

Data Visualization

2025-01-09

```
library(dplyr)
library(ggplot2)

# Load necessary libraries
library(dplyr)
library(ggplot2)
getwd()

## [1] "/Users/brynhaden/Desktop/538"

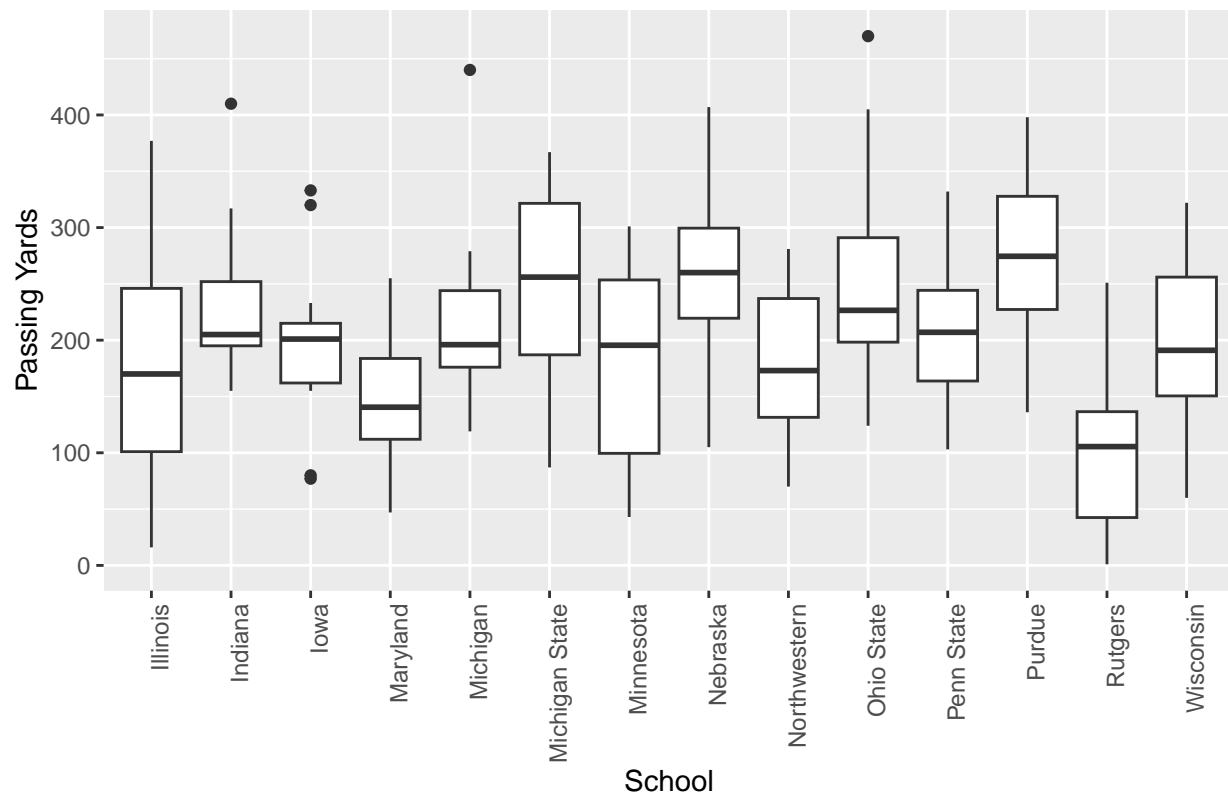
setwd("~/Desktop/538")
cleaned_stats <- read.csv("CleanedGameStats.csv")
View(cleaned_stats)

# Visual #1: Side by side boxplots for one variable for all schools in a chosen conference
# Big Ten Conference
big_ten_schools <- c("Illinois", "Indiana", "Iowa", "Maryland", "Michigan", "Michigan State",
                    "Minnesota", "Nebraska", "Northwestern", "Ohio State", "Penn State",
                    "Purdue", "Rutgers", "Wisconsin")

# Filter the dataset for Big Ten schools
big_ten_stats <- cleaned_stats %>%
  filter(Home %in% big_ten_schools)

# Create a boxplot for HPassYds
ggplot(big_ten_stats, aes(x = Home, y = HPassYds)) +
  geom_boxplot() +
  labs(title = "Passing Yards for Big Ten Schools (Home Games)",
       x = "School",
       y = "Passing Yards") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

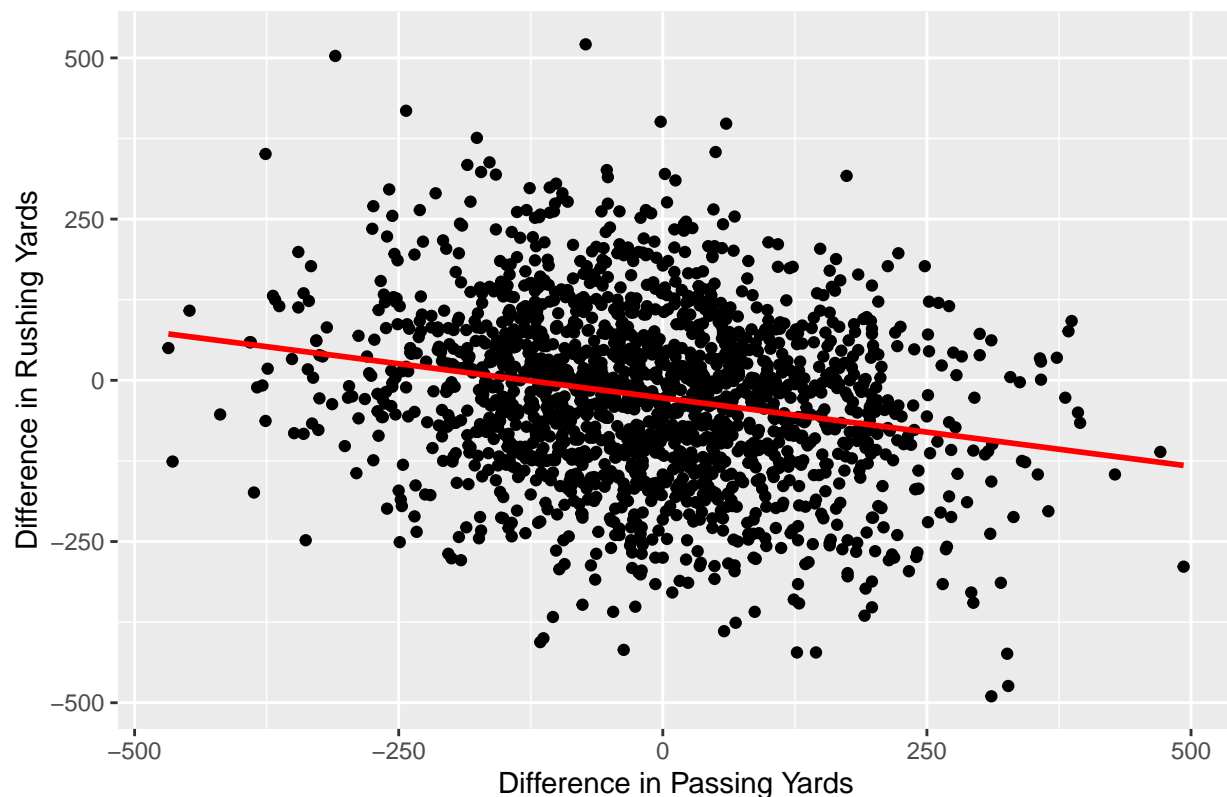
Passing Yards for Big Ten Schools (Home Games)



```
# Visual #2: Scatterplot with linear regression
# New variables: Difference in Passing Yards and Difference in Rushing Yards
cleaned_stats <- cleaned_stats %>%
  mutate(PassYds_Diff = HPassYds - APassYds,
         RushYds_Diff = HRushYds - ARushYds)

# Create a scatterplot with linear regression line
ggplot(cleaned_stats, aes(x = PassYds_Diff, y = RushYds_Diff)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Difference in Passing Yards vs. Difference in Rushing Yards",
       x = "Difference in Passing Yards",
       y = "Difference in Rushing Yards")
```

Difference in Passing Yards vs. Difference in Rushing Yards



```
# Visual #3: Bar plot summarizing one variable for each school in a different conference
#ACC Conference for this example
acc_schools <- c("Boston College", "Clemson", "Duke", "Florida State", "Georgia Tech",
                 "Louisville", "Miami", "North Carolina", "NC State", "Pittsburgh",
                 "Syracuse", "Virginia", "Virginia Tech", "Wake Forest")

# Filter the dataset for ACC schools
acc_stats <- cleaned_stats %>%
  filter(Home %in% acc_schools)

# Summarize the average number of home points for each school in the ACC
acc_home_points <- acc_stats %>%
  group_by(Home) %>%
  summarize(Average_Home_Points = mean(HKickPts, na.rm = TRUE))

# Create a bar plot for the average number of home points
ggplot(acc_home_points, aes(x = reorder(Home, Average_Home_Points), y = Average_Home_Points)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Home Points for ACC Schools",
       x = "School",
       y = "Average Home Points") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Average Home Points for ACC Schools

