Group Project Progress Report

Isaac Swope, Kyle Barber, Tim Smith

Data

- **National Football League Team Statistics**
  - Author: Ron Yurko
  - Published: May 24, 2023
  - Data shows all 32 NFL team's regular season from 1999 – 2022
  - All data collected from the nflreadr package in R
  - Case: A team in a specific regular season
  - 56 total variables
    - Team
    - Season
    - Regular season team performance (offense and defense)

Insight

- First, we looked at the raw data and saw what wrangling steps we needed to take.
  - Added division and conference columns
  - Air yards & NAs
- After looking over the data we wanted to compare the performance between teams using various tables and visualizations
- We evaluated the attributes and statistics we wanted to compare
  - Wins, defensive pass yards, ETC

Plan

- Tables
  1. Create a table with the teams as the rows and Avg. Wins, Avg. Defensive Pass Yards, and Avg. Defensive Run Yards in the columns
  2. Create a table with teams as rows and descriptive statistics on wins as columns
  3. Create a table with teams as the rows and Avg. Wins, Avg. Losses, and Avg. Point Differential as the columns ordered by highest to lowest average wins.
- Visualizations
  1. Time Series Plot Separated by Division with Year as the x variable and wins as the y variable.
  2. Time Series Plot Separated by Division with Year as the x variable and offensive pass yards as the y variable.
  3. Bar charts of total team wins separated by division and conference
  4. Scatter plot to show team efficiency from 1999 - 2022 by showing points scored on the x-axis and points against on the y-axis. Each team will have an image representing their point on the graph.

Time Series Plot

NFC teams and their wins from year to year

```
NFL_Clean%>%
  filter(Conf == "NFC")%>%
  filter(Div %in% c("East", "West"))%>%
ggplot(
    mapping = aes(
      x = season,
      y = wins,
      color = team
    )
) +
geom_line(linewidth = 1) +
facet_wrap(~Div) +
labs(#Step 4: add labels and title to the data visualization and create the colors for the lines----
    x = "Year",
    y = "Wins",
    color = "team",
    shape = "team",
    title = "NFC teams and their wins from year to year"
) +
scale_color_manual(
    values = c("#332288", "#117733", "#44AA99","#88CCEE", "#DDCC77", "#AA4499","#800080", "#FFA500" )
)+
theme_bw() +
theme(
  legend.position = "bottom"
)
```
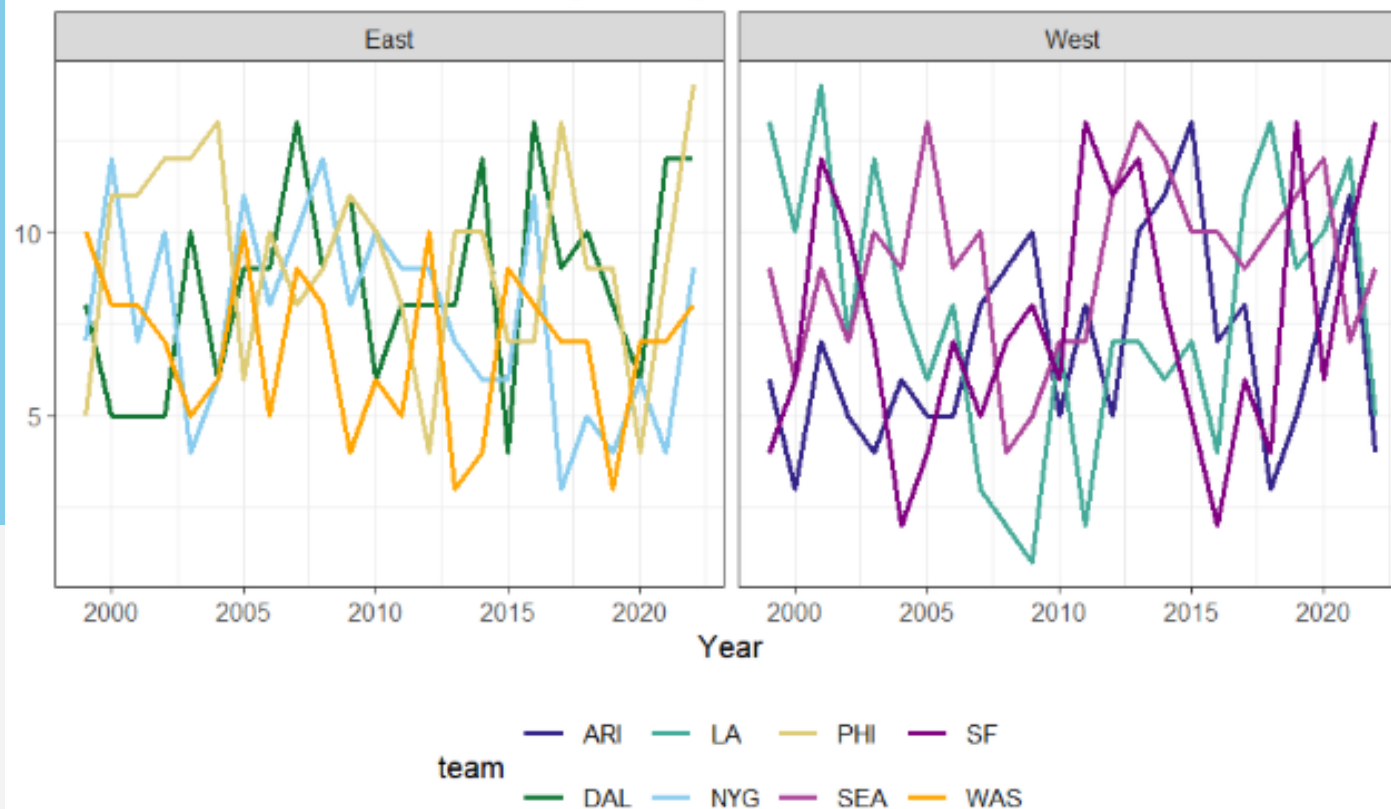
Efficiency Scatterplot
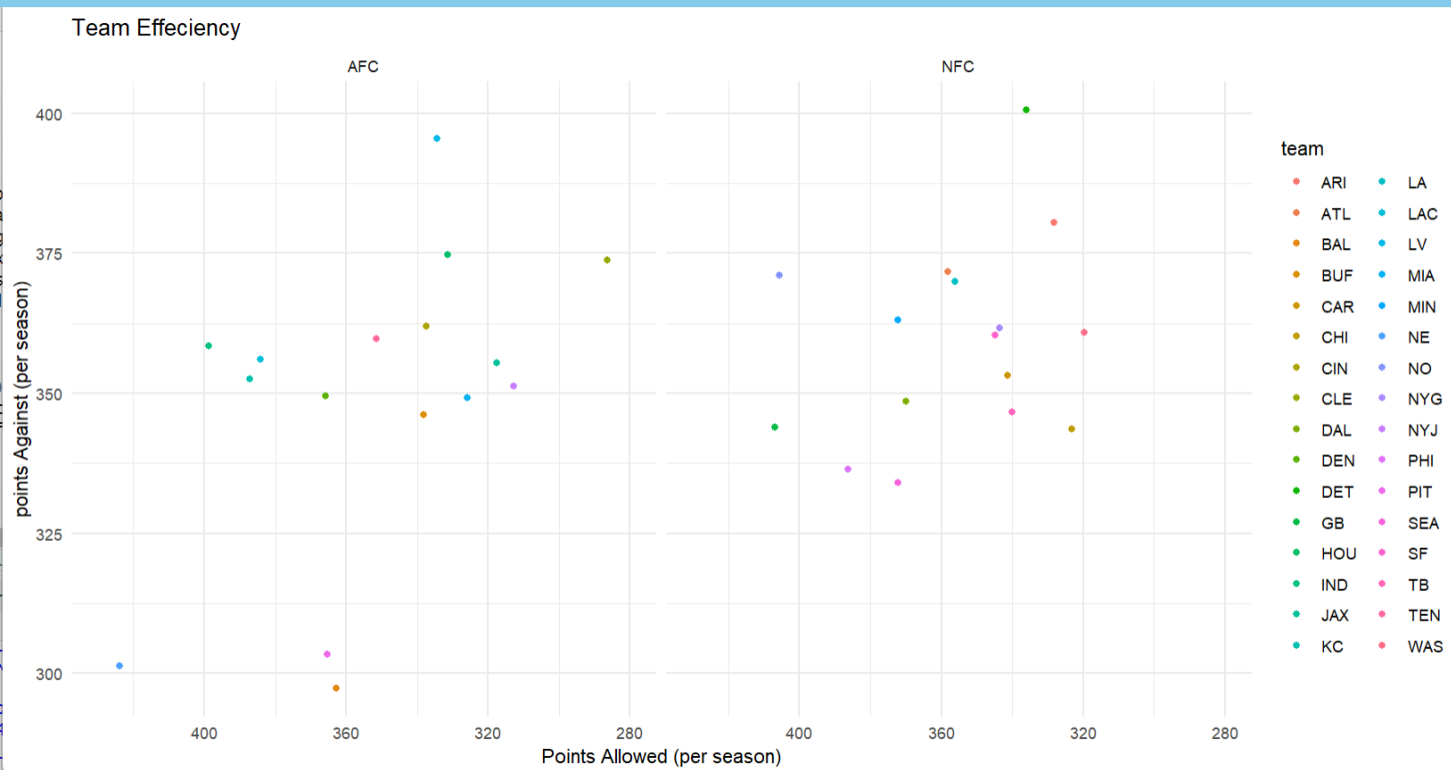

```
NFL_Clean_2 <- NFL_Clean%>%
    select(team, Conf, score_differential, points_scored,
            points_allowed, wins, losses)

ARI_data <- NFL_Clean_2%>%
    filter(
        team == 'ARI'
    )%>%
    summarise(
        points_scored = mean(points_scored),
        points_allowed = mean(points_allowed),
        wins = mean(wins),
        losses = mean(losses),
        point_differential = mean(score_differential)
    )%>%
    mutate(
        team = 'ARI',
        conf = 'NFC'
    )
```

```
ggplot(NFL_Clean_3) +
    aes(x = points_scored, y = points_allowed, colour = team) +
    geom_point() +
    xlim(430, 280) +
    scale_color_hue(direction = 1) +
    labs(
        title = "Team Effeciency",
        x = "Points Allowed (per season)",
        y = "points Against (per season)"
    ) +
    theme_minimal() +
    facet_wrap(vars(conf))
```

Efficiency Tables

AFC Team Effeciency			
team	point_differential	wins	losses
JAX	-38	6	10
IND	40	10	6
HOU	-43	7	9
TEN	-8	8	8
PIT	62	10	6
BAL	65	9	7
CLE	-87	5	11
CIN	-25	7	9
DEN	16	9	8
KC	35	9	7
LAC	28	8	8
LV	-61	7	10
NE	123	11	5
BUF	-8	8	8
NYJ	-39	7	9
MIA	-23	8	8

NFC Team Effeciency			
team	point_differential	wins	losses
MIN	9	8	8
CHI	-20	8	9
DET	-65	6	10
GB	63	10	6
NO	35	9	7
CAR	-12	8	8
ATL	-14	8	8
TB	-7	8	9
DAL	21	9	8
NYG	-18	8	8
WAS	-41	7	9
PHI	50	9	7
ARI	-52	7	9
SF	-16	8	8
SEA	38	9	7
LA	-14	8	8

```
AFC_table <- NFL_Clean_3%>%
  filter(conf == 'AFC')%>%
  select(team, point_differential, wins, losses)%>%
  mutate(
    point_differential = round(point_differential, digits = 0),
    wins = round(wins, digits = 0),
    losses = round(losses, digits = 0)
  )%>%
  kable(caption = "AFC Team Effeciency")%>%
  kableExtra::kable_classic()
```

```
NFC_table <- NFL_Clean_3%>%
  filter(conf == 'NFC')%>%
  select(team, point_differential, wins, losses)%>%
  mutate(
    point_differential = round(point_differential, digits = 0),
    wins = round(wins, digits = 0),
    losses = round(losses, digits = 0)
  )%>%
  kable(caption = "NFC Team Effeciency")%>%
  kableExtra::kable_classic()
```

Challenges

- 32 Teams is too difficult to see in our data visualizations at once, so we needed to separate them by conference and division using the facet function

- We needed to focus on a couple of variables as 54 quantitative variables is too many for one visualization.

- There was so much information in our data set that it was difficult to decide which data visualizations to use.

- There were a couple of variables that had incomplete data such as offensive average air yards, which was NA.

Questions