of Team Wins



Now it would be appropriate to provide a table showing statistics for wins by season. The following table Table 2 gives seasons played from 1999-2022, minimum wins in a season, first quartile, median wins, third quartile, maximum wins in a season, simple arithmetic mean of wins, and standard deviation for each team. One thing to note is that Houston has played three seasons less than the other teams, as they joined the NFL in the 2002 season. Standard deviation represented by SD is the how much a team's amount of wins usually varies from season to season.

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

Table 2: Team Wins Statistics

Team	Seasons	Minimum	Quartile 1	Median	Quartile 3	Max	Mean	SD
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
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

Team Efficiency

Figure 3: NFL Team Efficiency Scatter Plot

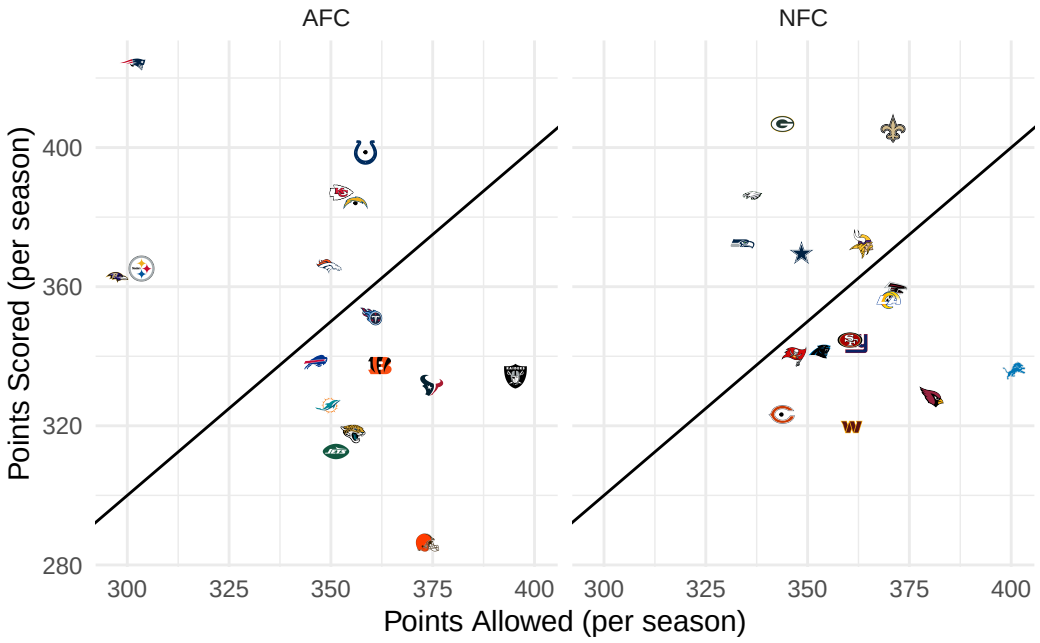


Figure 3 shows a scatter plot of each NFL team's efficiency by showing the relationship between a team's average points allowed per season (x-axis) vs average points scored per season (y-axis) from

the 1999 NFL regular season to the 2022 NFL regular season. Each point on the graph represents an individual team average and differentiates the teams by team logo, ("NFL Logos | Football Team Logos," n.d.). The top left corner of the plot shows the most efficient teams. The bottom right corner of the plot shows the least efficient teams. The diagonal line in the middle of both plots demonstrates a efficiency threshold. If a team is above the line they are considered efficient. If a team is below the line they are considered inefficient. If a team is on the line they are neutral. The plot is broken up into two sides, one representing the AFC teams and the other representing the NFC teams.

Table 3: AFC Efficiency Table

Team	image	Average Point Differential	Win Rate	Loss Rate
NE		122.5833	11.2500	4.8333
PIT		61.7500	10.0417	5.9167
IND		40.1667	9.7500	6.2917
BAL		65.3750	9.4583	6.6250
KC		34.5417	8.9583	7.1250
DEN		16.2083	8.5417	7.5417
TEN		-8.2917	8.4583	7.6250
LAC		28.0000	8.1667	7.9167
MIA		-23.2917	7.7500	8.3333
BUF		-8.0000	7.7083	8.3333
CIN		-24.5833	7.2917	8.5833
NYJ		-38.5417	7.0417	9.0417
HOU		-43.1905	6.7619	9.2857
LV		-61.0417	6.5417	9.5417
JAX		-37.8333	6.4167	9.6667
CLE		-87.4583	5.2917	10.7500

Table 4: NFC Efficiency Table

Team	Logo	Avg. Point Dif	Win Rate	Loss Rate
GB		62.92	9.96	6.04
PHI		49.71	9.25	6.75
SEA		38.21	9.12	6.92
NO		34.71	8.92	7.17
DAL		21.46	8.58	7.50
MIN		9.21	8.46	7.54
ATL		-13.54	7.79	8.25
LA		-13.83	7.67	8.38
NYG		-18.12	7.67	8.38
CAR		-11.92	7.58	8.46
SF		-15.71	7.54	8.50
TB		-6.54	7.54	8.54
CHI		-20.33	7.50	8.58
ARI		-52.21	6.92	9.08
WAS		-41.25	6.83	9.17
DET		-64.54	5.79	10.21

Refrences

NFL Logos | Football Team Logos. (n.d.). In *Freebie Supply*. Retrieved December 17, 2025, from https://freebiesupply.com/s/nfl-logos/

Yurko, R. (2023). *National Football League Team Statistics – SCORE Sports Data Repository*. https://data.scorenetwork.org/football/nfl-team-statistics.html

Code Appendix

```
#Tidyverse Coding Styling


library(tidyverse)
library(ggimage)
library(kableExtra)
library(patchwork)
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_d
```

```
  arrange(season, team)%>%
    separate_wider_delim(
    cols = 'Conf',
    delim = '_',
    names = c('Conf', 'Div')
  )

#author: Timothy Smith
#reviewed: Isaac Swope, Kyle Barber

plot1<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(# Step 1: Create the first graph using ggplot and geom point with x as offense pass yard
    mapping = aes(
      x = offense_total_yards_gained_pass,
      y = wins,
      color = Conf
    )#Step 2: Make the points size 1 and the transparency to be .2 ----
  ) +
  geom_point(size = .6) +
  labs(#Step 3: 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")
# Step 4: Create the regression line ----
  plot2<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(# Step 5: Repeat Step 1 but x is the total rush yards----
    mapping = aes(
      x = offense_total_yards_gained_run,
      y = wins,
      color = Conf
    )#Step 6: repeat step 2 for the second plot ----
  ) +
  geom_point(size=.6) +
  labs(#Step 7: repeat step 3 for the second plot----
    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")
  #Step 8: Repeat step 4 for second plot----
  plot3<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(#Step 9: Repeat step 1 but with x as the defensive allowed pass yards ----
    mapping = aes(
      x = defense_total_yards_gained_pass,
      y = wins,
      color = Conf
    )#Step 10: Repeat step 2 but for third plot----
  ) +
  geom_point(size=.6) +
  labs(#Step 11: repeat step 3 but for third plot----
    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")
  #Step 12: repeat step 4 but for third plot----

  plot4<-NFL_Clean%>%
  group_by(Div)%>%
ggplot(#Step 13: repeat step 1 but x is the allowed rush yards----
    mapping = aes(
      x = defense_total_yards_gained_run,
      y = wins,
      color = Conf
```

```
    )# Step 14: repeat step 2 but for fourth plot----
  ) +
  geom_point(size=.6) +
  labs(#Step 15: repeat step 3 but for fourth plot----
    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")
#Step 16: repeat step 4 but for fourth plot----

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

#Step 17: create a grid with all of the plots together----


#author: Timothy Smith
#reviewed: Isaac Swope, Kyle Barber

NFL_O_D <- NFL_Clean %>%
  group_by(team)%>%
  summarize(#Step 1: Determine the average values of the 5 main statistics that I want to work
    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)
  ) %>%
  arrange(desc(Mean_Wins))

    names(NFL_O_D) <- c(#Step 2: replace the Names of the columns----
    "Teams",
    "Average Wins",
    "Mean Offensive Pass Yards",
    "Mean Offense Run Yards",
    "Mean Defensive Pass Yards",
    "Mean Defensive Run Yards"
    )
```

```r
NFL_O_D %>%
  kable(#Step 3: Create the table using the kable function----
    booktabs = TRUE,
    align = c("l", rep("c",10)),
    digits = 2,
    format.args = list(big.mark = ",")
  )

#author: Isaac Swope
#reviewed: Kyle Barber

#we need to sum and order by wins so we use fct_reorder and then .desc=TRUE to force descending
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
#reviewed: Timothy Smith, Kyle Barber

winsTable <- NFL_Clean %>%
  group_by(team)%>% #we want to group by team
  summarize( # calculating the selected statistics
    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) #rounding to 2 decimal places
    )
  )

names(winsTable) <- c( #renaming columns
    "Team",
    "Seasons",
    "Minimum",
    "Quartile 1",
```

```
    "Median",
    "Quartile 3",
    "Max",
    "Mean",
    "SD"
    )

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

#Author: Kyle Barber
#reviewed:

NFL_Clean_3 <- NFL_Clean %>% ## Wrangle data
  select(team, Conf, score_differential, points_scored,
         points_allowed, wins, losses) %>% ### Choose necessary columns
  group_by(team, Conf) %>%
  summarise(
    points_scored = mean(points_scored),
    points_allowed = mean(points_allowed),
    wins = mean(wins),
    losses = mean(losses),
    point_differential = mean(score_differential),
    .groups = "drop"
  ) ### Get the mean for each variable
NFL_Clean_3$logo <- paste0("Logos/", NFL_Clean_3$team, ".png") ### Add logo column

ggplot(NFL_Clean_3) + ## Make plot
  aes(x = points_allowed, y = points_scored) + ### Set aesthetic
  geom_point(size = 0.1) + ### Make scatter plot
  geom_image(aes(image = logo)) + ### add in images
  labs(
    x = "Points Allowed (per season)",
    y = "Points Scored (per season)"
  ) + ### Set x and y axis labels
  geom_abline(slope = 1, intercept = 0) + ### Add efficiency line
  theme_minimal() +
  facet_wrap(vars(Conf)) ### Separate plots by conference

#Author: Kyle Barber
#Reviewed:

AFC_table <- NFL_Clean_3%>% ## Clean data
  filter(Conf == 'AFC')%>% ### Choose necessary confrence
```

```
  select(team, point_differential, wins, losses, logo)%>% ### Choose necessary variables
  mutate(
    point_differential = round(point_differential, digits = 4),
    wins = round(wins, digits = 4),
    losses = round(losses, digits = 4),
    image = ""
  )%>% ### Round all numeric variables
  arrange(desc(wins)) %>% ### Arrange by descending win rate
  rename( ### Rename variables
    Team = team,
    `Average Point Differential` = point_differential,
    `Win Rate` = wins,
    `Loss Rate` = losses
  )

AFC_table %>% ## Make table
  select(Team, image, `Average Point Differential`, `Win Rate`, `Loss Rate`) %>%### Choose nec
  kable( ### Set format
    booktabs = TRUE,
    align = c("l", rep("c", 10)),
    format = "latex"
  )%>%
  kable_paper()%>%
  column_spec(2, image = spec_image(path = AFC_table$logo, 50, 50)) ### Add images

#Author: Kyle Barber
#Reviewed:

NFC_table <- NFL_Clean_3%>% ## Clean data
  filter(Conf == 'NFC')%>% ### Choose necessary conference
  select(team, point_differential, wins, losses, logo)%>% ### Choose necessary variables
  mutate( ### round each numeric variable
    point_differential = round(point_differential, digits = 2),
    wins = round(wins, digits = 2),
    losses = round(losses, digits = 2),
    image = ""
  )%>%
  arrange(desc(wins)) %>% ### Arrange by win rate descending
  rename( ### Rename variables
    Team = team,
    `Average Point Differential` = point_differential,
    `Win Rate` = wins,
    `Loss Rate` = losses
  )

NFC_table %>% ## Make table
  select(Team, image, `Average Point Differential`, `Win Rate`, `Loss Rate`) %>% ### select ne
```

```
kable( ### Set format
  booktabs = TRUE,
  align = c("l", rep("c", 10)),
  format = "latex",
  col.names = c( ### Set column names
    "Team",
    "Logo",
    "Avg. Point Dif",
    "Win Rate",
    "Loss Rate"
  )
)%>%
kable_paper()%>%
column_spec(2, image = spec_image(path = NFC_table$logo, 50, 50)) ### Add images
```